# The spatial structure of correlated neuronal variability

Robert Rosenbaum[1,2], Matthew A Smith[3–5], Adam Kohn[6,7], Jonathan E Rubin[5,8] & Brent Doiron[5,8]

**Shared neural variability is ubiquitous in cortical populations. While this variability is presumed to arise from overlapping synaptic input, its precise relationship to local circuit architecture remains unclear. We combine computational models and *in vivo* recordings to study the relationship between the spatial structure of connectivity and correlated variability in neural circuits. Extending the theory of networks with balanced excitation and inhibition, we find that spatially localized lateral projections promote weakly correlated spiking, but broader lateral projections produce a distinctive spatial correlation structure: nearby neuron pairs are positively correlated, pairs at intermediate distances are negatively correlated and distant pairs are weakly correlated. This non-monotonic dependence of correlation on distance is revealed in a new analysis of recordings from superficial layers of macaque primary visual cortex. Our findings show that incorporating distance-dependent connectivity improves the extent to which balanced network theory can explain correlated neural variability.**



The spiking activity of cortical neurons is often characterized by their average response over a large number of trials, prompting a wealth of theoretical studies relating the structure of neuronal networks to their trial-averaged firing rate dynamics[1]. However, trial averages do not capture the stochastic and irregular dynamics characteristic of cortical populations and the nervous system in general[2]. Indeed, trial-to-trial fluctuations are central to contemporary theories of cortical computation[3,4]. A deep mechanistic understanding of neuronal variability remains an open challenge.

Early theoretical studies deduced that variable spiking activity could arise through a balancing of strong, yet opposing, excitatory and inhibitory synaptic inputs[5,6]. Expanding on this conjecture, van Vreeswijk and Sompolinsky[7] showed that networks of recurrently coupled model neurons robustly create a state where strong excitation is approximately balanced by inhibition, creating a push-pull dynamic that generates irregular spiking activity. More recently, balanced networks have been implicated in theories of optimal coding[8], working memory[9] and stimulus tuning[10]. Numerous experimental studies have established that excitation is often approximately balanced by inhibition in cortical circuits[11–17]. In sum, balanced networks provide a parsimonious model of the irregular spiking activity observed in cortical circuits.

Early balanced network models produced asynchronous activity through sparse connectivity[7,18]. However, several experimental studies reveal that local cortical networks are densely connected, with connection probabilities between nearby neurons sometimes exceeding 40 percent[19–22]. These data imply substantial overlap between local synaptic inputs, which could, in principle, synchronize cortical networks. However, counter to intuition, balanced networks with dense connectivity show weak spike train correlations[23]. This 'asynchronous state' results from the correlated excitatory (e) or inhibitory (i) afferents to neuron pairs being actively cancelled by a strong negative e–i correlation, establishing weak correlations even when connectivity is not sparse[23].

Consistent with the predicted asynchronous state, some multiunit extracellular recordings show noise correlations that are nearly zero on average[24]. However, a majority of population recordings in cortex reveal comparatively large correlations[25,26]. Several studies suggest that the magnitude of noise correlations is dependent on many factors[27], including arousal[28], attention[29], anesthetic state[23,24,30,31] and cortical layer[32,33]. Finally, while *in vivo* whole-cell recordings reveal strong positive e–e and i–i correlations coexisting with strong e–i correlations[13], these correlation sources do not always perfectly cancel as predicted by some theoretical models[28]. Taken together, these studies show that cortical circuits can exhibit both weak and moderate noise correlations, at odds with predictions from the current theory of balanced networks[23].

In this study, we generalize the theory of correlations in densely connected, balanced networks to include the widely observed dependence of synaptic connection probability on distance[21,34]. We show that spatially broad recurrent projections disrupt the asynchronous state, producing a signature spatial correlation structure: nearby pairs of neurons are positively correlated on average, pairs at intermediate distances are negatively correlated and distant pairs are weakly correlated. These positive and negative correlations cancel so that the average correlation between pairs sampled randomly over a large range of distances is nearly zero. We uncover this non-monotonic dependence of correlation on distance in recordings from superficial layers of macaque primary visual cortex, but only after correcting for a latent source of shared fluctuations. Our findings decouple balanced excitation and inhibition from asynchronous network

[1]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana, USA. [2]Interdisciplinary Center for Network Science and Applications, University of Notre Dame, Notre Dame, Indiana, USA. [3]Department of Ophthalmology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [4]Fox Center for Vision Restoration, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [5]Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania, USA. [6]Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, Yeshiva University, Bronx, New York, USA. [7]Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Yeshiva University, Bronx, New York, USA. [8]Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Correspondence should be addressed to R.R. (robert.rosenbaum@nd.edu) or B.D. (bdoiron@pitt.edu).

Received 28 March; accepted 28 September; published online 31 October 2016; doi:10.1038/nn.4433

activity, greatly extending the applicability of balanced network theory to explaining cortical dynamics.

## RESULTS

We consider a network of excitatory and inhibitory exponential integrate-and-fire model neurons. Neurons provide recurrent, lateral synaptic input to one another and receive feedforward synaptic input from a nonlocal presynaptic population. A detailed mathematical analysis of correlations in the limit of large network size is provided in **Supplementary Note 1** and **Supplementary Figures 1–3**. Below we provide an outline of these theoretical results and confirm their predictions using computer simulations. We first use a simplified network model to demonstrate how the asynchronous state considered in previous theoretical work[23] is broken by heterogeneous input correlations[35]. We then consider a more realistic model where neurons belong to a continuous spatial domain and connection probability depends on distance.

### Homogeneous input correlations are cancelled by balanced networks

To demonstrate the mechanisms affecting correlations in recurrent networks, we first simulated a simplified network of $N = 20{,}000$ neurons, half excitatory and half inhibitory, that all receive the same fluctuating feedforward input and are each connected with probability 0.25 (**Fig. 1a**). Despite the fact that neuron pairs share all of their feedforward input and 25% of their recurrent synaptic input on average, spiking activity was asynchronous, with an average pairwise spike count correlation of $6.8 \times 10^{-5}$ (**Fig. 1b**,**c**).

This small average correlation is a defining characteristic of the asynchronous state. Mathematically, this state is realized when spike count covariances in the network satisfy[23]

$$C_{SS} \sim O(1/N)$$

where $C_{SS}$ denotes the average spike count covariance between pairs of neurons in the recurrent network, $N$ is the number of neurons in the network and $\sim O(1/N)$ denotes asymptotic proportionality to $1/N$ for large $N$. Note that covariance and correlation scale identically with network size in balanced networks, so we discuss them interchangeably[23].

Our theoretical analysis proceeds by noting that spike count covariance is inherited from synaptic input covariance[26] and therefore the two scale similarly with $N$ in the asynchronous state,

$$C_{II} \sim O(1/N)$$

where $C_{II}$ denotes the average covariance between neurons' synaptic inputs.

Synaptic inputs can be decomposed into their feedforward and recurrent sources, $I = F + R$, so that neurons' input covariances decompose as

$$C_{II} = C_{FF} + C_{RR} + 2C_{RF}$$

where $C_{FF}$ is the average covariance between neurons' feedforward input currents, $C_{RR}$ between their recurrent inputs and $C_{RF}$ between one neuron's recurrent and the other neuron's feedforward synaptic input. Recurrent synaptic input, $R$, is composed of positive contributions from lateral excitatory synaptic inputs and negative contributions from inhibitory ($R = e - i$).

Shared input fluctuations are visualized by averaging the inputs to several neurons, so that the unshared contributions average out (**Fig. 1d**). Overlapping inputs cause $C_{FF}$ and $C_{RR}$ to be positive (**Fig. 1d**). If feedforward input correlation is moderate, $C_{FF} \sim O(1)$, then recurrent input tracks the feedforward input so that $C_{RF}$ is negative and nearly perfectly cancels the positive sources of correlations (i.e., $2C_{RF} = -(C_{FF} + C_{RR}) + O(1/N)$; see **Supplementary Note 1**). As a result, the covariance between the total synaptic inputs is weak, $C_{II} \sim O(1/N)$ (**Fig. 1d**). This cancellation arises naturally in

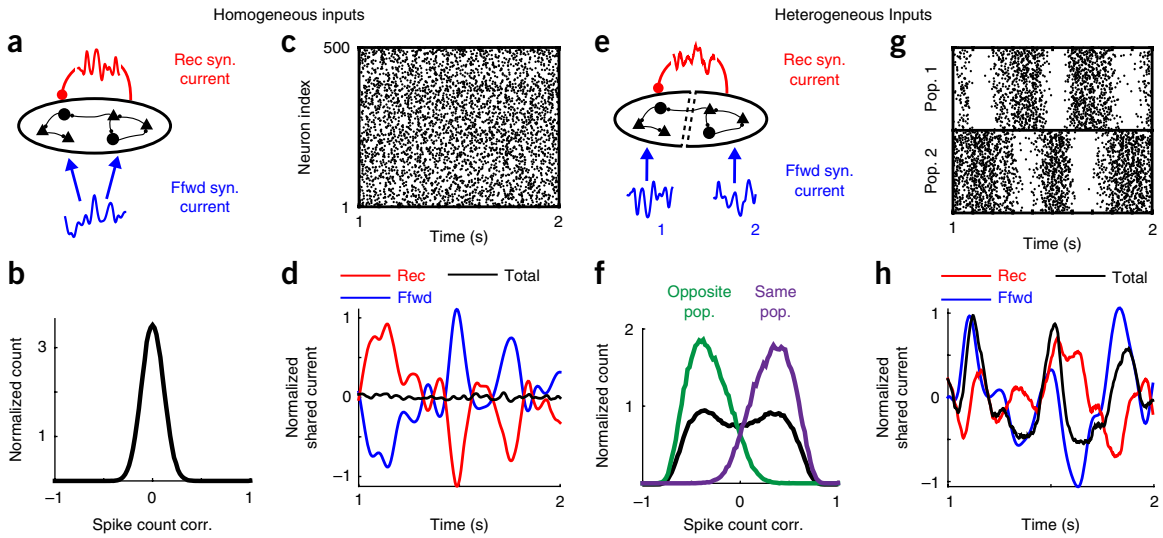



**Figure 1** Heterogeneous feedforward input breaks asynchrony in balanced recurrent networks. (**a**) Homogeneous network diagram. A population of 20,000 recurrently connected excitatory and inhibitory neurons receives globally correlated feedforward input. (**b**) Normalized histogram of pairwise spike count correlations between 1,000 randomly selected neurons. All histograms are normalized by their integral. (**c**) Raster plot of 500 randomly chosen neurons plotted over 1 s. (**d**) Shared fluctuations in the feedforward (blue) and recurrent (red) synaptic inputs cancel so that shared fluctuations in the total synaptic currents (black) are weak. Curves were computed by averaging the synaptic input currents to 500 neurons, convolving with a Gaussian-shaped kernel ($\sigma = 15$ ms), subtracting the mean and dividing by the neurons' rheobase. (**e**–**h**) Same as **a**–**d** except neurons were separated into two populations with separate feedforward inputs. Currents in **h** are from neurons in population 2. Histograms in **f** show correlations from neuron pairs randomly selected from both populations (black), from the same population (purple) and from opposite populations (green). Rec, recurrent; ffwd, feedforward; syn, synaptic; corr., correlation; pop., population.

the balanced state and does not require a precise tuning of model parameters[23]. Since spiking correlations are inherited from synaptic input covariance[26], this cancellation of input covariances leads to small, $O(1/N)$, spike count correlations.

### Heterogeneous input correlations can disrupt the asynchronous state

To study the impact of heterogeneity on correlations in balanced networks, we modified the above model by dividing the neurons into two populations. Each population received a separate feedforward input (**Fig. 1e**). The two feedforward input sources were statistically identical but uncorrelated. Recurrent connectivity was not changed: neurons were randomly connected without respect to population membership (identically to **Fig. 1a**). This input heterogeneity dramatically changed the structure of correlations in the network. Pairs of neurons in the same population had strongly positive spike count correlations on average (0.34), while neuron pairs from opposite populations were negatively correlated with a nearly identical correlation magnitude (−0.34), and the average correlation between all pairs was nearly zero ($4.2 \times 10^{-4}$; **Fig. 1f,g**).

The mechanism responsible for this change in correlations can be understood by again separating the synaptic input covariance into its recurrent and feedforward sources, but generalizing the decomposition to account for neuron 'distance' to obtain

$$C_{II}(d) = C_{FF}(d) + C_{RR}(d) + 2C_{RF}(d) \quad (1)$$

Here $C_{II}(d)$ is the average covariance between input currents to pairs of neurons separated by distance $d$, where $d = 0$ for neurons in the same population and $d = 1$ for opposite population pairs, and similarly for the other terms. Feedforward input is only correlated between neurons in the same population, so $C_{FF}(0) > 0$, but $C_{FF}(1) = 0$.

In contrast, recurrent connections do not respect population membership and thus neither do the statistics of recurrent input,

$$C_{RR}(0) = C_{RR}(1) \text{ and } C_{RF}(0) = C_{RF}(1)$$

Since covariances $C_{RR}(d)$ and $C_{RF}(d)$ do not depend on $d$ but $C_{FF}(d)$ does, cancellation cannot be achieved in equation (1) for both $d = 1$ and $d = 0$ simultaneously. In other words, the one 'copy' of shared recurrent synaptic input cannot cancel both versions of the feedforward synaptic inputs. The loss of cancellation causes the total synaptic current shared by neurons in the same population to inherit shared fluctuations from their feedforward inputs (**Fig. 1h**), giving rise to positive $O(1)$ correlations between same-population pairs. A competitive dynamic introduces negative correlations between neurons in opposite populations (**Fig. 1f,g**). A similar mechanism was considered in a recent theoretical study[35].

For illustrative purposes, we considered a simplified network model with discrete subpopulations, but correlations and connectivity in many cortical circuits depend on continuous quantities such as physical distance or tuning similarity[19,21,36]. Next, we generalize these findings to more biologically realistic networks with connection probabilities that depend on neuron distance.

### A spatially extended network model

We next considered a network of $N_e = 40{,}000$ excitatory and $N_i = 10{,}000$ inhibitory model neurons arranged on a square-shaped domain modeling a portion of a cortical layer. The neurons receive feedforward synaptic input from a separate layer of Poisson-spiking neurons and are connected with a probability that decays with distance (**Fig. 2**). Specifically, the probability of a connection between two neurons in the recurrent network obeys

$$\Pr(\text{connection}) \propto g(d;\alpha_{rec}^2)$$

where $d$ is the distance between the neurons measured along the two-dimensional network, $g(d;\sigma^2) \propto \exp(-d^2/(2\sigma^2))$ is a Gaussian-shaped function and $\alpha_{rec}$ approximately represents the average length of a recurrent synaptic projection. Similarly, the probability of a synaptic projection from a neuron in the feedforward layer to a neuron in the recurrent layer decays with distance similarly to a Gaussian with width parameter $\alpha_{ffwd}$, where distance is measured parallel to the cortical surface (**Fig. 2**). We next show that the asynchronous state is realized when $\alpha_{rec} < \alpha_{ffwd}$, then show that the asynchronous state cannot be realized when $\alpha_{rec} > \alpha_{ffwd}$.

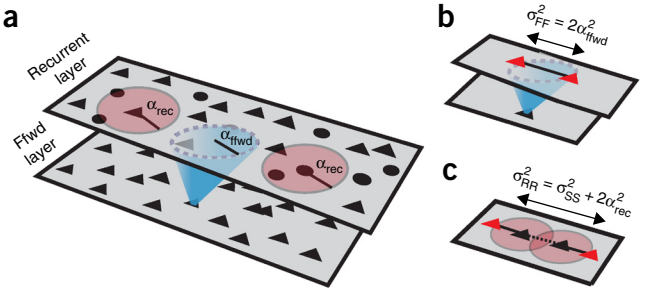



**Figure 2** Correlation and projection widths in spatially extended networks. (**a**) Network schematic. Black triangles and circles represent excitatory and inhibitory neurons. Red disks indicate recurrent synaptic projections. Recurrent connection probability decays with distance with width parameter $\alpha_{rec}$. Blue cone denotes feedforward (ffwd) synaptic projections from a separate layer, with width parameter $\alpha_{ffwd}$. (**b**) Correlations introduced by overlapping feedforward input to neurons in the recurrent layer (shared blue input to red triangles) decay with distance twice as slowly as feedforward connection probability ($\sigma^2_{FF} = 2\alpha_{ffwd}$). (**c**) The spatial width of correlations between two neurons' recurrent inputs ($\sigma^2_{RR}$; input from black triangles to red triangles) is equal to width of spike train correlations ($\sigma^2_{SS}$; dashed line) plus twice the width of recurrent projections ($2\alpha_{rec}$; solid lines).

### The asynchronous state in spatially extended cortical circuits

As above, asynchronous spiking requires a cancellation between input covariances (cf. equation (1)), except that $d$ now represents continuous instead of binary distance. Therefore, conditions on asynchrony require first an understanding of how input covariances depend on pairwise neuron distance.

Overlapping feedforward synaptic projections introduce $O(1)$ correlations between the feedforward inputs to neuron pairs. Since nearby pairs share more feedforward inputs, these correlations are distance dependent. Specifically, synaptic divergence causes feedforward input correlations to be $O(1)$ and twice as broad as synaptic projection widths (**Fig. 2b**),

$$C_{FF}(d) \propto g(d;\alpha_{ffwd}^2).$$

The fact that $C_{FF}(d) \sim O(1)$ at first seems to preclude the possibility of an asynchronous state because $C_{FF}(d)$ is one component of $C_{II}(d)$ in equation (1) and the asynchronous state requires $C_{II}(d) \sim O(1/N)$. However, the asynchronous state is realized under a cancellation between positive ($C_{FF}$ and $C_{RR}$) and negative ($C_{RF}$) sources of correlations in equation (1). Cancellation at all distances requires

that all correlation sources have the same shape (**Supplementary Note 1**), meaning that

$$C_{RF}(d), C_{RR}(d) \propto g(d; \alpha_{ffwd}^2) \quad (2)$$

The implications of this requirement on the spatial profile of spiking correlations are clarified by noting that recurrent synaptic input is generated by spike trains in the recurrent network. Synaptic divergence causes the correlations between neurons' recurrent synaptic inputs to be broader in space than the correlations between spike trains according to (**Fig. 2c**):

$$\sigma_{RR}^2 = \sigma_{SS}^2 + \alpha_{rec}^2$$

Here $\sigma_{RR}$ is the width of correlations between neurons' recurrent synaptic input currents and $\sigma_{SS}$ is the width of spike train correlations in the recurrent network. In general, we use $\alpha$ to denote the widths of synaptic projections and $\sigma$ to denote the widths of correlations.

Correlations between recurrent inputs are constrained by the cancellation required in the asynchronous state. Specifically, equation (2) requires that the width of correlations between recurrent synaptic inputs satisfy

$$\sigma_{RR}^2 = 2\alpha_{ffwd}^2$$

Combining the two expressions for $\sigma_{RR}$ above yields

$$\sigma_{SS}^2 = 2(\alpha_{ffwd}^2 - \alpha_{rec}^2) \quad (3)$$

The existence of a real solution to equation (3) requires that $\alpha_{ffwd} > \alpha_{rec}$; in other words, the spatial width of the recurrent projections must be narrower than the width of feedforward projections for the asynchronous state to exist. Further, equation (3) implies that $\sigma^2_{SS} > \sigma^2_{FF} - \alpha^2_{rec}$, so that spike train correlations are spatially narrower than correlations between feedforward input currents. Thus, recurrent dynamics actively sharpen the spatial profile of correlations in the asynchronous state (compare to the sharpening of tuning curves in previous work[37]).

To test these theoretical findings, we performed network simulations with feedforward synaptic projections broader than recurrent projections (**Fig. 3a**). The simulations confirmed that $C_{RR}(d)$ and $C_{RF}(d)$ decayed similarly with distance to $C_{FF}(d)$ (**Fig. 3b**). This allowed a cancellation between positive and negative sources of correlations, so that correlations between neurons' total synaptic currents and between their spike trains were weak over all distances (**Fig. 3b–e**). Despite their small average, spike count correlations had a larger standard deviation (s.d. = 0.11; **Fig. 3d**), consistent with results for nonspatial networks[23] (**Fig. 1b**). Neurons in the network receive strong excitation that is canceled by strong inhibition on average (**Fig. 3f** and **Supplementary Fig. 4a**,**b**), confirming that the network maintains a balanced state. Correlations computed from simulations agreed with closed-form mathematical predictions (**Fig. 3e**; see **Supplementary Note 1** for equations). Additional simulations confirmed that mean correlations decay toward zero at increasing network size (**Supplementary Note 2** and **Supplementary Fig. 4c**,**d**).

### Broad lateral connections produce a signature spatial correlation structure

As noted above, the cancellation between positive and negative correlations necessary for the asynchronous state cannot be realized when recurrent projections are broader than feedforward ($\alpha_{rec} > \alpha_{ffwd}$)

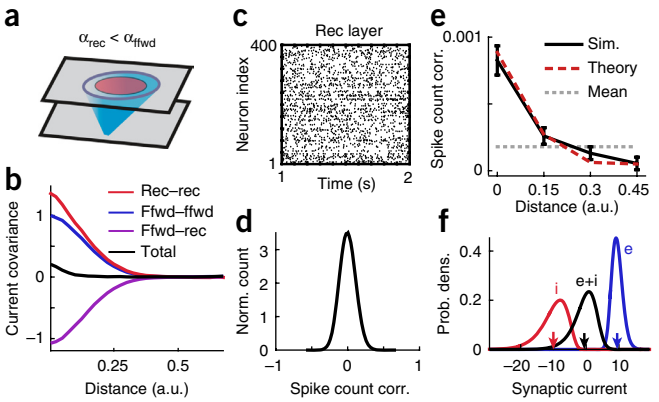



**Figure 3** The asynchronous state in a spatially extended network model. (**a**) Network schematic. As in **Figure 2** with recurrent projections narrower than feedforward projections ($\alpha_{rec} = 0.5\alpha_{ffwd}$). (**b**) Average covariance between different sources of synaptic currents to excitatory neuron pairs as a function of distance. Positive covariance between neurons' feedforward input currents (blue) and between their recurrent input currents (red) cancel with negative covariance between one neuron's feedforward and the other neuron's recurrent input (purple) to produce weak covariance between their total input (black). Curves were computed from input currents to 400 randomly selected excitatory neurons and were normalized by the peak feedforward input covariance. (**c**) Spike rasters of the 400 excitatory neurons comprising the center 20 × 20 square of neurons in the recurrent layer. (**d**) Normalized histogram of pairwise spike count correlation between 5,000 randomly selected neurons. (**e**) Mean spike count correlation between neurons (± s.e.m.) as a function of their distance. Solid black curve computed from the correlations between 5,000 randomly sampled neurons. Dashed red curve is from mathematical calculations (see **Supplementary Note 1**). Gray dashed line shows mean across all sampled pairs. (**f**) Distribution of excitatory (blue), inhibitory (red) and total (black) synaptic currents across the membranes of 400 randomly selected excitatory neurons, measured in units of the neurons' rheobase. Arrows indicate mean values. Rec, recurrent; ffwd, feedforward; corr., correlation; sim., simulation; prob. dens., probability density; e, excitatory; i, inhibitory.

because equation (3) cannot be solved in this case. Instead, neuron pairs inherit correlations from overlapping feedforward inputs so that $C_{SS}(d) \sim O(1)$.

We confirmed this prediction by numerical simulations identical to those discussed above, but with recurrent projections broader than feedforward (**Fig. 4a**). As predicted, recurrent input correlations were too spatially broad to cancel with the more sharply decaying feedforward correlations (**Fig. 4b**), so that the total input correlation between nearby neurons was large (**Fig. 4b**; compare to **Fig. 3b**). This effect introduced moderately strong correlations between nearby spike trains (**Fig. 4c–e**) that did not decay to zero at increasing network size (**Supplementary Note 2** and **Supplementary Fig. 4e**,**f**). Nevertheless, the network maintained excitatory–inhibitory balance (**Fig. 4f** and **Supplementary Fig. 4a**,**b**).

Since recurrent inputs must cancel feedforward inputs in balanced networks, $C_{RF}(d)$ is negative (**Figs. 1**, **3** and **4** and **Supplementary Note 1**). Moreover, broad recurrent projections cause $C_{RF}(d)$ to decay slowly with distance (**Fig. 4b**). Through equation (1), this imparts a non- monotonicity in the dependence of $C_{II}(d)$ on $d$ (**Fig. 4b**), and spike count correlations inherit this non-monotonic shape (**Fig. 4e**).

Following the same argument made for the homogeneous network, the spike count correlations averaged over neuron pairs at all distances is $O(1/N)$ (**Fig. 4e** and **Supplementary Note 1**). However, as noted above, the average correlation over each distance cannot be $O(1/N)$. Hence, there must be a cancellation between positive and negative

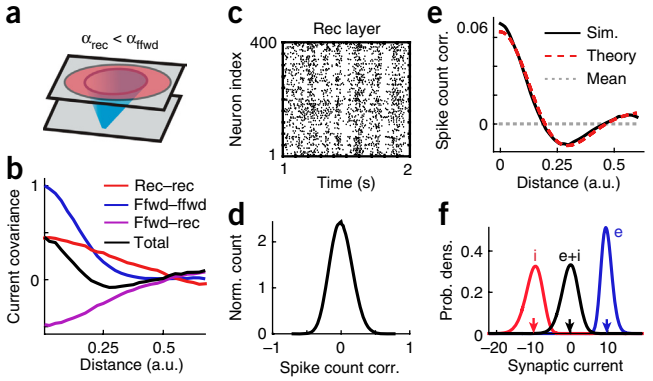



**Figure 4** Broad recurrent projections lead to a correlated balanced state. (**a**–**f**) Same as **Figure 3a**–**e** except recurrent projections were changed to be broader than feedforward projections ($\alpha_{rec} = 2.5\alpha_{ffwd}$). This change prevents the recurrent network from canceling positive feedforward input correlations (**b**), resulting in population-wide spike count correlations with increased s.d. (**c**,**d**) and with positive correlations between nearby neurons but negative correlations between more distant neurons (**e**). Nevertheless, the network maintains balance (**f**).

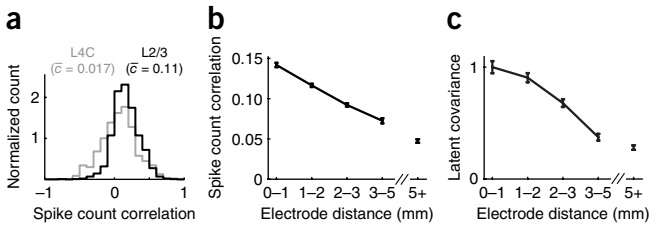



**Figure 5** Dependence of correlations on layer and distance in macaque V1. (**a**) Histogram of pairwise correlations between neurons in superficial (putative L2/3, black) and middle (putative L4C, gray) layers of macaque primary visual cortex. Legends give average correlation. (**b**) Average pairwise correlation between putative L2/3 neurons (± s.e.m.) as a function of the distance between the electrodes on which the neurons were recorded. (**c**) Average latent covariance between putative L2/3 neurons (± s.e.m.). Data points in **c** were normalized by the peak at 0–1 mm.

correlations at different distances. As in **Figure 1e**–**h**, a competitive dynamic causes nearby neurons to be positively correlated and more distant neurons to be negatively correlated. This competitive dynamic does not extend beyond the reach of recurrent projections, so sufficiently distant neurons are weakly correlated. Hence, correlation decreases and then increases with distance. This non-monotonicity can be explained more precisely using a mathematical theory of correlation transfer (**Fig. 4e** and **Supplementary Note 1**). The heterogeneity of positive and negative correlations at different distances increases the standard deviation of pairwise correlations, but only modestly (s.d. = 0.16, **Fig. 4d**; compare to **Fig. 3d**).

In summary, when recurrent projections are spatially narrower than feedforward projections ($\alpha_{rec} < \alpha_{ffwd}$, as in **Fig. 3**), correlations are weak between pairs of neurons at all distances. When recurrent projections are broader than feedforward ($\alpha_{rec} > \alpha_{ffwd}$, as in **Fig. 4**), nearby neurons are positively correlated, neurons at moderate distances are negatively correlated and distant neurons are weakly correlated. Moreover, the average correlation between pairs of neurons sampled randomly at all distances is small. The non-monotonic dependence of correlation on distance is a distinct signature of correlations arising from broad recurrent projections. We next investigated whether this correlation structure predicted by our theory is present in cortical recordings.

### Spatial correlation structure in a visual cortical circuit

We next asked whether our theoretical characterization of correlations in spatially extended networks can explain correlations in a cortical circuit. Layers 2/3 and layer 4C of macaque primary visual cortex (L2/3 and L4C) provide an ideal circuit for testing our predictions. Pairs of neurons in L2/3 exhibit moderately large noise correlations that decay with distance, but neurons in L4C, which are a primary source of interlaminar input to L2/3, exhibit extremely weak pairwise noise correlations[32,33] (**Fig. 5**, with data from previous studies[33,36]).

Neurons in L4C receive much of their feedforward input from thalamic projections, which form spatially broad synaptic fields, around 1 mm in diameter, but lateral projections within macaque L4C form narrower, sub-millimeter synaptic fields[34]. Thus, our theoretical prediction that correlations are weak when $\alpha_{ffwd} > \alpha_{rec}$ is consistent with the weak pairwise correlations observed between L4C neurons *in vivo* (as in **Fig. 3**).

Interlaminar projections from L4C to L2/3 have a similar sub-millimeter width to excitatory intralaminar projections within L4C, and lateral projections from inhibitory basket cells in L2/3 form sub-millimeter synaptic fields similar to those in L4C[34]. Excitatory neurons in L2/3, however, form long-range lateral synaptic projections with synaptic fields spanning several millimeters[34]. Our theoretical results can be generalized to this setting, where inhibitory and excitatory projections have different spatial profiles (**Supplementary Note 1**). This extension predicts the same correlation structure reported in **Figure 4**. However, correlations measured in L2/3 are positive on average over a broad range of distances[36] (**Fig. 5b**), in disagreement with this prediction.

We hypothesized that this inconsistency could be explained by recent studies showing that much of the correlated variability measured in L2/3 arises from a low-dimensional shared source of latent variability[30,31,38–40]. We conjectured that this shared variability increases pairwise correlations in L2/3 at all distances, thereby 'washing out' the negative correlations predicted by our theory. To search for low-dimensional variability in our data, we used Gaussian process factor analysis[31,41] (GPFA), a statistical algorithm that extracts shared fluctuations from a population of spike trains (see Online Methods). Applying this algorithm to our L2/3 recordings revealed a source of one-dimensional covariability that decays with distance (**Fig. 5c**). This distance dependence implies that nearby neurons are affected similarly by the latent variable.

To test whether one-dimensional latent variability explains the discrepancy between our theoretical predictions and data, we built a two-layer network model representing a 10 mm by 10 mm square of cortex (**Fig. 6a**). The first layer, representing L4C, was similar to the model in **Figure 3**, with the profile of feedforward and recurrent projections chosen to match experimentally constrained thalamic and lateral projection widths[34]. The second layer, representing L2/3, was similar to **Figure 4**, with feedforward synaptic input from excitatory neurons in the L4C model and recurrent projection widths also chosen to match anatomical measurements. To capture latent variability in L2/3, the feedforward synaptic input to each neuron in the second layer was modulated by a time-varying, multiplicative gain modulation. We chose a multiplicative source of variability to be consistent with the properties of low-dimensional variability previously reported in macaque V1 (ref. 31), but an additive source of latent variability would produce similar overall results. The gain modulation contributes an $O\left(\sqrt{N}\right)$ source of covariance to the feedforward inputs that the recurrent network cannot cancel[23]. To capture the distance dependence of latent variability (**Fig. 5c**), the magnitude of the gain modulation was heterogeneous across the network in such a way that nearby neurons received similar modulations and more distant neurons received less similar modulations.

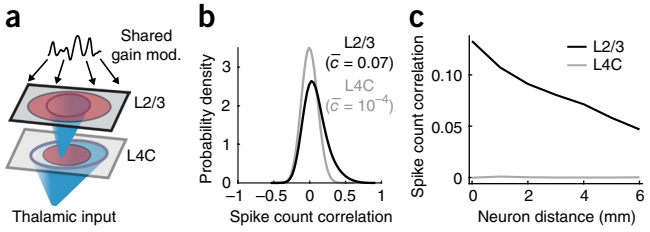



**Figure 6** Dependence of correlations on layer and distance in a spatially extended, multilayer network model. (**a**) Network schematic. Thalamic input to L4C is broader than recurrent projections within L4C. Projections from L4C to L2/3 are narrower than recurrent excitatory (but not inhibitory) projections within L2/3. Neurons in L2/3 also receive a shared gain modulation (mod.). (**b**) Histograms of pairwise correlations between randomly selected neurons in each layer. (**c**) Average pairwise correlation between neurons in each layer as a function of the distance between the neurons.

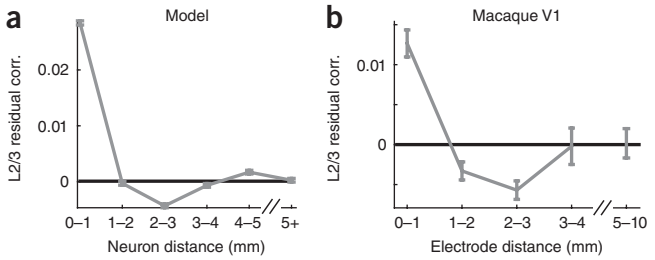



**Figure 7** Residual correlations in macaque V1 and in a model. (**a**) Residual correlation (corr.) between neurons within the model L2/3 network as a function of distance. (**b**) Residual correlations between putative L2/3 neurons in macaque primary cortex (same data as **Fig. 5b**). Residual correlation approximates spike count correlations after a single source of shared latent variability is removed. Both plots show mean ± s.e.m. Correlations decreased in the first two bins ($P < 10^{-12}$; unpaired *t*-test), increased from the third to fourth bin ($P = 0.019$) and from the third to the fifth bin ($P = 0.0067$).

Simulations of this two-layer model revealed that correlations between neurons in L4C were extremely small on average (**Fig. 6b**,**c**), consistent with our theoretical predictions (**Fig. 3**) and consistent with *in vivo* recordings (**Fig. 5**). Correlations in the model L2/3 layer were moderately large and positive over all distances (**Fig. 6b**,**c**), comparable to those in *in vivo* recordings (**Fig. 5**).

Thus, our model recovers the coarse structure of correlations in L4C and L2/3. However, our explanation of positive correlations in L2/3 is unsatisfying because the addition of globally shared variability destroys the distinct non-monotonic relationship between correlation and distance predicted by our theory (compare **Fig. 4e** to **Fig. 6b**). We next asked whether this structure could be recovered by filtering out globally shared variability. To accomplish this, we computed the residual correlation matrix estimated by GPFA. Residual correlations approximate the spike count correlations with the contribution from low-dimensional variability removed[31].

Residual correlations computed between the simulated L2/3 spike trains exhibited the predicted non-monotonic dependence on distance, corroborating the ability of the GPFA algorithm to extract low-dimensional variability and leave the structure of residual correlations intact. We next computed the mean residual correlation in macaque L2/3 as a function of electrode distance. In doing so, we observed the same non-monotonic dependence of residual correlation on distance predicted by our theory (**Fig. 7b**; further statistical analysis in **Supplementary Note 3** and **Supplementary Fig. 5**).

In summary, combining theoretical analysis and computer simulations of a multilayer network reveals a parsimonious model of the sources of shared variability in a visual cortical circuit *in vivo*. Under this model, positive correlations introduced by shared thalamic inputs to L4C neurons are actively canceled by negative correlations arising from recurrent circuitry so that pairs of L4C neurons at all distances exhibit weak average spike count correlations[32,33]. Correlations between neurons in L2/3 are introduced by overlapping feedforward inputs from L4C and a low-dimensional source of variability. Correlations arising from overlapping L4C projections are filtered by recurrent circuitry in L2/3 to promote a non-monotonic dependence of correlation on distance. This non-monotonic correlation structure is washed out by low-dimensional latent variability, but can be recovered using GPFA to estimate and remove this variability.

## DISCUSSION

Previous theoretical work on spatially homogeneous balanced networks with dense connectivity shows that they produce very weak spike train correlations[23]. We have generalized this theory to account for heterogeneous inputs and distance-dependent connection probability. In this framework we have made two notable discoveries.

First, in agreement with the original findings, when lateral synaptic projections are spatially narrower than incoming feedforward projections, correlations are extremely weak on average at all distances. This theoretical finding can explain the weak pairwise correlations observed between neurons in middle layers of macaque primary visual cortex[32,33]. However, correlations measured in cortical recordings are not always weak[25]. Second, networks with broader lateral than feedforward projections produce correlations that do not decay to zero at increasing network size.

In previous studies of balanced networks with spatially homogeneous or clustered connectivity[23,42], the asynchrony condition $C_{SS} \sim O(1/N)$ is satisfied and population averaged pairwise correlations vanish in the large network limit. In contrast, spatially extended networks with broad lateral projections violate the asynchrony condition, and consequently the expected pairwise correlations at a specific distance do not vanish. Nonetheless, mean excitatory and inhibitory currents balance and firing rates are moderate even when the asynchrony condition is violated (**Supplementary Fig. 4a**,**b** and **Supplementary Note 1**). This represents a novel solution for balanced networks that, for the first time, formally decouples network-wide asynchrony from excitatory–inhibitory balance.

We focused on the dependence of correlations on distance, but correlations also depend on tuning similarity. Partitioning L2/3 neuron pairs by tuning similarity reveals that correlations are strongest between similarly tuned neurons[36] (**Supplementary Fig. 6a**,**b**). Modifying our computational model to capture tuning-dependent correlations produced a non-monotonic dependence of residual correlation on tuning similarity in some parameter regimes, but the relevant parameters have not been measured experimentally (**Supplementary Fig. 6c–e** and **Supplementary Note 4**). Nevertheless, the modified theory could explain negative correlations previously observed in computer simulations of networks with tuning-specific connectivity[32] and the finding that negative correlations are more frequent between disparately tuned neurons in V1 (ref. 43).

As with nearly any computational model, many of the parameters used in our simulations may not reflect their corresponding values in specific cortical areas of specific species. However, our theoretical analysis does not depend on the precise values of these

parameters. Our finding that the asynchronous state requires $\alpha_{rec} < \alpha_{ffwd}$ is a fundamental property of networks with balanced excitation and inhibition.

We used a simplified model of a visual cortical circuit. In reality, pyramidal neurons in V1 form both local and long-range projections, connection probability in primate V1 depends on both distance and tuning similarity, and these dimensions are coupled[44]. Moreover, connectivity properties of inhibitory neurons depend on their subtype[45]. We modeled unidirectional connections from L4 to L2/3, but L4 also receives indirect feedback from L2/3 through deeper cortical layers. Spike trains in our model feedforward layer were modeled by homogeneous Poisson processes, in contrast to the oscillatory firing rates evoked by drifting grating stimuli in the data we analyzed. Our model can be extended to account for these additional features without affecting our overall conclusions.

Our findings have important implications for the interpretation of correlations in neural recordings. The average (residual) correlation between cell pairs sampled across a large range of distances could be extremely small, even when nearby pairs are positively correlated with moderate magnitude (**Figs. 4** and **7**). Hence, subtracting low-dimensional latent variability and partitioning neuron pairs by distance can reveal correlation structure that would otherwise not be apparent. A previous study[31] computed residual correlations as a function of distance in primate V1, but did not report a non-monotonic dependence. While we cannot be certain why their findings differ from ours, the accurate estimation of residual correlations with GPFA depends on the amount of data used to estimate shared variability. Our data are well-suited for this purpose, as they contain over 800 pairs of units per recording on average.

There is a long history of computational models of cortical circuits that consider either networks with spatially dependent coupling[1] or balanced excitation and inhibition in spatially homogeneous networks[7,23]. Only recently has the spatial structure of cortical connectivity been included in networks with balanced excitation and inhibition[9,37,46], and guiding theoretical principles are lacking. Our theory has taken this spatial structure into account and produced two core predictions for cortical circuits with long range lateral connections: first, nearby neurons exhibit significant positive correlations; second, the dependence of pairwise correlation on pairwise distance is non-monotonic. These predictions are clearly falsifiable and hence represent strong tests of our theory. The superficial layers of visual cortex have long-range lateral connections[34], making them a suitable test bed for our theory of correlations. After accounting for a source of global variability, both of our predictions were verified from population recordings in macaque V1 (**Fig. 7**). Further, a similar noise correlation structure has been reported in recordings from mouse V1 (ref. 47). The successful validation of our predictions marks our theory as a promising framework for studying the structure of neural variability in cortical circuits. Nevertheless, there are many aspects of cortical dynamics that remain unexplained by balanced networks, such infrequent yet large membrane fluctuations during spontaneous dynamics[15,48]. Capturing these dynamics in cortical models with balanced architectures remains an open challenge.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*


## ACKNOWLEDGMENTS

We are grateful to T.S. Lee and A. Movshon for research support. This work was supported by National Science Foundation grants NSF-DMS-1517828 (R.R.), NSF-DMS-1313225 (B.D.), NSF-DMS-1517082 (B.D.), NSF-DMS-1612913 (J.E.R.), NSF-DMS-1516288 (J.E.R.) and NSF-DMS-1312508 (J.E.R.); National Institute of Health grants R01NS070865 (B.D., J.E.R.), CRCNS-R01DC015139 (B.D.), R01EY016774 (A.K.), R01EY022928 (M.A.S.) and P30EY008098 (M.A.S.); two grants from the Simons Foundation collaboration on the global brain (SCGB#325293MC;BD, B.D. and 364994AK, A.K.); by the Eye and Ear Foundation of Pittsburgh (M.A.S.); and by Research to Prevent Blindness (A.K., M.A.S.).


## AUTHOR CONTRIBUTIONS

R.R. and B.D. conceived the project; R.R. performed the simulations, data analysis and mathematical calculations; M.A.S. and A.K. provided the experimental data. J.E.R. and B.D. supervised the project. All authors contributed to writing the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.




1. Ermentrout, B. Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.* **61**, 353 (1998).
2. Faisal, A.A., Selen, L.P. & Wolpert, D.M. Noise in the nervous system. *Nat. Rev. Neurosci.* **9**, 292–303 (2008).
3. Pouget, A., Beck, J.M., Ma, W.J. & Latham, P.E. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16**, 1170–1178 (2013).
4. Shamir, M. Emerging principles of population coding: in search for the neural code. *Curr. Opin. Neurobiol.* **25**, 140–148 (2014).
5. Softky, W.R. & Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334–350 (1993).
6. Shadlen, M.N. & Newsome, W.T. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569–579 (1994).
7. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
8. Denève, S. & Machens, C.K. Efficient codes and balanced networks. *Nat. Neurosci.* **19**, 375–382 (2016).
9. Lim, S. & Goldman, M.S. Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. *J. Neurosci.* **34**, 6790–6806 (2014).
10. Hansel, D. & van Vreeswijk, C. The mechanism of orientation selectivity in primary visual cortex without a functional map. *J. Neurosci.* **32**, 4049–4064 (2012).
11. Shu, Y., Hasenstaub, A. & McCormick, D.A. Turning on and off recurrent balanced cortical activity. *Nature* **423**, 288–293 (2003).
12. Haider, B., Duque, A., Hasenstaub, A.R. & McCormick, D.A. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* **26**, 4535–4545 (2006).
13. Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci.* **11**, 535–537 (2008).
14. Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E. & Froemke, R.C. Developmental sensory experience balances cortical excitation and inhibition. *Nature* **465**, 932–936 (2010).
15. Graupner, M. & Reyes, A.D. Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. *J. Neurosci.* **33**, 15075–15085 (2013).
16. Zhou, M. *et al.* Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nat. Neurosci.* **17**, 841–850 (2014).
17. Xue, M., Atallah, B.V. & Scanziani, M. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* **511**, 596–600 (2014).
18. Amit, D.J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* **7**, 237–252 (1997).
19. Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
20. Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69**, 1188–1203 (2011).
21. Levy, R.B. & Reyes, A.D. Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *J. Neurosci.* **32**, 5609–5619 (2012).
22. Oswald, A.M., Doiron, B., Rinzel, J. & Reyes, A.D. Spatial profile and differential recruitment of GABAB modulate oscillatory activity in auditory cortex. *J. Neurosci.* **29**, 10321–10334 (2009).
23. Renart, A. *et al.* The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
24. Ecker, A.S. *et al.* Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010).
25. Cohen, M.R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).

26. Doiron, B., Litwin-Kumar, R., Rosenbaum, R., Ocker, G.K. & Josić, K. The mechanics of state-dependent neural correlations. *Nat. Neurosci.* **19**, 383–393 (2016).
27. Kohn, A., Zandvakili, A. & Smith, M.A. Correlations and brain states: from electrophysiology to functional imaging. *Curr. Opin. Neurobiol.* **19**, 434–438 (2009).
28. Poulet, J.F. & Petersen, C.C. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**, 881–885 (2008).
29. Cohen, M.R. & Maunsell, J.H. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
30. Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K.D. & de la Rocha, J. Stochastic transitions into silence cause noise correlations in cortical circuits. *Proc. Natl. Acad. Sci. USA* **112**, 3529–3534 (2015).
31. Ecker, A.S. *et al.* State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014).
32. Hansen, B.J., Chelaru, M.I. & Dragoi, V. Correlated variability in laminar cortical circuits. *Neuron* **76**, 590–602 (2012).
33. Smith, M.A., Jia, X., Zandvakili, A. & Kohn, A. Laminar dependence of neuronal correlations in visual cortex. *J. Neurophysiol.* **109**, 940–947 (2013).
34. Lund, J.S., Angelucci, A. & Bressloff, P.C. Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cereb. Cortex* **13**, 15–24 (2003).
35. Wimmer, K. *et al.* The dynamics of sensory integration in a hierarchical network explains choice probabilities in MT. *Nat. Commun.* **6**, 6177 (2015).
36. Smith, M.A. & Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* **28**, 12591–12603 (2008).
37. Rosenbaum, R. & Doiron, B. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Phys. Rev. X* **4**, 021039 (2014).
38. Yu, B., Kohn, A. & Smith, M.A. Estimating shared firing rate fluctuations in neural populations. *Soc. Neurosci.* abstr. 483.18 (2011).
39. Lin, I.C., Okun, M., Carandini, M. & Harris, K.D. The nature of shared cortical variability. *Neuron* **87**, 644–656 (2015).
40. Schölvinck, M.L., Saleem, A.B., Benucci, A., Harris, K.D. & Carandini, M. Cortical state determines global variability and correlations in visual cortex. *J. Neurosci.* **35**, 170–178 (2015).
41. Yu, B.M. *et al.* Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009).
42. Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**, 1498–1505 (2012).
43. Chelaru, M.I. & Dragoi, V. Negative correlations in visual cortical networks. *Cereb. Cortex* **26**, 246–256 (2016).
44. Bosking, W.H., Zhang, Y., Schofield, B. & Fitzpatrick, D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17**, 2112–2127 (1997).
45. Pfeffer, C.K., Xue, M., He, M., Huang, Z.J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013).
46. Kriener, B., Helias, M., Rotter, S., Diesmann, M. & Einevoll, G.T. How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime. *Front. Comput. Neurosci.* **7**, 187 (2014).
47. Rikhye, R.V. & Sur, M. Spatial correlations in natural scenes modulate response reliability in mouse visual cortex. *J. Neurosci.* **35**, 14661–14680 (2015).
48. Tan, A.Y., Chen, Y., Scholl, B., Seidemann, E. & Priebe, N.J. Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature* **509**, 226–229 (2014).

## ONLINE METHODS

**Description of computational model.** We modeled a square of cortex with $N$ neurons, $N_e$ of which are excitatory and $N_i$ inhibitory. The membrane potential of neuron $j$ from the excitatory ($a$ = e) or inhibitory ($a$ = i) population obeyed exponential integrate-and-fire (EIF) dynamics,

$$C_m \frac{dV_j^a}{dt} = I_L\left(V_j^a\right) + f\left(V_j^a\right) + I_j^a(t)$$

Each time that $V_j^a$ exceeds a threshold at $V_{th}$, the neuron spikes and the membrane potential is held for a refractory period $\tau_{ref}$, then reset to a fixed value $V_{re}$. The leak current is given by

$$I_L(V) = -g_L(V - E_L)$$

and a spike-generating current is defined by

$$f(V) = g_L \Delta_T \exp[(V - V_T)/\Delta_T]$$

For excitatory neurons, $\tau_m = C_m/g_L$ = 15 ms, $E_L$ = −60 mV, $V_T$ = −50 mV, $V_{th}$ = −10 mV, $\Delta_T$ = 2 mV, $V_{re}$ = −65 mV and $\tau_{ref}$ = 1.5 ms. Inhibitory neurons were the same except $\tau_m$ = 10 ms, $\Delta_T$ = 0.5 mV and $\tau_{ref}$ = 0.5 ms.

Synaptic input currents were defined by

$$C_m^{-1} I_j^a(t) = F_j^a(t) + R_j^a(t)$$

where $F^a(t)$ is the feedforward input and $R^a(t)$ the recurrent input to neuron $j$ in population $a$ = e, i. The feedforward input was modeled differently for different figures, as described below. The recurrent input was defined by

$$R_j^a(t) = \sum_{b=e,i} \sum_{k=1}^{N_b} \frac{J_{jk}^{ab}}{\sqrt{N}} \sum_n \eta_b\left(t - t_n^{b,k}\right) \quad (4)$$

where $t_n^{b,k}$ is the $n$th spike time of neuron $k$ in population $b$ = e, i. The $1/\sqrt{N}$ scaling of synaptic weights is a defining feature of the balanced network formalism and captures the balance between excitatory and inhibitory currents as well as intrinsically generated temporal variability for large $N$ (ref. 7). Each term $J_{jk}^{ab}$ represents the synaptic weight from presynaptic neuron $k$ in population $b$ to postsynaptic neuron $j$ in population $a$. For all simulations, we modeled synaptic kinetics using

$$\eta_b(t) = \exp(-t/\tau_b)/\tau_b \text{ for } t > 0$$

where $\tau_e$ = 6 ms and $\tau_i$ = 5 ms. All networks were 'dense' in the sense that connection probabilities are $O(1)$ (ref. 23).

For the model in **Figure 1**, there were $N$ = 20,000 neurons, half of which were excitatory and half inhibitory. For each (presynaptic) neuron, we randomly and uniformly chose 2,500 excitatory and 2,500 inhibitory postsynaptic neurons in the network. Postsynaptic neurons were sampled with replacement, so that a single presynaptic neuron could make multiple contacts with a postsynaptic neuron. The synaptic weight of each connection depended on the pre- and postsynaptic neuron types (excitatory or inhibitory). Specifically,

$$J_{jk}^{ab} = (\text{number of contacts}) \times j_{ab} \quad (5)$$

where $j_{ee}$ = 12.5 mV, $j_{ie}$ = 20 mV and $j_{ii} = j_{ei}$ = −50 mV. Note that synaptic weights were scaled by $\sqrt{N}$ = 141 in equation (4), so that the actual synaptic weight of each contact was on the order of 0.1 mV.

For **Figure 1a–d**, the feedforward input to each neuron was given by the sum of an input bias and a smoothly varying signal,

$$F_j^a(t) = \sqrt{N} m_a + \sigma_s s(t)$$

Here, $s(t)$ is a shared source of smooth, unbiased Gaussian noise defined by its auto-covariance function,

$$\text{cov}(s(t), s(t+\tau)) = \exp\left(-\tau^2/2\tau_s^2\right),$$

$\tau_s$ = 40 ms sets the correlation timescale and $\sigma_s$ = 0.1 mV/ms scales the magnitude of the fluctuations. The terms $m_e$ = 0.015 mV/ms and $m_i$ = 0.01 mV/ms introduce a static bias to the input current. The model for **Figure 1e–h** was identical, except that two independent realizations, $s_1(t)$ and $s_2(t)$, of the shared input were generated. Half of the neurons received $s_1(t)$ and the other half received $s_2(t)$. Firing rates for **Figure 1a–d** were 7.6 Hz on average for excitatory neurons and 3.8 Hz for inhibitory neurons. For **Figure 1e–h**, average firing rates were 7.4 Hz for excitatory and 3.8 Hz for inhibitory neurons.

To model the spatially extended recurrent network in **Figures 3** and **4**, we arranged $N_e$ = 40,000 excitatory and $N_i$ = 10,000 inhibitory EIF model neurons on a uniform grid covering a two- dimensional square domain. The feedforward layer was modeled by a population of $N_F$ = 5,625 excitatory Poisson-spiking neurons on a uniform grid covering a square that is parallel to the recurrent network. Feedforward input to the recurrent layer was defined by

$$F_j^a(t) = \sum_{k=1}^{N_F} \frac{J_{jk}^{aF}}{\sqrt{N}} \sum_n \eta_e\left(t - t_n^{F,k}\right)$$

where $t_n^{F,k}$ is the $n$th spike time of neuron $k$ in the feedforward layer. Each spike train in the feedforward layer was modeled as independent Poisson processes with rate $r_F$ = 5 Hz.

To simplify calculations, we measured distance in units of the side-length of the square domain. In these units, the domain is represented as the unit square, $\Gamma = [0, 1] \times [0, 1]$. Neurons were connected randomly and the probability that two neurons were connected depended on their distance measured periodically on $\Gamma$. The precise algorithm for generating connections is described in **Supplementary Note 5**. This algorithm assures that the expected number of synaptic contacts from a presynaptic neuron at coordinates $\mathbf{y} = (y_1, y_2)$ in population $b$ to a postsynaptic neuron at $\mathbf{x} = (x_1, x_2)$ in population $a$ is given by

$$p_{ab}(\mathbf{x} - \mathbf{y}) = \frac{K_{ab}^{out}}{N_a} g(x_1 - y_1; \alpha_b)\, g(x_2 - y_2; \alpha_b)$$

where $g(u;\alpha)$ is a wrapped Gaussian distribution[37]. Out-degrees were $K_{ee}^{out} = K_{ei}^{out} = 2{,}000$, $K_{ie}^{out} = K_{ii}^{out} = 500$, $K_{eF}^{out} = 10{,}000$ and $K_{eF}^{out} = 800$. It follows that the network-wide average number of synaptic inputs to excitatory and inhibitory neurons in the recurrent network was $K$ = 3,715. Synaptic weights were determined by equation (5) where $j_{ee}$ = 40 mV, $j_{ie}$ = 120 mV, $j_{ei} = j_{ii}$ = −400 mV, and $j_{eF} = j_{iF}$ = 120 mV. Note again that these terms were divided by $\sqrt{N} \approx 224$ as indicated in equation (4), so that the actual synaptic weights were between 0.18 mV and 1.8 mV.

Excitatory and inhibitory recurrent projection widths were $\alpha_{rec} = \alpha_e = \alpha_i$ = 0.05 for **Figure 3** and $\alpha_{rec}$ = 0.25 for **Figure 4**. Feedforward connection widths were $\alpha_{ffwd}$ = 0.1 in both figures. For the simulations in **Figure 3**, average firing rates were 3.9 Hz for excitatory and 6.2 Hz for inhibitory neurons. For the simulations in **Figure 4**, average rates were 4.0 Hz for excitatory and 6.1 Hz for inhibitory neurons.

The first layer (L4C) in the model in **Figure 6** was identical to the model in **Figure 3** except that $\alpha_{ffwd}$ = 0.1, $\alpha_e$ = 0.05 and $\alpha_i$ = 0.03. The length units used in **Figures 6** and **7a** were determined by interpreting the network domain, $\Gamma$, as a 10 mm by 10 mm square. Thus, in physical dimensions, $\alpha_{ffwd}$ = 1 mm, $\alpha_e$ = 0.5 mm and $\alpha_i$ = 0.3 mm. Average firing rates in the L4C layer were 3.7 Hz for excitatory and 6.1 Hz for inhibitory neurons.

Connectivity in the second layer (L2/3) in **Figure 6** was identical to that in the L4C layer except that $\alpha_e$ = 0.15 (or 1.5 mm), $\alpha_{ffwd}$ = 0.05 (or 0.5 mm). The spike times from equation (6) for the L2/3 layer were given by the spike times of neurons in the L4C layer, so that $N_F$ = 50,000. The magnitude of feedforward connectivity was also modified by setting $K_{eF}$ = 1,406, $K_{iF}$ = 113 and $j_{eF} = j_{iF}$ = 220 mV. A shared gain modulation was implemented by altering feedforward input currents according to

$$F_j^a(t) \leftarrow F_j^a(t)\left[1 + w_L(\mathbf{x}) L(t)\right]$$

The shared gain modulation, $L(t)$, is a realization of unbiased Gaussian noise defined by its auto-covariance function

$$\text{cov}(L(t), L(t+\tau)) = \exp\left(-\tau^2/2\tau_L^2\right)$$

with correlation timescale $\tau_L$ = 40 ms. The dimensionless weight factor $w_L(\mathbf{x})$ depended on the coordinates, $\mathbf{x} = (x_1, x_2) \in \Gamma$, of the neuron and was given by

$$w_L(x_1, x_2) = 0.5 g(x_1 - c; \sigma_L)\, g(x_2 - c; \sigma_L)$$

where $c = 0.5$ and $\sigma_L = 0.25$ (in physical dimensions, $c = 5$ mm and $\sigma_L = 2.5$ mm) so that neurons near the center of the network receive a more intense gain modulation and the strength of the modulation decays slowly with distance from the center. This imparted a long-range distance dependence to the correlations (**Fig. 6c**), similar to that observed in the cortical recordings (**Fig. 5b**). Average firing rates in the L2/3 layer were 5.7 Hz for excitatory and 9.1 Hz for inhibitory neurons.

All simulations and numerical computations were performed on a MacBook Pro running OS X 10.9.5 with a 2.3 GHz Intel Core i7 processor. All simulations were written in a combination of C and Matlab (Matlab R 2014b, MathWorks). The differential equations defining the neuron model were solved using a simple forward Euler method with time step 0.1 ms and all simulations were run for a duration of 22 s.

**Experimental methods.** Anesthesia was induced with ketamine (10 mg/kg) and maintained during preparatory surgery with isoflurane (1.0–2.0% in 95% O2). Anesthesia during recordings was maintained with sufentanil citrate (6–18 $\mu g \cdot kg^{-1} \cdot h^{-1}$). Vecuronium or pancuronium bromide (0.1–0.15 $mg \cdot kg^{-1} \cdot h^{-1}$) was used to suppress eye movements. Physiological signs were monitored to ensure adequate anesthesia and animal well-being. Vital signs (EEG, ECG, blood pressure, end-tidal $P_{CO2}$, temperature, and airway pressure) were monitored continuously. We used supplementary lenses to bring the retinal image into focus. At the end of the recording session, animals were killed and tissue was processed histologically to verify recording locations. All procedures were approved by the Institutional Animal Care and Use Committee of the Albert Einstein College of Medicine.

The pairwise correlations used to make the distributions in **Figure 5a** were computed using the same methods as those reported previously[33] and included 1,613 L2/3 pairs and 469 L4 pairs. The data were recorded from nine anesthetized, adult male macaque monkeys (*Macaca fascicularis*). Recording were made with a group of five to seven linearly arranged (305 µm spacing) platinum-tungsten electrodes or tetrodes (Thomas Recording), inserted normally to the cortical surface. The electrodes were advanced together through cortex, sampling in 200-µm intervals until all electrodes had exited into white matter. Location of the middle layer sites involved a number of criteria, including nominal depth, histology, and CSD analysis, as detailed previously[33].

Data for distance-dependent correlations reported in **Figures 5b**,**c** and **7b** were from recordings described previously[36]. The data were recorded from seven hemispheres of four adult male macaque monkeys (*M. fascicularis*). There was one experimental group of normal animals. The array consisted of a 10 × 10 grid of silicon microelectrodes (1 mm in length) spaced 400 µm apart. We inserted the array partially into cortex, resulting in recordings confined mostly to layers 2–3. In two cases, we recorded simultaneously with a group of seven linearly arranged (2 mm extent) platinum-tungsten microelectrodes or tetrodes (Thomas Recording), positioned so that the nearest electrode was ~5 mm anterior to one edge of the multielectrode array. In this configuration, the distances between these electrodes and the array ranged from ~5 to 10 mm. Neuronal receptive fields in all data sets were within 5° of the fovea. Waveform segments that exceeded a user-defined threshold were sorted offline (Plexon Offline Sorter). We quantified sort quality using the signal-to-noise ratio (SNR) of each candidate unit[49], keeping units with an SNR of at least 2.3 (for the data in **Fig. 5a**) or at least 2.75 (for **Figs. 5b** and **7b**). In both data sets, we eliminated neurons for which the best grating stimulus did not evoke a response of at least 2 spikes/s. Changing the SNR or responsivity threshold did not qualitatively change any of the results described herein.

Visual stimuli were generated with EXPO and displayed on a linearized CRT monitor (mean luminance 40 $cd/m^2$) with a resolution of 1,024 by 768 pixels and a refresh rate of 100 Hz. Stimuli were presented in a circular aperture surrounded by a gray field of average luminance. We mapped the spatial receptive field (RF) of units by presenting small (0.6°) drifting gratings at a range of spatial positions. We then centered our stimuli on the aggregate RF of the recorded units.

Stimuli were viewed binocularly and presented for 1.28 s, separated by 1.5 s intervals of isoluminant gray screen (except in one penetration, for which the interval was 10 s). We presented full-contrast drifting sinusoidal gratings at 8 or 12 orientations spaced equally (22.5° or 30° increments). The spatial frequency (1–1.3 cpd) and temporal frequency (3 or 6.25 Hz) values were chosen to correspond to the typical preference of parafoveal V1 neurons[50]. The position and size (3.9–10°) of the grating were sufficient to cover the receptive fields of all the neurons. Stimulus orientation was block randomized, each stimulus orientation was repeated 100–200 times per recording session and each of the eight recording sessions yielded 21–68 units (210–2,278 pairs) that met our SNR and responsivity thresholds. This yielded a total of 318 units and 6,907 simultaneously recorded pairs.

**Statistical methods.** To compute spike count correlations for **Figures 1, 3, 4** and **6** we first randomly sampled a subset of the excitatory neurons in the recurrent network. The number of neurons sampled is indicated in each figure caption. Neurons with firing rates less than 1 Hz were excluded from the correlation analysis. Pearson correlation coefficients were computed by counting spikes over 250 ms windows and then computing the Pearson correlation coefficients between all pairs. The first 2 s of each simulation was excluded from the correlation analysis.

To obtain the residual correlations in **Figure 7a**, we randomly sampled 500 neurons with firing rates above 2 Hz from the L2/3 layer simulations. After omitting the first 2 s of the simulation, spike counts were computed using 50-ms bins, then partitioned into a sequence of 250-ms windows. To each window, we directly applied the GPFA algorithm used in a previous study[31] and downloaded from http://toliaslab.org/publications/ecker-et-al-2014/. This is a modified version of the algorithm introduced in ref. 41. Briefly, GPFA extracts shared fluctuations from a population of spike trains by fitting spike count covariance matrices to a model of the form

$$\mathrm{cov}(\mathbf{y}) = \mathbf{c}\mathbf{c}^{\mathbf{T}} + Q$$

Here, $\mathbf{c}$ is a column vector of coefficients quantifying the contribution of latent variability to each neuron's spiking activity. Latent variability is modeled by a one-dimensional Gaussian stochastic process. Hence, $\mathbf{c}\mathbf{c}^{\mathbf{T}}$ is a rank-one matrix of latent covariances and $Q$ quantifies the residual covariability not captured by the latent variable. Latent covariance in **Figure 5c** was computed by averaging the entries of $\mathbf{c}\mathbf{c}^{\mathbf{T}}$ computed using GPFA. Residual correlations were computed by using the "window" option (so that spike count covariances were computed over 250-ms windows), then converting the resulting covariance matrix, $Q$, to a correlation matrix using the corrcov command in Matlab (MathWorks).

Spike count correlations in **Figure 5b** were computed across trials, counting spikes over 250-ms time bins and subtracting the trial-averaged rate of each unit over each stimulus orientation. Hence, all reported spike count correlations represent correlated trial-to-trial variability (noise correlations) with stimulus correlations removed. Latent covariance in **Figure 5c** and residual correlation in **Figure 7b** were computed from the mean-subtracted spike counts using 50-ms time bins and 250-ms window sizes (treating each 250-ms window as a separate trial), then using the same GPFA algorithms and procedures that were used for the simulated data in **Figure 7a** (see description above). The three $P$-values reported in **Figure 7b** were computed using one-sided, unpaired $t$-tests with $t = 8.1$, 2.1 and 2.5 and degrees of freedom 3,937, 2,404 and 2,508, respectively.

The four error bars in **Figure 3e** were computed from $n = 880{,}057$, 2,646,645, 4,420,219 and 4,550,579 pairs respectively. The five error bars in **Figure 5b** were computed from $n = 1{,}688$, 2,251, 1,948, 458 and 562 pairs, respectively, and the same for **Figures 5c** and **7b**. The six error bars in **Figure 7a** were computed from $n = 3{,}885$, 11,696, 19,493, 27,521, 35,336 and 26,819 pairs, respectively. No statistical methods were used to predetermine sample sizes, but our sample sizes are the same as those reported in previous publications[33,36]. Where $t$-tests were used, data distribution was assumed to be normal but this was not formally tested. Data collection and analysis were performed blind to the conditions of the experiments. No animals were excluded from the analysis.

**Code availability.** Computer code for all simulations and analysis of the resulting data is included in **Supplementary Software**.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request.

A **Supplementary Methods Checklist** is available.


49. Kelly, R.C. *et al.* Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *J. Neurosci.* **27**, 261–264 (2007).
50. De Valois, R.L., Albrecht, D.G. & Thorell, L.G. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559 (1982).





doi:10.1038/nn.4433

# Supplementary Mathematical Note
# for The spatial structure of correlated neuronal variability

Robert Rosenbaum, Matthew A. Smith, Adam Kohn,
Jonathan E. Rubin and Brent Doiron

September 29, 2016

## S.1 Mathematical analysis of firing rates and correlations in recurrent networks

In the Results, we provided an intuitive justification for why some network architectures can realize an asynchronous state and why other architectures cannot. We now provide a more mathematically rigorous derivation of these results. We first review the computation of firing rates in the network, using previously developed methods [12, 6]. We then describe a cross-spectral measure of covariability and associated notation that greatly simplifies calculations, then derive general formulae for the full matrix of pairwise cross-spectra between the inputs and spike trains of all neurons. Finally, we use these results to derive mean field cross-spectra and conditions on the existence of an asynchronous state and an approximation to the spike train correlations at large system size.

### S.1.1 Mean field theory of firing rates in balanced networks

We first review the mean field theory of firing rates in balanced networks and compare the theoretical predictions to the results from our simulations. We represent spike trains as sums of Dirac delta functions,

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

where $t_{j,n}^a$ is the $n$-th spike time of neuron $j$ in population $a \in \{\mathrm{e}, \mathrm{i}, F\}$. For convenience, we also normalize units to set $C_m = 1$ for all calculations. In this case, the synaptic input to neuron $j$ in population $a = \mathrm{e}, \mathrm{i}$ (see Eqs. (4) and (6) in Methods) can be re-written as

$$\begin{aligned} I_j^a(t) &= F_j^a(t) + R_j^a(t) \\ &= F_j^a(t) + \frac{1}{\sqrt{N}} \sum_{b=\mathrm{e},\mathrm{i}} \sum_{k=1}^{N_b} J_{jk}^{ab} (\eta_b * S_k^b)(t) \end{aligned} \tag{S.1}$$

where $F_j^a(t)$ is the feedforward input to neuron $j$ in population $a$, $R_j^a(t)$ is the recurrent input, $*$ denotes convolution and

$$\eta_b(t) = \frac{1}{\tau_b} e^{-t/\tau_b} \Theta(t), \quad b = \mathrm{e}, \mathrm{i}$$

is the postsynaptic current waveform with $\Theta(t)$ representing the Heaviside step function. The variable $J_{jk}^{ab}$ denotes the synaptic weight from neuron $k$ in population $b = \mathrm{e}, \mathrm{i}$ to a neuron $j$ in population $a = \mathrm{e}, \mathrm{i}$.

Taking averages over time and over indices ($j$ and $k$) in Eq. (S.1) gives the mean-field mapping of firing rates to synaptic currents,

$$\begin{aligned} \overline{I}_a &= \overline{F}_a + \overline{R}_a \\ &= \overline{F}_a + \sqrt{N}(w_{a\mathrm{e}}^0 r_\mathrm{e} + w_{a\mathrm{i}}^0 r_\mathrm{i}). \end{aligned} \tag{S.2}$$

In this expression, $\overline{I}_a$ is the average synaptic current across the membrane of excitatory ($a = \mathrm{e}$) or inhibitory ($a = \mathrm{i}$) neurons in the recurrent network. Similarly, $\overline{F}_a$ is the average input from the feedforward population, $\overline{R}_a$ is the average input current from the local recurrent population and $r_a$ is the average firing rate of neurons in population $a = \mathrm{e}, \mathrm{i}$. The terms

$$w_{ab}^0 = p_{ab}^0 j_{ab} q_b \propto \mathcal{O}(1)$$

represents a normalized mean-field connectivity strength where $p_{ab}^0$ is the average number of connections from presynaptic neurons in population $b = \mathrm{e}, \mathrm{i}$ to postsynaptic neurons in population $a = \mathrm{e}, \mathrm{i}$, $j_{ab}$ is the average synaptic strength of each connection and $q_b = N_b/N$ is the proportion of neurons in the network that are in population $b$. When a presynaptic neuron can only make one synaptic projection to a single postsynaptic neuron, $p_{ab}^0$ is the probability of connection. In the networks we consider, multiple connections are possible (though rare), so $p_{ab}^0$ is technically the expected number of connections. The $\sqrt{N}$ scaling in Eq. (S.2) comes from the fact that there each neuron receives $\mathcal{O}(N)$ inputs (since connection probability is $\mathcal{O}(1)$) and the synaptic weight of each input scales like $\mathcal{O}(1/\sqrt{N})$ [12].

For the networks considered in Figure 1, the mean feedforward inputs are explicitly scaled as $\mathcal{O}(\sqrt{N})$ since (see Experimental Procedures)

$$F_j^a(t) = \sqrt{N} m_a + \sigma_s s(t)$$

where $s(t)$ is unbiased so that

$$\overline{F}_a = \sqrt{N} m_a.$$

For the spatially extended networks, feedforward inputs come from a separate population of neurons with

$$F_j^a(t) = \sum_{k=1}^{N_F} \frac{J_{jk}^{aF}}{\sqrt{N}} (\eta_F * S_k^F)(t)$$

so that

$$\overline{F}_a = \sqrt{N} w_{aF}^0 r_F$$

where $r_F = 5$ Hz is the firing rates of neurons in the feedforward population and $w_{aF}^0$ is defined in the same way as $w_{ab}^0$ above. Therefore, for all networks we consider, the mean inputs can be written as

$$\begin{aligned} \begin{bmatrix} \overline{I}_\mathrm{e} \\ \overline{I}_\mathrm{i} \end{bmatrix} &= \sqrt{N} \begin{bmatrix} w_{\mathrm{ee}}^0 r_\mathrm{e} + w_{\mathrm{ei}}^0 r_\mathrm{i} + f_\mathrm{e} \\ w_{\mathrm{ie}}^0 r_\mathrm{e} + w_{\mathrm{ii}}^0 r_\mathrm{i} + f_\mathrm{i} \end{bmatrix} \\ &= \sqrt{N}(W_0 \vec{r} + \vec{f}) \end{aligned} \tag{S.3}$$

where $f_a = m_a$ is the scaled mean feedforward input current for the networks considered in Figure 1, $f_a = \sqrt{N} w^0_{aF} r_F$ for the spatially extended networks, $\vec{r} = [r_\mathrm{e}\ r_\mathrm{i}]^T$ and $\vec{f} = [f_\mathrm{e}\ f_\mathrm{i}]^T$. Note that $w^0_{ab}, f_a \propto \mathcal{O}(1)$. Hence, to avoid excessively large input current magnitudes ($\overline{I}_a \to \pm\infty$ as $N \to \infty$), the excitatory and inhibitory currents in Eq. (S.3) must balance so that

$$W_0 \vec{r} + \vec{f} \propto \mathcal{O}(1/\sqrt{N}). \tag{S.4}$$

Firing rates in the limit of large $N$ are given by taking $N \to \infty$ in Eq. (S.4) and solving for the rates to obtain [12, 6]

$$\begin{aligned} \lim_{N\to\infty} r_\mathrm{e} &= \frac{f_\mathrm{e} w^0_\mathrm{ii} - f_\mathrm{i} w^0_\mathrm{ei}}{w^0_\mathrm{ei} w^0_\mathrm{ie} - w^0_\mathrm{ee} w^0_\mathrm{ii}} \\ \lim_{N\to\infty} r_\mathrm{i} &= \frac{f_\mathrm{e} w^0_\mathrm{ei} - f_\mathrm{i} w^0_\mathrm{ee}}{w^0_\mathrm{ei} w^0_\mathrm{ie} - w^0_\mathrm{ee} w^0_\mathrm{ii}}. \end{aligned} \tag{S.5}$$

For these expressions to give positive firing rates and a stable balanced state, parameters must satisfy [12, 6]

$$\frac{f_\mathrm{e}}{f_\mathrm{i}} > \frac{w^0_\mathrm{ei}}{w^0_\mathrm{ii}} > \frac{w^0_\mathrm{ee}}{w^0_\mathrm{ie}}. \tag{S.6}$$

For the networks in Figure 1, Eq. (S.5) predicts firing rates $r_\mathrm{e} = 6.2$ Hz and $r_\mathrm{i} = 4.0$ Hz. The simulations in Figure 1a-d produced average firing rates $r_\mathrm{e} = 7.6$ Hz and $r_\mathrm{i} = 3.8$ Hz. The simulations in Figure 1e-h produced average firing rates $r_\mathrm{e} = 7.4$ Hz and $r_\mathrm{i} = 3.8$ Hz.

For the networks in Figures 3 and 4, Eq. (S.5) predicts firing rates of $r_\mathrm{e} = 3.6$ Hz and $r_\mathrm{i} = 5.7$ Hz. For the simulations in Figure 3, average firing rates were $r_\mathrm{e} = 3.9$ Hz and $r_\mathrm{i} = 6.2$ Hz. For the simulations in Figure 4, average rates were $r_\mathrm{e} = 4.0$ Hz and $r_\mathrm{i} = 6.1$ Hz.

For the simulations in Figure 6, Eq. (S.5) predicts firing rates of $r_\mathrm{e} = 3.6$ Hz, $r_\mathrm{i} = 5.7$ Hz for L4 neurons and $r_\mathrm{e} = 5.0$ Hz, $r_\mathrm{i} = 7.9$ Hz for L2/3 neurons (using the L4 excitatory neuron rates as the feedforward input rates to L2/3). The average rates of L4 neurons in simulations were $r_\mathrm{e} = 3.8$ Hz and $r_\mathrm{i} = 6.2$ Hz and the average rates for L2/3 neurons were $r_\mathrm{e} = 5.7$ Hz and $r_\mathrm{i} = 9.2$ Hz.

Eq. (S.4) implies that, in balanced networks, feedforward input currents ($f_\mathrm{e}$ and $f_\mathrm{i}$) are approximately balanced by recurrent sources of synaptic input ($w^0_\mathrm{ee} r_\mathrm{e} + w^0_\mathrm{ei} r_\mathrm{i}$ and $w^0_\mathrm{ie} r_\mathrm{e} + w^0_\mathrm{ii} r_\mathrm{i} + f_\mathrm{i}$). This provides an intuition for the fact that feedforward and recurrent input currents are negatively correlated ($C_{RF} < 0$, see Results): Fluctuations in the feedforward input currents must be cancelled by opposite-polarity fluctuations in the recurrent input currents. Therefore, recurrent and feedforward inputs are negatively correlated. This intuition falls short of a precise theory for correlated variability in balanced networks, however, since Eq. (S.4) represents mean synaptic currents, not fluctuations. We next provide a more rigorous mathematical analysis of correlated variability in balanced networks.

### S.1.2 Cross-spectral measures of covariability

We first define the measures of correlation used in our calculations and review some of their properties. For mathematical convenience, we measure covariance between random processes, $U(t)$ and $V(t)$, in terms of their cross-spectral density or "cross-spectrum." Specifically, we define the cross-spectral operator, $\langle \cdot, \cdot \rangle$, by

$$\langle U, V \rangle(f) = \lim_{T\to\infty} \frac{1}{T} U_T(f) V_T^*(f)$$

where $^*$ denotes the complex conjugate,

$$U_T(f) = \int_0^T \left[U(t) - \overline{U}\,\right] e^{-2\pi i f t} dt$$

is the finite-time Fourier transform of $U(t)$ and $\overline{U}$ is the steady-state mean of $U(t)$. $V_T(f)$ is defined similarly.

The cross-spectrum is a function of frequency, $f$, but for notational convenience, we omit this explicit dependence in many expressions below. The cross-spectrum can also be defined as the Fourier transform of the cross-covariance function (sometimes also called the cross-correlation) [13, 9],

$$\langle U, V\rangle(f) = \int_{-\infty}^{\infty} CCG(\tau) e^{-2\pi i f \tau} d\tau$$

where $CCG(\tau) = \mathrm{cov}(V(t), U(t+\tau))$. Thus, cross-covariance functions are given by taking an inverse Fourier transform of the cross-spectrum. For spike trains, $S_1(t)$ and $S_2(t)$, spike count covariances over finite windows are given by integrating the cross-spectrum against a frequency kernel [7],

$$\text{spike count covariance over window of size T} = T \int_{-\infty}^{\infty} \langle S_1, S_2\rangle(f) K_T(f) df \tag{S.7}$$

where

$$K_T(f) = \frac{\sin^2(Tf\pi)}{Tf^2\pi^2}. \tag{S.8}$$

Since $\lim_{T\to\infty} K_T(f) = \delta(f)$, it follows that whenever $T$ is much larger than the timescale of correlations between the spike trains, we have the approximation

$$\text{spike count covariance over large window of size T} \approx T\langle S_1, S_2\rangle(0). \tag{S.9}$$

Thus, the cross-spectral operator captures all common measures of spike train covariability. More details and derivations of these relationships can be found in [9, 13].

For vector processes, $U(t) = [U_1(t), \cdots, U_M(t)]$ and $V(t) = [V_1(t), \cdots, V_N(t)]$, we define the $M \times N$ cross-spectral matrix operator,

$$\langle U, V\rangle = \Big[\, \langle U_j, V_k\rangle \,\Big]_{j,k=1}^{M,N}$$

which is the analogue to a covariance matrix in frequency space [2]. This is a bilinear, Hermitian operator in the sense that $\langle U + X, V\rangle = \langle U, V\rangle + \langle X, V\rangle$ and

$$\langle AU, BV\rangle = A\langle U, V\rangle B^T$$

for real matrices $A$ and $B$. Moreover, suppose $A(t)$ and $B(t)$ are matrix functions and $(A * U)(t)$ denotes matrix convolution, *i.e.*, a matrix product where multiplication is replaced by convolution [11]. Then

$$\langle A * U, B * V\rangle = \widetilde{A}\langle U, V\rangle \widetilde{B}^*$$

where

$$\widetilde{A}(f) = \int_{-\infty}^{\infty} A(t) e^{-2\pi i f t} dt$$

is the Fourier transform of the matrix function $A(t)$ and $\widetilde{B}^*(f)$ denotes the conjugate transpose of the Fourier transform of $B(t)$. Finally note that $\langle U, V \rangle = \langle V, U \rangle^*$ where $^*$ denotes the conjugate transpose. These definitions and properties greatly simplify the calculations below.

### S.1.3 Mapping pairwise spike train correlations to synaptic input correlations

We now derive a pair of equations that will be used to analyze correlations in all of the network models considered in the main text. To simplify calculations, we define the vector processes

$$S(t) = [S_1^{\mathrm{e}}(t) \, , \ldots \, , S_{N_{\mathrm{e}}}^{\mathrm{e}}(t) \, , S_1^{\mathrm{i}}(t) \, , \ldots \, , S_{N_{\mathrm{i}}}^{\mathrm{i}}(t)]^T$$

which contains all of the spike trains in the network and where $^T$ denotes the transpose. The vector process, $I(t)$, representing all synaptic input currents, $F(t)$, representing all feedforward input currents and $R(t)$, representing all recurrent inputs, are defined similarly. For notational convenience, we will often omit the explicit dependence on $t$.

Using this notation,

$$R = \frac{1}{\sqrt{N}} JH * S$$

is the vector of recurrent inputs and Eq. (S.1) can be written in vector form as

$$I = F + \frac{1}{\sqrt{N}} JH * S \tag{S.10}$$

where

$$J = \begin{bmatrix} J_{\mathrm{ee}} & J_{\mathrm{ei}} \\ J_{\mathrm{ie}} & J_{\mathrm{ii}} \end{bmatrix}$$

is an $N \times N$ matrix of synaptic weights written in block form and composed of the $N_a \times N_b$ sub-matrices

$$J_{ab} = \left[ J_{j,k}^{ab} \right]_{j,k=1}^{N_a, N_b}$$

for $j = 1, \ldots, N_a$ and $k = 1, \ldots, N_b$. Similarly,

$$H(t) = \begin{bmatrix} H_{\mathrm{e}}(t) & 0 \\ 0 & H_{\mathrm{i}}(t) \end{bmatrix}$$

is composed of sub-matrices

$$H_a(t) = \eta_a(t) I_{N_a \times N_a}$$

where $I_{N_a \times N_a}$ is the $N_a \times N_a$ identity matrix. The matrix convolution symbol, $*$, in Eq. (S.10) indicates that elements of the matrix $H(t)$ are convolved with elements of the vector $S(t)$ instead of multiplied [11].

Using Eq. (S.10), the cross-spectral matrix between total synaptic input and feedforward synaptic input can easily be calculated using the properties of the cross-spectral operator outlined above. Specifically,

$$\begin{aligned}\langle I, F\rangle &= \left\langle F + \frac{1}{\sqrt{N}} JH * S,\ F \right\rangle \\ &= \langle F, F\rangle + \frac{1}{\sqrt{N}} J\widetilde{H}\langle S, F\rangle\end{aligned} \tag{S.11}$$

where $\widetilde{H}(f)$ is the Fourier transform of the matrix $H(t)$. Correlations between total input currents can be calculated similarly,

$$\begin{aligned}\langle I, I\rangle &= \langle F + R, F + R\rangle \\ &= \left\langle F + \frac{1}{\sqrt{N}} JH * S,\ F + \frac{1}{\sqrt{N}} JH * S \right\rangle \\ &= \langle F, F\rangle + \frac{1}{\sqrt{N}} J\widetilde{H}\langle S,\ F\rangle + \frac{1}{\sqrt{N}}\langle F, S\rangle \widetilde{H}^* J^T + \frac{1}{N} J\widetilde{H}\langle S, S\rangle \widetilde{H}^* J^T\end{aligned} \tag{S.12}$$

where $\widetilde{H}^*$ denotes the conjugate-transpose of $\widetilde{H}$ and $J^T$ the transpose of $J$. Eqs. (S.11) and (S.12) are used below to derive a mean-field theory of correlations for various network topologies.

The entire matrix of spike train cross-spectra can be approximated by combining Eqs. (S.11) and (S.12) with the assumption of linear relations between $\langle I, F\rangle$ and $\langle S, F\rangle$, and between $\langle I, I\rangle$ and $\langle S, S\rangle$ [11]. Below, we show that a mean-field theory for correlations in the limit of large $N$ can be obtained from Eqs. (S.11) and (S.12) without appealing to this linear response approximation. Instead, we only need to assume that neuronal transfer is $\mathcal{O}(1)$ in the sense that the statistics of the total synaptic currents, $I_j^a(t)$, scale similarly with $N$ to the statistics of the spike trains, $S_j^a(t)$. Mathematically speaking, we assume that

$$0 < \lim_{N\to\infty} \frac{\mathrm{avg}_{j,k}\langle S_j^a, F_k^b\rangle}{\mathrm{avg}_{j,k}\langle I_j^a, F_k^b\rangle} < \infty \quad \text{and} \quad 0 < \lim_{N\to\infty} \frac{\mathrm{avg}_{j,k}\langle S_j^a, S_k^b\rangle}{\mathrm{avg}_{j,k}\langle I_j^a, I_k^b\rangle} < \infty \tag{S.13}$$

for all combinations of $a, b = \mathrm{e}, \mathrm{i}$. Here, $\mathrm{avg}_{j,k}$ is the average over any subset of the indices $j = 1, \dots N_a$ and $k = 1, \dots N_b$ which contains a $\mathcal{O}(1)$ subset of all such indices. We additionally assume that the activity of individual neurons has $\mathcal{O}(1)$ statistics, which is a characteristic feature of balanced networks [12]. Specifically, we assume that

$$\lim_{N\to\infty} \mathrm{avg}_k\langle X_k^a, Y_k^a\rangle < \infty \tag{S.14}$$

for $X, Y = S, I, F$ and $a = \mathrm{e}, \mathrm{i}$.

### S.1.4 Correlations in the homogeneous network (Figure 1a-d)

We first consider correlations in a network where the statistics of connection probability and feedforward input only depend on neuron type (excitatory or inhibitory) and are otherwise uniform across the network, as in Figure 1a-d. These calculations represent a generalization of the results in [5] to networks with any neuron model for which synaptic integration is linear (whereas only binary

neurons models were considered in [5]). In addition, our calculations are more easily generalized to structured networks (as we show below) and are arguably more concise.

Define $p_{ab}$ to be the expected number of synaptic inputs received by a neuron in postsynaptic population $a = \mathrm{e}, \mathrm{i}$ from a neuron in presynaptic population $b = \mathrm{e}, \mathrm{i}$ and let $j_{ab}/\sqrt{N}$ be the strength of each such connection. Under these assumptions, the average connection strengths are

$$E\left[J_{jk}^{ab}/\sqrt{N}\right] = p_{ab}j_{ab}/\sqrt{N}.$$

Also assume that there are $N_b = q_b N$ neurons in population $b = \mathrm{e}, \mathrm{i}$ where $q_\mathrm{e} + q_\mathrm{i} = 1$ and each of $q_b$, $p_{ab}$ and $j_{ab}$ is $\mathcal{O}(1)$ as $N \to \infty$.

Since neurons are statistically identical within each population, we define the population-averaged cross-spectral matrices,

$$\langle \mathbf{I}, \mathbf{F} \rangle = \begin{bmatrix} \langle I_\mathrm{e}, F_\mathrm{e} \rangle & \langle I_\mathrm{e}, F_\mathrm{i} \rangle \\ \langle I_\mathrm{i}, F_\mathrm{e} \rangle & \langle I_\mathrm{i}, F_\mathrm{i} \rangle \end{bmatrix}$$

where

$$\langle I_a, F_b \rangle = \mathrm{avg}_{j,k}(\langle I_j^a, F_k^b \rangle)$$

is the average pairwise cross-spectrum between total input current to neurons in population $a = \mathrm{e}, \mathrm{i}$ and the feedforward input to neurons in population $b = \mathrm{e}, \mathrm{i}$. Implicitly, the $j = k$ term is left out of the average when $a = b$. All other $2 \times 2$ mean-field cross-spectral matrices $\langle \mathbf{F}, \mathbf{F} \rangle$, $\langle \mathbf{S}, \mathbf{F} \rangle$, $\langle \mathbf{S}, \mathbf{S} \rangle$, $\langle \mathbf{I}, \mathbf{I} \rangle$, etc. are defined analogously. Note that all of these expressions depend on frequency, $f$, but the dependence is omitted to simplify notation.

The asynchronous state is defined by the scaling laws

$$\langle \mathbf{I}, \mathbf{I} \rangle, \langle \mathbf{S}, \mathbf{S} \rangle \propto \mathcal{O}(1/N) \text{ and } \langle \mathbf{I}, \mathbf{F} \rangle, \langle \mathbf{S}, \mathbf{F} \rangle \propto \mathcal{O}(1/\sqrt{N})$$

whenever feedforward inputs are moderately correlated,

$$\langle \mathbf{F}, \mathbf{F} \rangle \propto \mathcal{O}(1).$$

For the specific model considered in Figure 1a-d, we have

$$\langle \mathbf{F}, \mathbf{F} \rangle = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \sigma_s^2 \widetilde{A}_s$$

where

$$\widetilde{A}_s(f) = \langle s, s \rangle(f) = \tau_s \sqrt{2\pi} e^{-2f^2\pi^2\tau_s^2}$$

is the power spectral density of $s(t)$, obtained from the Fourier transform of the auto-covariance, $A_s(\tau)$ (see Methods). To prove the self-consistency of the asynchronous state, we must show that these scaling laws are consistent with the relationship between synaptic inputs and spike trains given by Eq. (S.1) or, equivalently, Eq. (S.10).

We begin by computing the mean-field cross-spectra between total inputs and feedforward inputs to excitatory neurons, substituting Eq. (S.11) to obtain

$$\begin{aligned}
\langle I_{\mathrm{e}}, F_{\mathrm{e}}\rangle &= \mathrm{avg}_{j,k}\langle I_j^{\mathrm{e}}, F_k^{\mathrm{e}}\rangle \\
&= \mathrm{avg}_{j,k}\langle F_j^{\mathrm{e}} + R_j^{\mathrm{e}}, F_k^{\mathrm{e}}\rangle \\
&= \mathrm{avg}_{j,k}\left[\langle F_j^{\mathrm{e}}, F_k^{\mathrm{e}}\rangle + \frac{1}{\sqrt{N}}\sum_{b=\mathrm{e,i}}\widetilde{\eta}_b\sum_{l=1}^{N_b} J_{jl}^{\mathrm{e}b}\langle S_l^b, F_k^{\mathrm{e}}\rangle\right] \\
&= \langle F_{\mathrm{e}}, F_{\mathrm{e}}\rangle + \mathrm{avg}_{j,k}\frac{1}{\sqrt{N}}\sum_{b=\mathrm{e,i}}\widetilde{\eta}_b\sum_{l=1}^{N_b} J_{jl}^{\mathrm{e}b}\langle S_l^b, F_k^{\mathrm{e}}\rangle
\end{aligned}$$

For large $N_{\mathrm{e}}$ and $N_{\mathrm{i}}$, the inner sums can be replaced by their averages:

$$\begin{aligned}
\mathrm{avg}_{j,k}\sum_{l=1}^{N_{\mathrm{e}}} J_{jl}^{\mathrm{ee}}\langle S_l^{\mathrm{e}}, F_k^{\mathrm{e}}\rangle &\to (N_{\mathrm{e}}-1)\mathrm{avg}_{j,k,l\neq k}\left[J_{jl}^{\mathrm{ee}}\langle S_l^{\mathrm{e}}, F_k^{\mathrm{e}}\rangle\right] + \mathrm{avg}_k\left[J_{kk}^{\mathrm{ee}}\langle S_k^{\mathrm{e}}, F_k^{\mathrm{e}}\rangle\right] \\
&= Nq_{\mathrm{e}}p_{\mathrm{ee}}j_{\mathrm{ee}}\langle S_{\mathrm{e}}, F_{\mathrm{e}}\rangle + \mathcal{O}(1)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{avg}_{j,k}\sum_{l=1}^{N_{\mathrm{i}}} J_{jl}^{\mathrm{ei}}\langle S_l^{\mathrm{i}}, F_k^{\mathrm{e}}\rangle &\to N_{\mathrm{i}}\mathrm{avg}_{j,k,l} J_{jl}^{\mathrm{ei}}\langle S_l^{\mathrm{i}}, F_k^{\mathrm{e}}\rangle \\
&= Nq_{\mathrm{i}}p_{\mathrm{ei}}j_{\mathrm{ei}}\langle S_{\mathrm{i}}, F_{\mathrm{e}}\rangle
\end{aligned}$$

as $N\to\infty$.

Putting this together gives

$$\begin{aligned}
\langle I_{\mathrm{e}}, F_{\mathrm{e}}\rangle =& \langle F_{\mathrm{e}}, F_{\mathrm{e}}\rangle + \sqrt{N}\left(\widetilde{\eta}_{\mathrm{e}}q_{\mathrm{e}}p_{\mathrm{ee}}j_{\mathrm{ee}}\langle S_{\mathrm{e}}, F_{\mathrm{e}}\rangle + \widetilde{\eta}_{\mathrm{i}}q_{\mathrm{i}}p_{\mathrm{ei}}j_{\mathrm{ei}}\langle S_{\mathrm{i}}, F_{\mathrm{e}}\rangle\right) \\
&+ \mathcal{O}(1/\sqrt{N}).
\end{aligned}$$

The same computation can be applied to all other pairings of $\langle I_a, F_b\rangle$ for $a, b = \mathrm{e}, \mathrm{i}$ to finally obtain

$$\begin{aligned}
\langle \mathbf{I}, \mathbf{F}\rangle =& \langle \mathbf{F}, \mathbf{F}\rangle + \sqrt{N}W\langle \mathbf{S}, \mathbf{F}\rangle \\
&+ \mathcal{O}(1/\sqrt{N}).
\end{aligned} \tag{S.15}$$

where

$$W = \begin{bmatrix} w_{\mathrm{ee}} & w_{\mathrm{ei}} \\ w_{\mathrm{ie}} & w_{\mathrm{ii}} \end{bmatrix} \tag{S.16}$$

and

$$w_{ab}(f) = \widetilde{\eta}_b(f)q_b p_{ab} j_{ab} \propto \mathcal{O}(1).$$

Note that $w_{ab}(0) = w_{ab}^0$ and $W(0) = W_0$ from the mean-field firing rate calculations surrounding Eq. (S.2). Under our assumption that neuronal transfer is $\mathcal{O}(1)$ (see Eq. S.13), we have that

$$\langle \mathbf{I}, \mathbf{F}\rangle \propto \langle \mathbf{S}, \mathbf{F}\rangle.$$

as $N$ grows large. Thus, Eq. (S.15) implies that

$$\langle \mathbf{S}, \mathbf{F} \rangle \propto \langle \mathbf{F}, \mathbf{F} \rangle + \sqrt{N} W \langle \mathbf{S}, \mathbf{F} \rangle. \tag{S.17}$$

This at first appears inconsistent since $\langle \mathbf{F}, \mathbf{F} \rangle \propto \mathcal{O}(1)$ and $\langle \mathbf{S}, \mathbf{F} \rangle$ appears on both sides of Eq. (S.17) scaled differently with $N$. This apparent inconsistency is resolved in the asynchronous state [5], wherein

$$\langle \mathbf{S}, \mathbf{F} \rangle \propto \mathcal{O}(1/\sqrt{N})$$

and the two terms on the right hand side of Eq. (S.17) cancel in such a way that

$$\langle \mathbf{I}, \mathbf{F} \rangle = \langle \mathbf{F}, \mathbf{F} \rangle + \sqrt{N} W \langle \mathbf{S}, \mathbf{F} \rangle \propto \mathcal{O}(1/\sqrt{N}).$$

This cancellation requires that

$$\lim_{N \to \infty} \sqrt{N} W \langle \mathbf{S}, \mathbf{F} \rangle = -\langle \mathbf{F}, \mathbf{F} \rangle \tag{S.18}$$

and therefore that

$$\lim_{N \to \infty} \sqrt{N} \langle \mathbf{S}, \mathbf{F} \rangle = -W^{-1} \langle \mathbf{F}, \mathbf{F} \rangle$$

which gives the asymptotic scaling of correlations between feedforward inputs and spike trains in the asynchronous state, generalizing the derivation in [5].

Note that the matrix of average cross-spectra between excitatory and inhibitory neurons' recurrent and feedforward inputs is given by

$$\langle \mathbf{R}, \mathbf{F} \rangle = \sqrt{N} W \langle \mathbf{S}, \mathbf{F} \rangle = -\langle \mathbf{F}, \mathbf{F} \rangle + \mathcal{O}(1/\sqrt{N}).$$

In the models we consider, correlations between neurons' feedforward inputs are non-negative, $\langle \mathbf{F}, \mathbf{F} \rangle \geq 0$. This provides a mathematical explanation for our observation in the Results that feeedforward and recurrent inputs are negatively correlated.

We can use similar methods to compute the average pairwise cross-spectra between spike trains in the recurrent network. Taking averages over the excitatory and the inhibitory populations in Eq. (S.12) gives

$$\begin{aligned} \langle \mathbf{I}, \mathbf{I} \rangle =& \langle \mathbf{F}, \mathbf{F} \rangle + \sqrt{N} \left( W \langle \mathbf{S}, \mathbf{F} \rangle + \langle \mathbf{F}, \mathbf{S} \rangle W^* \right) + N W \langle \mathbf{S}, \mathbf{S} \rangle W^* + W A W^* \\ &+ \mathcal{O}(1/\sqrt{N}) \end{aligned} \tag{S.19}$$

where $W^*$ is the conjugate transpose of $W$. As before, the $\mathcal{O}(1/\sqrt{N})$ term captures the diagonal elements omitted from the averages that define $\langle \mathbf{S}, \mathbf{F} \rangle$. Finally, the term $A$ is defined by

$$A(f) = \begin{bmatrix} A_{\mathrm{e}}(f)/q_{\mathrm{e}} & 0 \\ 0 & A_{\mathrm{i}}(f)/q_{\mathrm{i}} \end{bmatrix}$$

where

$$A_a(f) = \mathrm{avg}_k \langle S_k^a, S_k^a \rangle$$

represents the average power spectral density of spike trains in population $a = \mathrm{e}, \mathrm{i}$. When spike trains in the network are approximately Poisson processes, the power spectral density is approximately equal to the firing rate $A_a(f) \approx r_a$.

Combining this with Eq. (S.18) and the fact that $\langle \mathbf{F}, \mathbf{F} \rangle = \langle \mathbf{F}, \mathbf{F} \rangle^*$ allows us to make the substitutions

$$\sqrt{N} W \langle \mathbf{S}, \mathbf{F} \rangle = \sqrt{N} \langle \mathbf{F}, \mathbf{S} \rangle W^* = -\langle \mathbf{F}, \mathbf{F} \rangle + \mathcal{O}(1/\sqrt{N})$$

Putting this together gives

$$\langle \mathbf{I}, \mathbf{I} \rangle \propto -\langle \mathbf{F}, \mathbf{F} \rangle + NW \langle \mathbf{S}, \mathbf{S} \rangle W^* + WAW^*$$

for large $N$. We again invoke our assumption that neuronal transfer is $\mathcal{O}(1)$ so that $\langle \mathbf{I}, \mathbf{I} \rangle \propto \langle \mathbf{S}, \mathbf{S} \rangle$ for large $N$ and therefore

$$\langle \mathbf{S}, \mathbf{S} \rangle \propto -\langle \mathbf{F}, \mathbf{F} \rangle + NW \langle \mathbf{S}, \mathbf{S} \rangle W^* + WAW^*. \tag{S.20}$$

This again presents an apparent inconsistency that is resolved in the asynchronous state where

$$\langle \mathbf{S}, \mathbf{S} \rangle \propto \mathcal{O}(1/N)$$

and the terms on the right hand side of Eq. (S.20) cancel so that

$$\lim_{N \to \infty} NW \langle \mathbf{S}, \mathbf{S} \rangle W^* = \langle \mathbf{F}, \mathbf{F} \rangle - WAW^*. \tag{S.21}$$

and

$$\lim_{N \to \infty} N \langle \mathbf{S}, \mathbf{S} \rangle = W^{-1} \langle \mathbf{F}, \mathbf{F} \rangle W^{-*} - A, \tag{S.22}$$

which gives the asymptotic scaling of correlations between spike trains in the asynchronous state, generalizing and simplifying the derivation in [5]. Here, $W^{-*}$ is the matrix inverse of $W^*$. The existence of this asynchronous state requires $W$ to be invertible or, if $W$ is not invertible, the matrix $\langle \mathbf{F}, \mathbf{F} \rangle$ must be in the range of the matrix operator $U \mapsto WUW^*$ so that Eq. (S.21) has a solution. For the parameters used in Figure 1a-d, $W$ is invertible.

This approach to deriving of the existence of the asynchronous state can be applied to any network model for which synaptic integration is linear, synaptic connection strengths scale like $\mathcal{O}(1/\sqrt{N})$, synaptic connection probability is $\mathcal{O}(1)$, and the inequalities in (S.13) are satisfied. Thus, these calculations represent a generalization of the results in [5], which were specific to binary neuron models. Below, we show that this approach facilitates the derivation of correlations in discretely heterogeneous and spatially extended networks.

### S.1.5 General treatment of correlations in discretely heterogeneous networks

Above, we considered a network with two populations, one excitatory and one inhibitory. Connectivity statistics and feedforward inputs were homogeneous within each population. We now generalize this approach to networks with an arbitrary number of populations. We then apply this generalized analysis to show that the network in Figure 1e-h cannot realize an asynchronous state.

Consider a network of $N$ neurons subdivided into $M$ sub-populations where the $m$th subpopulation contains $N_m = q_m N$ neurons for $m = 1, \ldots, M$. Generalizing Eq. (S.1), the input to neuron $j = 1, \ldots, N_m$ in population $m = 1, \ldots, M$ is given by

$$I_j^m(t) = F_j^m(t) + \sum_{n=1}^{M} \sum_{k=1}^{N_n} \frac{J_{jk}^{mn}}{\sqrt{N}} (\eta_n * S_k^n)(t) \tag{S.23}$$

where $S_k^n(t)$ is the $k$th neuron in population $n = 1, \ldots, M$. Define $p_{mn}$ to be the expected number of synaptic inputs received by a neuron in postsynaptic population $m = 1, \ldots, M$ from a neuron in presynaptic population $n = 1, \ldots, M$ and let $J_{jk}^{mn} = j_{mn}$ be the strength of each such connection. This assures that the expected value of $J_{jk}^{mn}$ is $p_{mn}j_{mn}$. Also assume that each of $q_m$, $p_{mn}$ and $j_{mn}$ is $\mathcal{O}(1)$ as $N \to \infty$. Now let

$$\langle F_m, F_n \rangle = \text{avg}_{j,k} \langle F_j^m, F_k^n \rangle \propto \mathcal{O}(1)$$

be the average cross-spectrum between feedforward inputs to neurons in population $m = 1, \ldots, M$ and $n = 1, \ldots, M$. We again implicitly omit terms for which $m = n$ and $j = k$ from the average. Define $\langle I_m, I_n \rangle$, $\langle I_m, F_n \rangle$, etc. analogously. Finally, define the $M \times M$ mean-field correlation matrices,

$$\langle \mathbf{F}, \mathbf{F} \rangle = \big[ \langle F_m, F_n \rangle \big]_{m,n}$$

and similarly for $\langle \mathbf{I}, \mathbf{I} \rangle$, $\langle \mathbf{I}, \mathbf{F} \rangle$, etc.

All of this notational machinery is useful because it allows us to take averages over each subpopulation in Eqs. (S.11) and (S.12) to obtain

$$\begin{aligned} \langle \mathbf{I}, \mathbf{F} \rangle =& \langle \mathbf{F}, \mathbf{F} \rangle + \sqrt{N} W \langle \mathbf{S}, \mathbf{F} \rangle \\ &+ \mathcal{O}(1/\sqrt{N}) \end{aligned} \tag{S.24}$$

and

$$\begin{aligned} \langle \mathbf{I}, \mathbf{I} \rangle =& \langle \mathbf{F}, \mathbf{F} \rangle + \sqrt{N} \left( W \langle \mathbf{S}, \mathbf{F} \rangle + \langle \mathbf{F}, \mathbf{S} \rangle W^* \right) + NW \langle \mathbf{S}, \mathbf{S} \rangle W^* + WAW^* \\ &+ \mathcal{O}(1/\sqrt{N}) \end{aligned}$$

which are identical to Eqs. (S.15) and (S.19) except that that the correlation matrices are $M \times M$ instead of $2 \times 2$. The $M \times M$ mean-field connectivity matrix is defined by

$$W = \big[ w_{mn} \big]_{m,n}$$

where

$$w_{mn} = \widetilde{\eta}_n q_n p_{mn} j_{mn} \propto \mathcal{O}(1)$$

is the mean-field connectivity from populations $n$ to $m$. The term $A(f)$ is a diagonal $M \times M$ matrix with diagonal elements $[A(f)]_{n,n} = A_n(f)/q_n$ where $A_n(f) = \text{avg}_k \langle S_k^n, S_k^n \rangle(f)$ is the average power spectral density of neurons in population $n$.

Thus, the analysis and derivation of Eqs. (S.15 – S.22) in Section S.1.4 is identical for heterogeneous networks, except with $M \times M$ matrices instead of $2 \times 2$. Specifically, the existence of the asynchronous state again requires that Eqs. (S.18) and (S.21) are solvable. These equations are necessarily solvable whenever $W$ is invertible and the solution is given by Eq. (S.22) in such cases. We next apply this generalized analysis to the network considered in Figure 1e-h.

### S.1.6 Correlations in the discretely heterogeneous network from Figure 1e-h

For the example considered in Figure 1e-h, there are two excitatory and two inhibitory populations, so the analysis from Section S.1.5 is applicable with $M = 4$. We can enumerate these populations as $\text{e}_1$, $\text{i}_1$, $\text{e}_2$ and $\text{i}_2$ where $\text{e}_1$ and $\text{e}_2$ are excitatory, $\text{i}_1$ and $\text{i}_2$ are inhibitory. All neurons in populations $\text{e}_1$

and $\mathrm{i}_1$ receive the same time-varying feedforward input, $s_1(t)$. Similarly, all neurons in populations $\mathrm{e}_2$ and $\mathrm{i}_2$ receive the time-varying input $s_2(t)$ which is uncorrelated with $s_1(t)$. Therefore, the cross-spectral matrix between feedforward inputs is given by

$$\langle \mathbf{F}, \mathbf{F} \rangle = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \sigma_s^2 \widetilde{A}_s$$

where $\widetilde{A}_s(f)$ is the cross-spectral density of $s_1(t)$ and also of $s_2(t)$.

Connectivity in the network only depends on neuron type (excitatory or inhibitory) so that

$$W = \frac{1}{2} \begin{bmatrix} w_{\mathrm{ee}} & w_{\mathrm{ei}} & w_{\mathrm{ee}} & w_{\mathrm{ei}} \\ w_{\mathrm{ie}} & w_{\mathrm{ii}} & w_{\mathrm{ie}} & w_{\mathrm{ii}} \\ w_{\mathrm{ee}} & w_{\mathrm{ei}} & w_{\mathrm{ee}} & w_{\mathrm{ei}} \\ w_{\mathrm{ie}} & w_{\mathrm{ii}} & w_{\mathrm{ie}} & w_{\mathrm{ii}} \end{bmatrix}$$

where each $w_{ab}$ is the same connection strength from the homogeneous network in Figure 1a-d, defined above. Note that this matrix is comprised of four blocks, each identical to the $W$ matrix for the $2 \times 2$ homogeneous network in Figure 1a-d. The multiplication by $1/2$ is necessary because there are half as many neurons in each population, e.g. $q_{\mathrm{e}_1} = q_{\mathrm{e}}/2$.

The symmetry of the network also implies that the average power spectral density is the same in population $\mathrm{e}_1$ as $\mathrm{e}_2$ and similarly for $\mathrm{i}_1$ and $\mathrm{i}_2$. Therefore,

$$A(f) = \frac{1}{2} \begin{bmatrix} A_{\mathrm{e}}(f)/q_{\mathrm{e}} & 0 & 0 & 0 \\ 0 & A_{\mathrm{i}}(f)/q_{\mathrm{i}} & 0 & 0 \\ 0 & 0 & A_{\mathrm{e}}(f)/q_{\mathrm{e}} & 0 \\ 0 & 0 & 0 & A_{\mathrm{i}}(f)/q_{\mathrm{i}} \end{bmatrix}$$

where $A_a(f)$ is the average cross-spectral density of excitatory ($a = \mathrm{e}$) or inhibitory ($a = \mathrm{i}$) spike trains and $q_a = 2q_{a_1} = 2q_{a_2} = 1/2$ is the proportion of neurons that are excitatory ($a = \mathrm{e}$) or inhibitory ($a = \mathrm{i}$).

Note that $W$ is a singular matrix, $\langle \mathbf{F}, \mathbf{F} \rangle$ is not in the range of $U \mapsto WU$ and therefore also not in the range of $U \mapsto WUW^*$. For example, note that for any $4 \times 4$ matrix $V$ (including any $V = UW^*$), the first row of the product $WV$ is the same as its third row, which is not true of $\langle \mathbf{F}, \mathbf{F} \rangle$. Therefore, the asynchronous state cannot be obtained in the heterogeneous two-population network considered in Figure 1a-d since Eqs. (S.18) and (S.21) do not admit solutions. In other words, it is not mathematically possible for all elements of the $4 \times 4$ matrix of spike train correlations,

$$\langle \mathbf{S}, \mathbf{S} \rangle = \begin{bmatrix} \langle S_{\mathrm{e}_1}, S_{\mathrm{e}_1} \rangle & \langle S_{\mathrm{e}_1}, S_{\mathrm{i}_1} \rangle & \langle S_{\mathrm{e}_1}, S_{\mathrm{e}_2} \rangle & \langle S_{\mathrm{e}_1}, S_{\mathrm{i}_2} \rangle \\ \langle S_{\mathrm{i}_1}, S_{\mathrm{e}_1} \rangle & \langle S_{\mathrm{i}_1}, S_{\mathrm{i}_1} \rangle & \langle S_{\mathrm{i}_1}, S_{\mathrm{e}_2} \rangle & \langle S_{\mathrm{i}_1}, S_{\mathrm{i}_2} \rangle \\ \langle S_{\mathrm{e}_2}, S_{\mathrm{e}_1} \rangle & \langle S_{\mathrm{e}_2}, S_{\mathrm{i}_1} \rangle & \langle S_{\mathrm{e}_2}, S_{\mathrm{e}_2} \rangle & \langle S_{\mathrm{e}_2}, S_{\mathrm{i}_2} \rangle \\ \langle S_{\mathrm{i}_2}, S_{\mathrm{e}_1} \rangle & \langle S_{\mathrm{i}_2}, S_{\mathrm{i}_1} \rangle & \langle S_{\mathrm{i}_2}, S_{\mathrm{e}_2} \rangle & \langle S_{\mathrm{i}_2}, S_{\mathrm{i}_2} \rangle \end{bmatrix}$$

to be $\mathcal{O}(1/N)$ because the assumption that all elements are $\mathcal{O}(1/N)$ leads to a mathematical contradiction. Thus, the lack of a solution to Eqs. (S.18) and (S.21) explains why spike trains are correlated in Figure 1e-h.

To understand why the average correlation between all pairs in Figure 1e-h is approximately zero, we can apply the mean-field correlation theory to pairwise correlations averaged across populations 1 and 2. Specifically, let $\langle F_{\mathrm{e}}, F_{\mathrm{e}} \rangle$ be the average correlation between all excitatory neurons, regardless of population membership and similarly for $\langle F_{\mathrm{i}}, F_{\mathrm{e}} \rangle$, etc. Now re-define

$$\langle \mathbf{F}, \mathbf{F} \rangle = \begin{bmatrix} \langle F_{\mathrm{e}}, F_{\mathrm{e}} \rangle & \langle F_{\mathrm{e}}, F_{\mathrm{i}} \rangle \\ \langle F_{\mathrm{i}}, F_{\mathrm{e}} \rangle & \langle F_{\mathrm{i}}, F_{\mathrm{i}} \rangle \end{bmatrix} \tag{S.25}$$

to be the matrix correlations averaged over populations 1 and 2. Since half of the neuron pairs receive perfectly correlated input and half receive perfectly uncorrelated input, we have

$$\langle \mathbf{F}, \mathbf{F} \rangle = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \sigma_s^2 \widetilde{A}_s. \tag{S.26}$$

The average mean-field connectivity between excitatory and inhibitory pairs is identical to the connectivity for the homogeneous network from Section S.1.4, so $W$ is the same $2 \times 2$ matrix given in Eq. (S.16). Defining the $2 \times 2$ population-averaged correlation matrices, $\langle \mathbf{I}, \mathbf{F} \rangle$, $\langle \mathbf{I}, \mathbf{I} \rangle$, etc. analogously to $\langle \mathbf{F}, \mathbf{F} \rangle$ in Eq. (S.25), we can apply the same mean-field correlation analysis that was applied to the homogeneous network in Section S.1.4. The only difference is that $\langle \mathbf{F}, \mathbf{F} \rangle$ is divided by two, *c.f.* Eq. (S.26). Since $W$ is invertible, this analysis yields a self-consistent asynchronous solution.

We conclude that correlations averaged across all pairs from both populations (1 and 2) are weak, but correlations averaged while respecting population membership are not. For example, the average correlation between all excitatory neurons in the network is $\mathcal{O}(1/N)$, but the average correlation between excitatory neurons in population 1 is $\mathcal{O}(1)$. However, note that since all populations have the same number of neurons, the average covariance between all excitatory pairs is formed by the average of the covariances between same- and opposite-population paris,

$$4\langle S_{\mathrm{e}}, S_{\mathrm{e}} \rangle = \underbrace{\langle S_{\mathrm{e}_1}, S_{\mathrm{e}_1} \rangle + \langle S_{\mathrm{e}_2}, S_{\mathrm{e}_2} \rangle}_{\text{same pop. pairs}} + \underbrace{\langle S_{\mathrm{e}_1}, S_{\mathrm{e}_2} \rangle + \langle S_{\mathrm{e}_2}, S_{\mathrm{e}_1} \rangle}_{\text{opposite pop. pairs}}. \tag{S.27}$$

Also note that the network is symmetric with respect to population membership in the sense that both populations are statistically identical to one another. Thus, $\langle S_{\mathrm{e}_1}, S_{\mathrm{e}_1} \rangle = \langle S_{\mathrm{e}_2}, S_{\mathrm{e}_2} \rangle$ and $\langle S_{\mathrm{e}_1}, S_{\mathrm{e}_2} \rangle = \langle S_{\mathrm{e}_2}, S_{\mathrm{e}_1} \rangle$. The only way to have $\langle S_{\mathrm{e}}, S_{\mathrm{e}} \rangle \propto \mathcal{O}(1/N)$ and $\langle S_{\mathrm{e}_m}, S_{\mathrm{e}_n} \rangle \propto \mathcal{O}(1)$ while respecting this symmetry is to have

$$\langle S_{\mathrm{e}_1}, S_{\mathrm{e}_1} \rangle \propto -\langle S_{\mathrm{e}_1}, S_{\mathrm{e}_2} \rangle + \mathcal{O}(1/N)$$

so that the contributions from same- and opposite-population pairs cancel in Eq. (S.27) up to order $1/N$. Indeed, this structure is observed in Figure 1f.

### S.1.7 Correlations in continuously indexed networks with distance-dependent connectivity

To compute the spatial shape of correlations and the conditions on the asynchronous state, we first need to compute the shape of correlations between neurons' feedforward inputs. Since spike trains in the feedforward layer are uncorrelated, correlations between the feedforward input to neurons

arise solely through overlapping presynaptic pools. Recall from Methods that the feedforward input to neuron $j = 1, \ldots, N_a$ in population $a = \mathrm{e}, \mathrm{i}$ is defined by

$$F_j^a(t) = \sum_{k=1}^{N_F} \frac{J_{jk}^{aF}}{\sqrt{N}} (\eta_F * S_k^F)(t)$$

where $S_k^F(t)$ is the spike train of neuron $k$ in the feedforward layer, $*$ denotes convolution, $\eta_F(t)$ is a postsynaptic current waveform (for which we set $\eta_F = \eta_\mathrm{e}$ in all simulations) and $J_{jk}^{aF}$ is a synaptic weight.

Our first goal is to write the cross-spectral density between feedforward inputs as a function of the distance between the postsynaptic neurons in the recurrent layer. A similar strategy will later be applied to compute correlations between spike trains in the recurrent network. First consider the cross-spectral density between excitatory neuron $j$ and excitatory neuron $k$,

$$\begin{aligned}\langle F_j^a, F_k^b \rangle &= \left\langle \sum_{l=1}^{N_F} \frac{J_{jl}^{aF}}{\sqrt{N}} \eta_F * S_l^F, \sum_{m=1}^{N_F} \frac{J_{km}^{bF}}{\sqrt{N}} \eta_F * S_m^F \right\rangle \\ &= \frac{|\widetilde{\eta}_F|^2}{N} \sum_{l,m=1}^{N_F} J_{jl}^{aF} J_{km}^{bF} \langle S_l^F, S_m^F \rangle.\end{aligned}$$

Since spike trains in the feedforward population are uncorrelated, the terms in this sum are only non-zero when $l = m$ (since otherwise $\langle S_l^F, S_m^F \rangle = 0$). Also, since each $S_m^F(t)$ is a Poisson process with rate $r_F$, the power-spectral density is $\langle S_m^F, S_m^F \rangle = r_F$. Putting these two facts together, we have

$$\langle F_j^a, F_k^b \rangle = q_F |\widetilde{\eta}_F|^2 r_F \frac{1}{N_F} \sum_{m=1}^{N_F} J_{jm}^{aF} J_{km}^{bF} \tag{S.28}$$

where $q_F = N_F/N$.

To obtain a spatially continuous description of correlations, now consider decomposing the square domain, $\Gamma$, into a uniform grid of $M$ small squares, each with side-length $\sqrt{\delta}$ where $\delta = 1/M$ is the area of each grid square. Now, let $\mathbf{x}$ and $\mathbf{y}$ be the two-dimensional coordinates of the center of two different grid squares and define the average cross-spectral density between feedforward input to excitatory neurons in those squares,

$$\begin{aligned}\{F_\mathrm{e}, F_\mathrm{e}\}(\mathbf{x}, \mathbf{y}, f) &:= \mathrm{avg}_{j \in \mathrm{e}(\mathbf{x}), k \in \mathrm{e}(\mathbf{y})} \langle F_j^\mathrm{e}, F_k^\mathrm{e} \rangle(f) \\ &= q_F |\widetilde{\eta}_F(f)|^2 r_F \frac{1}{N_F} \sum_{m=1}^{N_F} \mathrm{avg}_{j \in \mathrm{e}(\mathbf{x}), k \in \mathrm{e}(\mathbf{y})} J_{jm}^{\mathrm{e}F} J_{km}^{\mathrm{e}F}\end{aligned} \tag{S.29}$$

where $j \in \mathrm{e}(\mathbf{x})$ indicates that the excitatory neuron with index $j$ lies in the grid square centered at $\mathbf{x} \in \Gamma$, and similarly for $k \in \mathrm{e}(\mathbf{y})$. The second line follows from Eq. (S.28) and from permuting the average with the sum. The postsynaptic current waveform is given by $\eta_F(t) = e^{-t/\tau_F}/\tau_F$ for $t > 0$ which has Fourier transform,

$$\widetilde{\eta}_F(f) = \frac{1}{1 + 2\pi i f \tau_F}.$$

Recall that for all simulations we set $\tau_F = \tau_\mathrm{e}$.

Now note that the average in Eq. (S.29) is taken over $N_{\mathrm{e}}^2/M^2 \gg 1$ pairs of indices, $j$ and $k$. In the limit of large $N$, the average synaptic weights can be replaced by their expected values to get

$$\{F_{\mathrm{e}}, F_{\mathrm{e}}\}(\mathbf{x}, \mathbf{y}, f) = q_F |\widetilde{\eta}_F(f)|^2 r_F \frac{1}{N_F} \sum_{\mathbf{z}} j_{\mathrm{e}F}^2 p_{\mathrm{e}F}(\mathbf{x} - \mathbf{z}) p_{\mathrm{e}F}(\mathbf{y} - \mathbf{z}) \tag{S.30}$$

where the sum is over all neurons in the feedforward network with coordinates indicated by $\mathbf{z} \in \Gamma$. Recall from Methods that the expected number of connections from a neuron at coordinates $\mathbf{z}$ in the feedforward layer to an excitatory neuron at coordinates $\mathbf{x}$ in the recurrent network is given by

$$p_{\mathrm{e}F}(\mathbf{x} - \mathbf{z}) = \frac{K_{\mathrm{e}F}^{\mathrm{out}}}{N_{\mathrm{e}}} G(\mathbf{x} - \mathbf{z}; \alpha_F^2)$$

is the probability of connection from a neuron in the feedforward input layer at coordinates $\mathbf{z} = (z_1, z_2) \in \Gamma$ to an excitatory neuron in the recurrent layer at coordinates $\mathbf{x} = (x_1, x_2) \in \Gamma$ where

$$G(\mathbf{u}; \sigma^2) = g(u_1; \sigma^2) g(u_2; \sigma^2)$$

is a two-dimensional wrapped Gaussian function composed of one-dimensional wrapped Gaussians,

$$g(u; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} e^{-(u+k)^2/(2\sigma^2)}.$$

Now note that the sum in Eq. (S.30) contains $N_F$ equally-spaced elements, $\mathbf{z}$, so coupled with the $1/N_F$ coefficient, it represents a Riemmann sum on $\Gamma$. For large $N_F$, the sum can be written as an integral to obtain

$$\begin{aligned}
\{F_{\mathrm{e}}, F_{\mathrm{e}}\}(\mathbf{x}, \mathbf{y}, f) &= q_F |\widetilde{\eta}_F(f)|^2 r_F j_{\mathrm{e}F}^2 \iint_{\Gamma} p_{\mathrm{e}F}(\mathbf{x} - \mathbf{z}) p_{\mathrm{e}F}(\mathbf{y} - \mathbf{z}) d\mathbf{z} \\
&= q_F |\widetilde{\eta}_F(f)|^2 r_F j_{\mathrm{e}F}^2 \iint_{\Gamma} p_{\mathrm{e}F}(\mathbf{v}) p_{\mathrm{e}F}(\mathbf{y} - \mathbf{x} - \mathbf{v}) d\mathbf{v}
\end{aligned}$$

where the second line follows from the change of coordinates $\mathbf{v} = \mathbf{x} - \mathbf{z}$. This demonstrates that the $\{F_a, F_b\}(\mathbf{x}, \mathbf{y}, f)$ depends only on $\mathbf{u} = \mathbf{y} - \mathbf{x}$. With a slight abuse of notation, we can re-define $\{\cdot, \cdot\}$ to be distance-dependent by making the substitution $\{F_a, F_b\}(\mathbf{x} - \mathbf{y}, f) \leftarrow \{F_a, F_b\}(\mathbf{x}, \mathbf{y}, f)$. With this re-definition, we can re-write the integral as

$$\{F_{\mathrm{e}}, F_{\mathrm{e}}\}(\mathbf{u}, f) = q_F |\widetilde{\eta}_F(f)|^2 r_F j_{\mathrm{e}F}^2 \iint_{\Gamma} p_{\mathrm{e}F}(\mathbf{v}) p_{\mathrm{e}F}(\mathbf{u} - \mathbf{v}) d\mathbf{v} \tag{S.31}$$

where $\{F_{\mathrm{e}}, F_{\mathrm{e}}\}(\mathbf{u}, f)$ now denotes the average cross-spectral density between the feedforward input to pairs of excitatory neurons at coordinates $\mathbf{x} \in \Gamma$ and $\mathbf{y} \in \Gamma$ for which $\mathbf{x} - \mathbf{y} = \mathbf{u}$. Since $p_{\mathrm{e}F}(\mathbf{v})$ is a two-dimensional wrapped Gaussian as indicated above, the integral in Eq. (S.31) is a two-dimensional circular convolution [6]. The circular convolution of a wrapped Gaussian with another wrapped Gaussian simply sums the variances so that

$$\{F_{\mathrm{e}}, F_{\mathrm{e}}\}(\mathbf{u}, f) = q_F |\widetilde{\eta}_F(f)|^2 r_F j_{\mathrm{e}F}^2 \left( \frac{K_{\mathrm{e}F}^{\mathrm{out}}}{N_{\mathrm{e}}} \right)^2 G(\mathbf{u}; 2\alpha_F^2).$$

Thus, the cross-spectral density (and therefore covariance and correlation) between the feedforward inputs to two excitatory neurons only depends on their distance, $u$, measured periodically on $\Gamma$, and decays like a Gaussian with width parameter, $2\alpha_F^2$, justifying these claims made in the main text. We can apply the same computation to excitatory-inhibitory and inhibitory-inhibitory pairs to finally obtain

$$\begin{aligned}\{\mathbf{F},\mathbf{F}\}(\mathbf{u},f) &:= \begin{bmatrix} \{F_\mathrm{e},F_\mathrm{e}\}(\mathbf{u},f) & \{F_\mathrm{e},F_\mathrm{i}\}(\mathbf{u},f) \\ \{F_\mathrm{i},F_\mathrm{e}\}(\mathbf{u},f) & \{F_\mathrm{i},F_\mathrm{i}\}(\mathbf{u},f) \end{bmatrix} \\ &= \begin{bmatrix} |\overline{w}_{\mathrm{e}F}(f)|^2 & \overline{w}_{\mathrm{e}F}(f)\overline{w}^*_{\mathrm{i}F}(f) \\ \overline{w}_{\mathrm{i}F}(f)\overline{w}^*_{\mathrm{e}F}(f) & |\overline{w}_{\mathrm{i}F}(f)|^2 \end{bmatrix} r_F G(\mathbf{u};2\alpha_F^2)\end{aligned} \tag{S.32}$$

where $\{F_a,F_b\}(\mathbf{u},f)$ is the average cross-spectral density between the feedforward input to neurons in population $a$ and neurons in population $b$ separated by the vector $\mathbf{u}$, and where the mean-field feedforward connection strength is

$$\overline{w}_{aF}(f) = \sqrt{q_F}\overline{p}_{aF}j_{aF}\widetilde{\eta}_F(f)$$

with

$$\overline{p}_{aF} = \frac{K_{aF}^{\mathrm{out}}}{N_a}$$

representing the network-averaged outgoing connection probability. Note that this implies $\{\mathbf{F},\mathbf{F}\}(\mathbf{u},f) \propto \mathcal{O}(1)$ since all parameters in Eq. (S.32) are $\mathcal{O}(1)$.

Recall that the wrapped Gaussian, $G(\mathbf{u};2\alpha_F^2)$, represents a two-dimensional Gaussian with distance measured periodically on $\Gamma$. Therefore, Eq. (S.32) implies that correlations between feedforward inputs decay with distance like a Gaussian with a width parameter $2\alpha_F^2$ where $\alpha_F^2$ is the width parameter for the decay of feedforward connection probability. This result justifies the claim made in the results that correlations between feedforward inputs are twice as broad as the feedforward synaptic projections.

Now that we have computed the spatial profile of the cross-spectral densities between feedforward inputs to neurons, we use similar techniques to analyze the spatial structure of correlations between spike trains in the recurrent network. First consider the same discretization of the network into $M$ grid squares that we used above. For such a discretization, Eq. (S.23) and the remaining analysis in Section S.1.5 are applicable where each population represents neurons in one grid square. In this context, the same calculations used to compute $\{\mathbf{F},\mathbf{F}\}$ above can be used to write Eq. (S.24) as a Riemmann sum that converges to the integral equation

$$\begin{aligned}\{\mathbf{I},\mathbf{F}\} =& \{\mathbf{F},\mathbf{F}\} + \{\mathbf{R},\mathbf{F}\} \\ =& \{\mathbf{F},\mathbf{F}\} + \sqrt{N}\mathcal{W}\{\mathbf{S},\mathbf{F}\} + \mathcal{O}(1/\sqrt{N})\end{aligned} \tag{S.33}$$

where

$$\{\mathbf{I},\mathbf{F}\} = \begin{bmatrix} \{I_\mathrm{e},F_\mathrm{e}\} & \{I_\mathrm{e},F_\mathrm{i}\} \\ \{I_\mathrm{i},F_\mathrm{e}\} & \{I_\mathrm{i},F_\mathrm{i}\} \end{bmatrix}$$

is a matrix of spatially-dependent cross-spectra with components,

$$\{I_a,F_b\}(\mathbf{u},f) = \{I_a,F_b\}(\mathbf{x}-\mathbf{y},f) = \mathrm{avg}_{j\in a(\mathbf{x}),k\in b(\mathbf{y})}\langle I_j^a,F_k^b\rangle(f)$$

with the average taken over neurons, $j$, in population $a$ near coordinates $\mathbf{x} \in \Gamma$ and distinct neurons, $k$, in population $b$ near coordinates $\mathbf{y} \in \Gamma$ for which $\mathbf{u} = \mathbf{x}-\mathbf{y}$ (as in the definition of $\{F_{\mathrm{e}}, F_{\mathrm{e}}\}$ above). All other combinations, $\{\mathbf{S}, \mathbf{F}\}$, $\{\mathbf{R}, \mathbf{F}\}$, etc. are defined similarly. The term $\mathcal{W}$ in Eq. (S.33) is a matrix of integral operators defined by

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_{\mathrm{ee}} & \mathcal{W}_{\mathrm{ei}} \\ \mathcal{W}_{\mathrm{ie}} & \mathcal{W}_{\mathrm{ii}} \end{bmatrix}$$

with each component an integral operator defined by

$$\mathcal{W}_{ab}h(\mathbf{v}, f) = \iint_{\Gamma} w_{ab}(\mathbf{u}, f)h(\mathbf{v}-\mathbf{u}, f)d\mathbf{u} \tag{S.34}$$

where

$$w_{ab}(\mathbf{u}, f) = q_b j_{ab} \widetilde{\eta}_b(f) p_{ab}(\mathbf{u}).$$

is the mean-field connectivity from neurons in population $b = \mathrm{e}, \mathrm{i}$ at coordinates $\mathbf{x}$ to neurons in population $a = \mathrm{e}, \mathrm{i}$ at coordinates $\mathbf{y}$. The Fourier transform of the postsynaptic current waveform is given by

$$\widetilde{\eta}_b(f) = \frac{1}{1 + 2\pi i f \tau_b}$$

for $b = \mathrm{e}, \mathrm{i}$. Also recall that connection probability is defined by

$$p_{ab}(\mathbf{u}) = \overline{p}_{ab} G(\mathbf{u}; \alpha_b^2)$$

where

$$\overline{p}_{ab} = \frac{K_{ab}^{\mathrm{out}}}{N_a}$$

is the network-average number of connections probability from a neuron in population $b = \mathrm{e}, \mathrm{i}$ to a neuron in population $a = \mathrm{e}, \mathrm{i}$.

Under our assumption that neuronal transfer is $\mathcal{O}(1)$, we have that $\{\mathbf{I}, \mathbf{F}\} \propto \{\mathbf{S}, \mathbf{F}\}$ so that cancellation is required in Eq. (S.33). Specifically, in the asynchronous state,

$$\{\mathbf{R}, \mathbf{F}\} = \sqrt{N}\mathcal{W}\{\mathbf{S}, \mathbf{F}\} = -\{\mathbf{F}, \mathbf{F}\} + \mathcal{O}(1/\sqrt{N})$$

for large $N$. This explains why correlations between recurrent and feedforward inputs are approximately a negative reflection of feedforward-feedforward input correlations in Figure 3b.

These calculations are analogous to the calculations for networks with discrete populations considered above, except matrix equations like Eqs. (S.18) and (S.24) are replaced by integral equations like Eq. (S.33). We next show that these integral equations can be transformed into a sequence of matrix equations by transitioning to the spatial Fourier domain.

Since $p_{ab}(\mathbf{u}) = p_{ab}(\mathbf{x}-\mathbf{y})$ is defined using periodic boundary conditions, i.e. it is defined in terms of a wrapped Gaussian, the integral in Eq. (S.34) represents a circular convolution. This implies that the spatial Fourier series of the convolution is the product of the Fourier series. Specifically,

$$\iint_{\Gamma} [\mathcal{W}_{ab}h](\mathbf{v}, f) e^{-2\pi i \mathbf{n}\cdot\mathbf{v}} d\mathbf{v} = \widetilde{w}_{ab}(\mathbf{n}, f)\widetilde{h}(\mathbf{n}, f)$$

where

$$\begin{aligned}\widetilde{w}_{ab}(\mathbf{n},f) &= \iint_{\Gamma} w_{ab}(\mathbf{u},f)e^{-2\pi i\mathbf{n}\cdot\mathbf{u}}d\mathbf{u} \\ &= q_b j_{ab}\widetilde{\eta}_b(f)\iint_{\Gamma} p_{ab}(\mathbf{u})e^{-2\pi i\mathbf{n}\cdot\mathbf{u}}d\mathbf{u} \\ &= q_b j_{ab}\widetilde{\eta}_b(f)\widetilde{p}_{ab}(\mathbf{n})\end{aligned} \tag{S.35}$$

is the two-dimensional discrete Fourier transform of $w_{ab}(\mathbf{u},f)$ and similarly for $\widetilde{h}(\mathbf{n},f)$. Note that $\widetilde{w}_{ab}(\mathbf{0},0) = w^0_{ab}$ from the mean-field firing rate calculations surrounding Eq. (S.2) where $\mathbf{0} = (0,0)$ is the zero-vector. In these expressions, $\mathbf{n} = (n_1, n_2)$ represents a vector of discrete spatial frequency modes composed of a pair of integers and $\mathbf{n}\cdot\mathbf{u} = n_1u_1 + n_2u_2$ is the dot product. The inverse transform is given by the two-dimensional Fourier series,

$$w_{ab}(\mathbf{u},f) = \sum_{n_1,n_2=-\infty}^{\infty} \widetilde{w}_{ab}(\mathbf{n},f)e^{2\pi i\mathbf{n}\cdot\mathbf{u}}$$

and similarly for $\widetilde{h}(\mathbf{n},f)$. Therefore, for each spatial mode, $\mathbf{n}$, and each temporal frequency, $f$, the integral equation in Eq. (S.33) can be re-written as a system of matrix equations,

$$\begin{aligned}\langle\mathbf{I},\mathbf{F}\rangle(\mathbf{n},f) =& \langle\mathbf{F},\mathbf{F}\rangle(\mathbf{n},f) + \sqrt{N}W(\mathbf{n},f)\langle\mathbf{S},\mathbf{F}\rangle(\mathbf{n},f) \\ &+ \mathcal{O}(1/\sqrt{N})\end{aligned} \tag{S.36}$$

where $W(\mathbf{n},f)$ is a $2\times 2$ matrix of complex numbers for defined by

$$W(\mathbf{n},f) = \begin{bmatrix} \widetilde{w}_{\mathrm{ee}}(\mathbf{n},f) & \widetilde{w}_{\mathrm{ei}}(\mathbf{n},f) \\ \widetilde{w}_{\mathrm{ie}}(\mathbf{n},f) & \widetilde{w}_{\mathrm{ii}}(\mathbf{n},f) \end{bmatrix}$$

and where we have abused notation to define the spatial cross-spectral operators in the Fourier domain by

$$\langle\mathbf{I},\mathbf{F}\rangle(\mathbf{n},f) = \iint_{\Gamma}\{\mathbf{I},\mathbf{F}\}(\mathbf{u},f)e^{-2\pi i\mathbf{n}\cdot\mathbf{u}}d\mathbf{u}$$

and similarly for $\langle\mathbf{S},\mathbf{F}\rangle(\mathbf{n},f)$, $\langle\mathbf{F},\mathbf{F}\rangle(\mathbf{n},f)$, etc. Note that $W(\mathbf{0},0) = W_0$ from the mean-field firing rate calculations surrounding Eq. (S.2).

Note that for each spatial and temporal frequency, $\mathbf{n}$ and $f$, Eq. (S.36) has the same form as Eq. (S.15). In summary, we used Eq. (S.11) and the continuous structure of the spatial network to derive Eq. (S.36) which is analogous to Eq. (S.15) for the homogeneous network, except that it depends on the spatial Fourier mode. In the same way, we can use Eq. (S.12) to derive

$$\begin{aligned}\langle\mathbf{I},\mathbf{I}\rangle =& \langle\mathbf{F},\mathbf{F}\rangle + \sqrt{N}\left(W\langle\mathbf{S},\mathbf{F}\rangle + \langle\mathbf{F},\mathbf{S}\rangle W^*\right) + NW\langle\mathbf{S},\mathbf{S}\rangle W^* + WAW^* \\ &+ \mathcal{O}(1/\sqrt{N})\end{aligned} \tag{S.37}$$

which is identical to Eq. (S.19) except that $W$ and all other terms depend on the spatial Fourier mode, $\mathbf{n}$. The matrix, $A(\mathbf{n},f)$, represents contributions from the power spectral densities, which are uniform across space and therefore only contribute to the $\mathbf{n} = (0,0)$ Fourier mode to yield

$$A(\mathbf{n},f) = \begin{bmatrix} A_{\mathrm{e}}(f)/q_{\mathrm{e}} & 0 \\ 0 & A_{\mathrm{i}}(f)/q_{\mathrm{i}} \end{bmatrix}\delta_{\mathbf{n}}$$

where $\delta_{\mathbf{n}} = 1$ when $\mathbf{n} = (0,0)$ and $\delta_{\mathbf{n}} = 0$ when $\mathbf{n} \neq (0,0)$.

From here, the analysis of correlation proceeds identically to that in Section S.1.4 to finally obtain $\langle \mathbf{S}, \mathbf{S} \rangle \propto \mathcal{O}(1/N)$ with

$$\lim_{N\to\infty} N\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n}, f) = W^{-1}(\mathbf{n}, f)\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n}, f)W^{-*}(\mathbf{n}, f) + A(\mathbf{n}, f) \tag{S.38}$$

analogous to Eq. (S.21). While we only considered simulations with Gaussian-shaped connectivity profiles in the Results, Eq. (S.38) can be applied to any network for which the Fourier transforms of connection probabilities in Eq. (S.35) can be computed. However, the existence of the asynchronous state requires the convergence of the Fourier series,

$$\{\mathbf{S}, \mathbf{S}\}(\mathbf{u}, f) = \sum_{n_1, n_2 = -\infty}^{\infty} \langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n}, f) e^{2\pi i \mathbf{n}\cdot\mathbf{u}} \tag{S.39}$$

where $\langle \mathbf{S}, \mathbf{S} \rangle$ is the solution from Eq. (S.38) and recall that $\{\mathbf{S}, \mathbf{S}\}(\mathbf{u}, f)$ is the mean-field cross-spectral matrix between the spike trains of neurons at coordinates $\mathbf{x}$ and $\mathbf{y}$ for which $\mathbf{u} = \mathbf{x} - \mathbf{y}$. We next consider the specific case of Gaussian-shaped connection probabilities and show that an asynchronous solution (i.e., the convergence of the series in Eq. (S.39)) requires that recurrent projections are narrower than feedforward: $\alpha_{\mathrm{e}}, \alpha_{\mathrm{i}} < \alpha_F$.

To complete the computation, we need to compute $\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n}, f)$ and $W(\mathbf{n}, f)$ explicitly, then take the inverse transform in Eq. (S.38). Applying a Fourier transform to Eq. (S.32) gives

$$\begin{aligned}
\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n}, f) &= \iint_{\Gamma} \{\mathbf{F}, \mathbf{F}\}(\mathbf{u}, f) e^{-2\pi i \mathbf{n}\cdot\mathbf{u}} d\mathbf{u} \\
&= \begin{bmatrix} |\overline{w}_{\mathrm{e}F}(f)|^2 & \overline{w}_{\mathrm{e}F}(f)\overline{w}^*_{\mathrm{i}F}(f) \\ \overline{w}_{\mathrm{i}F}(f)\overline{w}^*_{\mathrm{e}F}(f) & |\overline{w}_{\mathrm{i}F}(f)|^2 \end{bmatrix} r_F e^{-4\pi^2\alpha_F^2|\mathbf{n}|^2}
\end{aligned}$$

where $|\mathbf{n}|^2 = n_1^2 + n_2^2$. Similarly, taking the spatial Fourier series of $w_{ab}(\mathbf{u}, f)$ gives

$$\widetilde{w}_{ab}(\mathbf{n}, f) = \overline{w}_{ab}(f) e^{-2\pi^2\alpha_b^2|\mathbf{n}|^2}$$

where

$$\overline{w}_{ab}(f) = q_b \overline{p}_{ab} j_{ab} \widetilde{\eta}_b(f)$$

is the network-averaged mean-field connectivity for $a, b = \mathrm{e}, \mathrm{i}$. Note that $\overline{w}_{ab}(0) = w^0_{ab}$ from the mean-field firing rate calculations surrounding Eq. (S.2). Putting this together gives,

$$W(\mathbf{n}, f) = \begin{bmatrix} \overline{w}_{\mathrm{ee}}(f) e^{-2\pi^2\alpha_{\mathrm{e}}^2|\mathbf{n}|^2} & \overline{w}_{\mathrm{ei}}(f) e^{-2\pi^2\alpha_{\mathrm{i}}^2|\mathbf{n}|^2} \\ \overline{w}_{\mathrm{ie}}(f) e^{-2\pi^2\alpha_{\mathrm{e}}^2|\mathbf{n}|^2} & \overline{w}_{\mathrm{ii}}(f) e^{-2\pi^2\alpha_{\mathrm{i}}^2|\mathbf{n}|^2} \end{bmatrix}.$$

Substituting these expressions into Eq. (S.38) gives

$$\lim_{N\to\infty} N\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n}, f) = \begin{bmatrix} \overline{C}_{\mathrm{ee}}(f) e^{-2\pi^2(2\alpha_F^2 - 2\alpha_{\mathrm{e}}^2)|\mathbf{n}|^2} & \overline{C}_{\mathrm{ei}}(f) e^{-2\pi^2(2\alpha_F^2 - \alpha_{\mathrm{i}}^2 - \alpha_{\mathrm{i}}^2)|\mathbf{n}|^2} \\ \overline{C}_{\mathrm{ie}}(f) e^{-2\pi^2(2\alpha_F^2 - \alpha_{\mathrm{e}}^2 - \alpha_{\mathrm{i}}^2)|\mathbf{n}|^2} & \overline{C}_{\mathrm{ii}}(f) e^{-2\pi^2(2\alpha_F^2 - 2\alpha_{\mathrm{i}}^2)|\mathbf{n}|^2} \end{bmatrix} + A(\mathbf{n}, f) \tag{S.40}$$

where

$$\overline{C}_{ab}(f) = W^{-1}(\mathbf{0}, f)\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{0}, f)W^{-*}(\mathbf{0}, f)$$

and $\mathbf{0} = (0, 0)$.

The Fourier series in Eq. (S.39) only converges if the entries in the matrix in Eq. (S.40) decay to zero at large Fourier modes. This, in turn, requires that $2\alpha_F^2 - \alpha_a^2 - \alpha_b^2 > 0$ for all four combinations of $a, b = \mathrm{e}, \mathrm{i}$. Thus, if $\alpha_\mathrm{e} > \alpha_F$ or $\alpha_\mathrm{i} > \alpha_F$, the terms in Eq. (S.40) do not have a well-defined inverse transform. This indicates that the system of integral equations in Eq. (S.33) (and the analogous integral equations generalizing Eq. (S.19) to space) do not have solutions [10, 6]. Thus, the equations that define the asynchronous state are not solvable, and therefore the asynchronous state is not self-consistent, when recurrent projections (excitatory or inhibitory) are broader in space than feedforward projections. Compare to results obtained for mean firing rates in [6].

When recurrent projections are narrower in space than feedforward projections (i.e., when $\alpha_\mathrm{e}, \alpha_\mathrm{i} < \alpha_F$), we can compute the Fourier series explicitly to obtain

$$\begin{aligned}\lim_{N\to\infty} N\{\mathbf{S}, \mathbf{S}\}(\mathbf{u}, f) = &\begin{bmatrix} \overline{C}_{\mathrm{ee}}(f) G(\mathbf{u}; 2\alpha_F^2 - 2\alpha_\mathrm{e}^2) & \overline{C}_{\mathrm{ei}}(f) G(\mathbf{u}; 2\alpha_F^2 - \alpha_\mathrm{i}^2 - \alpha_\mathrm{i}^2) \\ \overline{C}_{\mathrm{ie}}(f) G(\mathbf{u}; 2\alpha_F^2 - \alpha_\mathrm{e}^2 - \alpha_\mathrm{i}^2) & \overline{C}_{\mathrm{ii}}(f) G(\mathbf{u}; 2\alpha_F^2 - 2\alpha_\mathrm{i}^2) \end{bmatrix} \\ &+ \begin{bmatrix} A_\mathrm{e}(f)/q_\mathrm{e} & 0 \\ 0 & A_\mathrm{i}(f)/q_\mathrm{i} \end{bmatrix}\end{aligned}$$

where recall that $G(\mathbf{u}; \sigma^2)$ is a two-dimensional wrapped Gaussian with width parameter $\sigma$ (see above). For the simulations in Figures 3 and 4, we take $\alpha_\mathrm{e} = \alpha_\mathrm{i} = \alpha_\mathrm{rec}$ and rename $\alpha_\mathrm{ffwd} = \alpha_F$. This simplifies the expression above to yield

$$\begin{aligned}\lim_{N\to\infty} N\{\mathbf{S}, \mathbf{S}\}(\mathbf{u}, f) = &\begin{bmatrix} \overline{C}_{\mathrm{ee}}(f) & \overline{C}_{\mathrm{ei}}(f) \\ \overline{C}_{\mathrm{ie}}(f) & \overline{C}_{\mathrm{ii}}(f) \end{bmatrix} G(\mathbf{u}; 2\alpha_\mathrm{ffwd}^2 - 2\alpha_\mathrm{rec}^2) \\ &+ \begin{bmatrix} A_\mathrm{e}(f)/q_\mathrm{e} & 0 \\ 0 & A_\mathrm{i}(f)/q_\mathrm{i} \end{bmatrix}.\end{aligned} \tag{S.41}$$

Thus, when $\alpha_\mathrm{rec} < \alpha_\mathrm{ffwd}$ (as in Figure 3), spike train cross-spectra are $\mathcal{O}(1/N)$ and decay with distance like a Gaussian with width parameter

$$\sigma_{SS} = 2\alpha_\mathrm{ffwd}^2 - 2\alpha_\mathrm{rec}^2.$$

To compute spike-count covariances as a function of distance, consider a pair of neurons at coordinates $\mathbf{x}$ and $\mathbf{y}$. Our theoretical results show that their spike count correlation is proportional to $G(\mathbf{u}, \sigma_{SS}^2)$ where $\mathbf{u} = \mathbf{x} - \mathbf{y}$. Now note that the wrapped Gaussian function can be written as,

$$G(\mathbf{u}; \sigma_{SS}^2) = \frac{1}{2\pi\sigma_{SS}^2} e^{-d^2/(2\sigma_{SS}^2)} + \mathcal{O}\left(e^{-1/(2\sigma_{SS}^2)}\right) \tag{S.42}$$

where the last term represents artifacts from multiple "wraps" of the wrapped Gaussian and

$$d = \sqrt{\min(u_1, 1-u_1)^2 + \min(u_2, 1-u_2)^2}$$

is distance, measured periodically on $\Gamma$. Since $e^{-1/(2\sigma_{SS}^2)} \approx 1 \times 10^{-87}$ for the parameters used in Figure 3, the last term in Eq. (S.42) can be ignored. Combining this with Eq. (S.7) gives an equation for spike count covariances as a function of distance,

$$\begin{aligned}\begin{bmatrix} C_{SS}^{\mathrm{ee}}(d) & C_{SS}^{\mathrm{ei}}(d) \\ C_{SS}^{\mathrm{ie}}(d) & C_{SS}^{\mathrm{ii}}(d) \end{bmatrix} = &\frac{T}{N} \int_{-\infty}^{\infty} \frac{e^{-d^2/(2\sigma_{SS}^2)}}{2\pi\sigma_{SS}^2} \begin{bmatrix} \overline{C}_{\mathrm{ee}}(f) & \overline{C}_{\mathrm{ei}}(f) \\ \overline{C}_{\mathrm{ie}}(f) & \overline{C}_{\mathrm{ii}}(f) \end{bmatrix} K_T(f) df \\ &+ \frac{T}{N} \begin{bmatrix} r_\mathrm{e} F F_\mathrm{e}/q_\mathrm{e} & 0 \\ 0 & r_\mathrm{i} F F_\mathrm{i}/q_\mathrm{i} \end{bmatrix} + o(1/N)\end{aligned} \tag{S.43}$$

where $C_{SS}^{ab}(d)$ is the average spike count covariance between spike trains from populations $a, b = \mathrm{e}, \mathrm{i}$ and $T$ is the window size over which spikes are counted, $r_a$ is the average firing rate of neurons in population $a$ and $FF_a$ is the average Fano factor computed using the same counting window, $T$. The term $o(1/N)$ represents terms that converge to zero faster than $1/N$, *i.e.*, $N \times o(1/N) \to 0$ as $N \to \infty$. We used the fact that the spike count variance is proportional to the Fano factor,

$$FF_a = \frac{\text{spike count variance}}{\text{spike count mean}} = \frac{1}{r_a}\int_{-\infty}^{\infty} A_a(f)K_T(f)df.$$

Eq. (S.43) can be integrated numerically to obtain spike count covariances over any counting window size, but is greatly simplified by considering spike counts over large time windows.

If the counting window, $T$, is much larger than the correlation timescale of the spike trains, *i.e.* if cross-covariance function between the spike trains converges nearly to zero by lag $\tau = T$ then, from Eq. (S.9), the spike count covariances are approximated by taking $f = 0$ in Eq. (S.41). This gives,

$$\begin{aligned}\begin{bmatrix} C_{SS}^{\mathrm{ee}}(d) & C_{SS}^{\mathrm{ei}}(d) \\ C_{SS}^{\mathrm{ie}}(d) & C_{SS}^{\mathrm{ii}}(d) \end{bmatrix} &\approx \frac{T}{N}\left[W_0^{-1}C_{FF}^0 W_0^{-T}\right]\frac{e^{-d^2/(2\sigma_{SS}^2)}}{2\pi\sigma_{SS}^2} \\ &\quad + \frac{T}{N}\begin{bmatrix} r_{\mathrm{e}}FF_{\mathrm{e}}/q_{\mathrm{e}} & 0 \\ 0 & r_{\mathrm{i}}FF_{\mathrm{i}}/q_{\mathrm{i}} \end{bmatrix}\end{aligned}$$

where

$$W_0 = \begin{bmatrix} q_{\mathrm{e}}p_{\mathrm{ee}}j_{\mathrm{ee}} & q_{\mathrm{i}}p_{\mathrm{ei}}j_{\mathrm{ei}} \\ q_{\mathrm{e}}p_{\mathrm{ie}}j_{\mathrm{ie}} & q_{\mathrm{i}}p_{\mathrm{ii}}j_{\mathrm{ii}} \end{bmatrix}$$

is the mean-field connectivity matrix from Eq. (S.3), $W_0^{-T}$ is the inverse of its transpose and

$$C_{FF}^0 = \langle \mathbf{F}, \mathbf{F}\rangle(\mathbf{0}, 0) = \begin{bmatrix} j_{eF}^2\overline{p}_{eF}^2 & j_{eF}\overline{p}_{eF}j_{iF}\overline{p}_{iF} \\ j_{iF}\overline{p}_{iF}j_{eF}\overline{p}_{eF} & j_{iF}^2 p_{iF}^2 \end{bmatrix} q_F r_F$$

quantifies the network-averaged zero-frequency cross-spectra between neurons' feedforward inputs.

However, this only gives the *covariance* not the *correlation* between spike counts. Moreover, it requires the computation of the Fano factors, which we have not derived. Under the assumption that spiking is Poisson-like in the network, $FF_a \approx 1$. Combining this with the fact that the spike count variance of spike trains in population $a = \mathrm{e}, \mathrm{i}$ is equal to $FF_a r_a T$, we obtain an approximation to the spike count correlation coefficients,

$$\begin{aligned}\begin{bmatrix} \rho_{SS}^{\mathrm{ee}}(d) & \rho_{SS}^{\mathrm{ei}}(d) \\ \rho_{SS}^{\mathrm{ie}}(d) & \rho_{SS}^{\mathrm{ii}}(d) \end{bmatrix} &\approx \frac{1}{N}\left[R^{-1}W_0^{-1}C_{FF}^0 W_0^{-T}R^{-1}\right]\frac{e^{-d^2/(2\sigma_{SS}^2)}}{2\pi\sigma_{SS}^2} \\ &\quad + \begin{bmatrix} 1/(Nq_{\mathrm{e}}) & 0 \\ 0 & 1/(Nq_{\mathrm{i}}) \end{bmatrix}\end{aligned} \tag{S.44}$$

where

$$R = \begin{bmatrix} \sqrt{r}_{\mathrm{e}} & 0 \\ 0 & \sqrt{r}_{\mathrm{i}} \end{bmatrix}$$

and firing rates, $r_a$, in the limit of large $N$ are given by Eq. (S.5). The $1/(q_a N)$ terms in Eq. (S.44) represent contributions from intrinsically generated variability to correlations [5], which are orders of magnitude smaller than contributions from feedforward input covariability, $C_{FF}^0$. We can therefore

safely omit these terms from our computations, since all of our examples had strongly correlated feedforward input. Performing the matrix arithmetic gives an approximation to the correlation between excitatory neurons as a function of distances,

$$\rho^{\mathrm{ee}}_{SS}(d) = \frac{\left(j_{\mathrm{ei}}j_{\mathrm{i}F}p_{\mathrm{ei}}p_{\mathrm{i}F} - j_{\mathrm{e}F}j_{\mathrm{ii}}p_{\mathrm{e}F}p_{\mathrm{ii}}\right)^2 q_F r_F}{N\left(j_{\mathrm{ei}}j_{\mathrm{ie}}p_{\mathrm{ei}}p_{\mathrm{ie}} - j_{\mathrm{ee}}j_{\mathrm{ii}}p_{\mathrm{ee}}p_{\mathrm{ii}}\right)^2 q_{\mathrm{e}}^2 r_{\mathrm{e}} 2\pi\sigma_{SS}^2} e^{-d^2/(2\sigma_{SS}^2)} \tag{S.45}$$

To generate the dashed red curve in Figure 3e, we used Eq. (S.45), except that we had to account for the coarse binning of neurons by distance. When computing the average correlation as a function of distance from simulations (the solid curve in Figure 3e), we coarsely binned neuron pairs by distance. For example, the first data point represents the average spike count correlation between sampled pairs with distances between 0 and 0.15. This coarse binning was necessary to reduce the statistical error in estimating extremely weak correlations from simulations. However, because of the two-dimensional domain, distant pairs are over-represented within each bin. For example, the first bin contains fewer pairs with a distance between 0 and 0.075 than it contains pairs with a distance between 0.075 and 0.15. This is further complicated by the fact that distance is measured periodically. To compute the dashed curve in Figure 3e, we computed the pairwise distance between all sampled neurons from the simulation, substituted all of these distances into Eq. (S.44), then computed the average value over each distance bin. This naturally corrects the sampling problem.

### S.1.8 Approximate correlations at finite network size when recurrent projections are broader than feedforward

Above, we concluded that spike count correlations are $\mathcal{O}(1/N)$ at all distances only when feedforward projections are spatially broader than recurrent projections ($\alpha_{\mathrm{e}}, \alpha_{\mathrm{i}} < \alpha_F$), and we also derived the asymptotic correlations in that case. We have not yet derived the correlations when recurrent projections are broader than feedforward ($\alpha_{\mathrm{e}}, \alpha_{\mathrm{i}} > \alpha_F$). Here, we provide a linear approximation at finite $N$ that explains the non-monotonic dependence of correlation on distance when $\alpha_{\mathrm{e}}, \alpha_{\mathrm{i}} > \alpha_F$.

The central idea behind this approximation comes from previous studies of correlations transfer [11, 1]. Our computations above do not depend on the precise transfer from input correlation to spiking correlation, and instead relied only on the assumption that this transfer is $\mathcal{O}(1)$ in the sense of Eqs. (S.13). We now show that accounting for correlation transfer allows us to derive an approximation to the correlation structure at finite $N$ that is applicable even when $\alpha_{\mathrm{e}}, \alpha_{\mathrm{i}} > \alpha_F$.

Consider two neurons in the network, receiving input currents, $I_j^a(t)$ and $I_k^b(t)$. When input correlations are weak, ($\langle I_j^a, I_k^b\rangle \ll \langle I_j^a, I_j^a\rangle$), the cross-spectrum between two neurons' spike trains is approximately linearly related to the cross-spectrum between their input currents [1],

$$\langle S_j^a, S_k^b\rangle \approx L_j^a \langle I_j^a, I_k^b\rangle L_k^{b*} \tag{S.46}$$

where $L_j^a(f)$ is the susceptibility function of neuron $j$ in population $a = \mathrm{e}, \mathrm{i}$ [11, 1]. Likewise, the cross-spectra between feedforward inputs and spike trains are approximated by

$$\langle S_j^a, F_k^b\rangle \approx L_j^a \langle I_j^a, F_k^b\rangle. \tag{S.47}$$

Combining these approximations with Eqs. (S.11) and (S.12) gives a linear approximation to the entire $N \times N$ matrix of cross-spectra in the network [11, 4],

$$\langle S, S\rangle = \left(L^{-1} - \frac{1}{\sqrt{N}}J\widetilde{H}\right)^{-1} \langle F, F\rangle \left(L^{-1} - \frac{1}{\sqrt{N}}J\widetilde{H}\right)^{-*}. \tag{S.48}$$

where $^{-*}$ denotes the inverse of the conjugate transpose, $L(f)$ is a diagonal $N \times N$ matrix of each neuron's susceptibility function and all other terms are $N \times N$ matrices defined above. Eq. (S.48), which is accurate to first order in the magnitude of correlations, has been derived for weakly coupled networks of integrate-and-fire neurons with white noise inputs [11] and for networks of linearly interacting point processes [4], except those studies do not include the $\sqrt{N}$ scaling that is characteristic of balanced networks.

In sparsely connected networks with a dominant source of Gaussian white noise input, the susceptibility functions can be computed using a Fokker-Planck formalism [11]. Unfortunately, this approach is not directly applicable to the networks considered in our study because neurons are densely and strongly connected and inputs are temporally correlated, so Fokker-Planck techniques cannot be used to compute the susceptibility functions.

However, recall from Eq. (S.9) that spike count covariances computed over large counting windows are given by the zero-frequency cross-spectra. Hence, if we are only interested in computing spike count covariances over large time windows, we only need to compute the susceptibility functions at zero frequency, $L_j^a(0)$. This is made possible by noting that a neuron's susceptibility function at $f = 0$ is simply the gain of the neuron, *i.e.* the derivative of the neuron's f-I curve evaluated at its steady-state firing rate in the network [1]. Thus, if we knew the gain of every neuron in a network, we could use Eq. (S.48) to compute the entire matrix of cross-spectra. However, we do not know the gains and, even if we did, this large $N \times N$ matrix computation would provide little intuition.

Instead, we extend the spatial mean-field approach developed above to transform the $N \times N$ equation (S.48) into the $2 \times 2$ mean-field equation,

$$\langle \mathbf{S}, \mathbf{S} \rangle = \left( G^{-1} - \sqrt{N} W \right)^{-1} \langle \mathbf{F}, \mathbf{F} \rangle \left( G^{-1} - \sqrt{N} W \right)^{-*}. \tag{S.49}$$

where

$$G = \begin{bmatrix} g_\mathrm{e} & 0 \\ 0 & g_\mathrm{i} \end{bmatrix}$$

and $g_\mathrm{e}$ and $g_\mathrm{i}$ are the average gains of the excitatory and inhibitory neurons in the network. The $2 \times 2$ matrices, $\langle \mathbf{S}, \mathbf{S} \rangle$, $W$ and $\langle \mathbf{F}, \mathbf{F} \rangle$, are the same as defined in the previous section except they are implicitly evaluated at $f = 0$. In particular,

$$W(\mathbf{n}) = \begin{bmatrix} \widetilde{w}_\mathrm{ee}(\mathbf{n}) & \widetilde{w}_\mathrm{ei}(\mathbf{n}) \\ \widetilde{w}_\mathrm{ie}(\mathbf{n}) & \widetilde{w}_\mathrm{ii}(\mathbf{n}) \end{bmatrix}$$

where

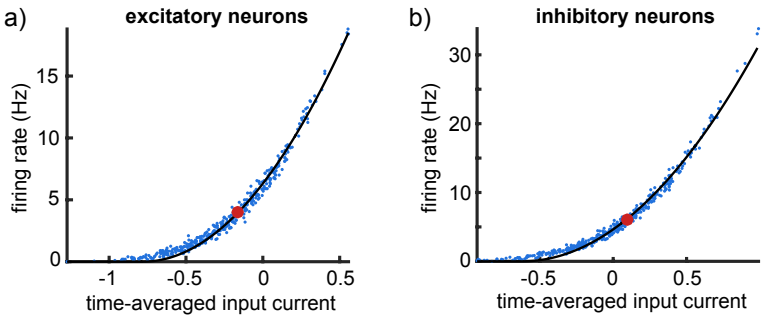
$$\widetilde{w}_{ab}(\mathbf{n}) = q_b j_{ab} \widetilde{p}_{ab}(\mathbf{n}).$$

Similarly,

$$\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n}) = \begin{bmatrix} |w_{\mathrm{e}F}(\mathbf{n})|^2 & w_{\mathrm{e}F}(\mathbf{n}) w_{\mathrm{i}F}^*(f) \\ w_{\mathrm{i}F}(\mathbf{n}) w_{\mathrm{e}F}^*(\mathbf{n}) & |w_{\mathrm{i}F}(\mathbf{n})|^2 \end{bmatrix} r_F$$

where $r_F$ is the firing rate of neurons in the feedforward layer and

$$w_{aF} = \sqrt{q_F} j_{aF} \widetilde{p}_{aF}(\mathbf{n})$$




Supplementary Figure 1: **Fitting f-I curves of neurons from simulations. a)** *Firing rate as a function of time-averaged total synaptic input from the simulation in Figure 4. Blue dots are from 400 randomly selected excitatory neurons, the black curve is the best fit thresholded quadratic and the red dot is placed at the population-average firing rates, at which the gain can be computed as the derivative of the black curve.* **b)** *Same as (a), but for inhibitory neurons.*

In these expressions, $q_b$ and $j_{ab}$ are as defined previously and

$$\widetilde{p}_{ab}(\mathbf{n}) = \iint p_{ab}(\mathbf{u}) e^{-2\pi i \mathbf{n}\cdot u} d\mathbf{u} \tag{S.50}$$

where $p_{ab}(\mathbf{u})$ is the probability of connection from a neuron in population $b = \mathrm{e}, \mathrm{i}, F$ to a neuron in population $a = \mathrm{e}, \mathrm{i}$ at coordinates $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{u} = \mathbf{x} - \mathbf{y}$.

Note that, if $W$ is invertible and the small contribution from $A$ is ignored, then Eq. (S.49) implies Eq. (S.38) in the large $N$ limit. In other words, Eq. (S.49) provides a finite-$N$ generalization to Eq. (S.38). Moreover, since $\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n})$ is a convergent Fourier series and $G^{-1}$ does not depend on $\mathbf{n}$, then the Fourier series in Eq. (S.39) necessarily converges at finite $N$ when Eq. (S.49) is used to compute the Fourier coefficients, $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$. Therefore, Eq. (S.49) gives a well-defined solution at finite $N$ even when recurrent projections are broader than feedforward ($\alpha_\mathrm{e}, \alpha_\mathrm{i} > \alpha_F$). Indeed, Eq. (S.49) gives a well-defined solution for any chosen connection probability profiles, not just Gaussian-shaped connectivity. One only needs to compute the Fourier transform of connection probability profiles from Eq. (S.50).

The only remaining question is how to compute the average gains, $g_\mathrm{e}$ and $g_\mathrm{i}$. To accomplish this, we sampled the synaptic input currents and firing rates of 400 excitatory and 400 inhibitory neurons (Supplementary Figure 1a,b; blue dots). We then fit the relationship between these mean inputs ($I$) and mean firing rates ($r$) to a thresholded quadratic f-I curve,

$$r = \begin{cases} a_1(I-\theta) + a_2(I-\theta)^2 & I > \theta \\ 0 & I \leq \theta \end{cases},$$

obtaining the best fit values of $a_1$, $a_2$ and $\theta$ using the curve fitting toolbox in Matlab (Supplementary Figure 1a,b; black curve). Once the f-I curve is fit, the gain is approximated by the derivative $g = dr/dI = a_1 + 2a_2 I$ evaluated at the mean firing rate (Supplementary Figure 1a,b; red dot).

Once the gains are approximated, Eq. (S.49) can be used to compute $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ at any given $\mathbf{n} = (n_1, n_2)$. The zero-frequency ($f = 0$) cross-spectral matrices as a function of neuron distance can then be computed numerically by numerically summing the Fourier series in Eq. (S.38). Spike

count correlations over large time windows are then given by applying Eq. (S.9) following the same approach used in Supplementary Section S.1.7.

This approach was used to compute the dashed red curve in Figure 4e. We also tested this approximation on several examples with various widths and shapes of the feedforward and recurrent connection probability profiles and found it to be highly accurate (Supplementary Figure 2).

We first considered an example that is identical to the one in Figure 4, except recurrent and feedforward connection probabilities were made narrower by a factor of two ($\sigma_{\text{rec}} = 0.125$, $\sigma_{\text{ffwd}} = 0.05$). This produced a similar non-monotonicity with a larger peak correlation (Supplementary Figure 2a; compare to Figure 4e).

We next considered an example where recurrent connections were only slightly broader than feedforward ($\alpha_{\text{rec}} = 0.055$, $\alpha_{\text{ffwd}} = 0.05$; Supplementary Figure 2b, bottom). The non-monotonicity persisted in this case, but was less dramatic and correlations were weaker overall (Supplementary Figure 2b, top). Similar effects were observed when feedforward and recurrent projections had identical widths ($\alpha_{\text{rec}} = \alpha_{\text{ffwd}} = 0.05$; Supplementary Figure 2c). When feedforward projections were slightly broader than recurrent ($\alpha_{\text{rec}} = 0.05$, $\alpha_{\text{ffwd}} = 0.055$), there was a weak non-monotonicity and the theoretical calculations were less accurate (Supplementary Figure 2d). Hence, at finite $N$, correlations can still depend non-monotonically on distance when $\alpha_{\text{ffwd}} > \alpha_{\text{rec}}$.

We next considered an example where, as in Figures 6-7 of the main text, the recurrent excitatory connections are broader than feedforward connections, but recurrent inhibition is narrower than feedforward ($\sigma_{\text{e}} = 0.15$, $\sigma_{\text{i}} = 0.05$, $\sigma_F = 0.075$). Since the asynchronous state requires that excitatory *and* inhibitory recurrent projections are narrower than feedforward (see above), the moderate magnitude and non-monotonic shape of correlations persisted (Supplementary Figure 2e).

We then considered an example where feedforward connection probabilities decay to half their peak value instead of decaying to zero,

$$p_{aF}(\mathbf{x}-\mathbf{y}) \propto 0.5G(\mathbf{x}-\mathbf{y};\alpha_{\text{ffwd}}) + 0.5$$

for $a = \text{e}, \text{i}$, but they decay to this value faster than recurrent projections decay to zero ($\alpha_{\text{rec}} = 0.1$ and $\alpha_{\text{ffwd}} = 0.05$; Supplementary Figure 2f, bottom). Feedforward connection probabilities could be viewed as "broader" than recurrent in this example since they do not decay to zero at large distances, but recurrent connection probabilities do. However, the moderate correlation magnitude and non-monotonic dependence of correlation on distance persists (Supplementary Figure 2f, top), suggesting that this connectivity is not consistent with the asynchronous state as $N \to \infty$. This can be understood by computing the Fourier series of feedforward connection probabilities,

$$\widetilde{p}_{aF}(\mathbf{n}) \propto 0.5e^{-4\pi^2\alpha_{\text{ffwd}}^2|\mathbf{n}|^2} + 0.5\delta_{\mathbf{n}}$$

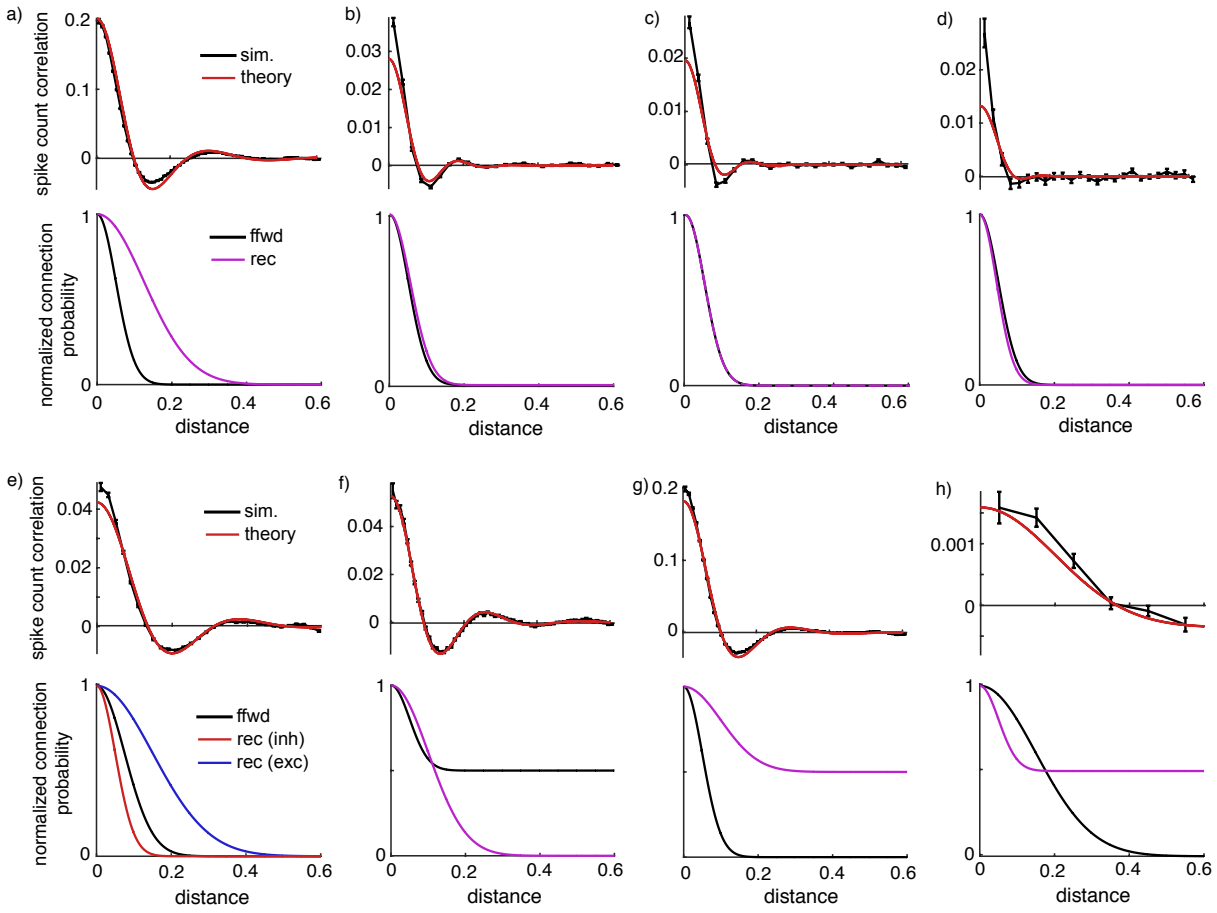
where $\delta_{\mathbf{n}} = 1$ when $\mathbf{n} = (0,0)$ and $\delta_{\mathbf{n}} = 0$ otherwise. Therefore, the overall dependence of $\widetilde{p}_{aF}(\mathbf{n})$ on $\mathbf{n} \neq (0,0)$ is unchanged. As a consequence, we still have

$$\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n}) \propto e^{-4\pi^2\alpha_{\text{ffwd}}^2|\mathbf{n}|^2}$$

and

$$W(\mathbf{n}) \propto e^{-2\pi^2\alpha_{\text{rec}}^2|\mathbf{n}|^2}.$$

so that the $N \to \infty$ asynchronous solution from Eq. (S.38) does not have a well-defined inverse transform when $\alpha_{\text{rec}} > \alpha_{\text{ffwd}}$.




Supplementary Figure 2: **Correlation as a function of distance for various connection probability profiles. a)** *Bottom: Connection probability from feedforward layer (black) and connection probability within the recurrent layer (purple) as a function of neuron distance. Each curve is normalized by its peak. Top: Mean spike count correlation between excitatory neurons as a function of neuron distance from simulations (black; ±SEM; from randomly sampling 5000 excitatory neurons) and from the theoretical calculation using Eq.* (S.49) *(red).* **b-h)** *Same as (a) except in (e) where recurrent excitatory (blue) and inhibitory (red) connection probabilities have different profiles.*

We next considered an example where recurrent connection probabilities decay to a non-zero value,

$$p_{ab}(\mathbf{x}-\mathbf{y}) \propto 0.5G(\mathbf{x}-\mathbf{y};\alpha_{\text{rec}}) + 0.5$$

for $a, b = \text{e}, \text{i}$, feedforward connection probabilities decay to zero,

$$p_{aF}(\mathbf{x}-\mathbf{y}) \propto G(\mathbf{x}-\mathbf{y};\alpha_{\text{ffwd}})$$

and where $\alpha_{\text{rec}} < \alpha_{\text{ffwd}}$ ($\alpha_{\text{rec}} = 0.05$ and $\alpha_{\text{ffwd}} = 0.15$; Supplementary Figure 2g, bottom). Perhaps unsurprisingly, the moderate magnitude of correlations and the non-monotonic dependence on distance persists in this case (Supplementary Figure 2g, top).

We next considered an example where recurrent connection probabilities decay to a non-zero value and feedforward connection probability decays to zero (as above), except with $\alpha_{\text{rec}} < \alpha_{\text{ffwd}}$ ($\alpha_{\text{rec}} = 0.05$ and $\alpha_{\text{ffwd}} = 0.15$; Supplementary Figure 2h, bottom). Even though recurrent connection probabilities are "broader" in the sense that they do not decay to zero, spike count correlations are extremely weak (Supplementary Figure 2h, top). This can again be understood by noting that only the zero-Fourier mode is affected by having recurrent connection probabilities decay to a non-zero values. Therefore Eq. (S.38) has a well-defined inverse transform when $\alpha_{\text{rec}} < \alpha_{\text{ffwd}}$, so the asynchronous solution is realized.

To better understand these results, we now derive a more precise condition that must be satisfied for correlations to be $\mathcal{O}(1/N)$ over every distance. We showed above that the cancellation required for $\mathcal{O}(1/N)$ correlations at every distance depends on the convergence of the Fourier series in Eq. (S.39) with coefficients defined by Eq. (S.38). This, in turn, requires that $\langle \mathbf{S}, \mathbf{S} \rangle$ from Eq. (S.38) decays to zero as $|\mathbf{n}| \to \infty$, and therefore only depends on the behavior of $\langle \mathbf{F}, \mathbf{F} \rangle$ and $W(\mathbf{n})$ at high Fourier modes ($|\mathbf{n}|$ large). Note that $W(\mathbf{n}) \propto \widetilde{p}_{ab}(\mathbf{n})$ for $a, b = \text{e}, \text{i}$ and $\langle \mathbf{F}, \mathbf{F} \rangle \propto [\widetilde{p}_{aF}(\mathbf{n})]^2$ so, from Eq. (S.38), $\mathcal{O}(1/N)$ correlations at every distance requires that

$$\lim_{|\mathbf{n}|\to\infty} \left| \frac{\widetilde{p}_{aF}(\mathbf{n})}{\widetilde{p}_{ab}(\mathbf{n})} \right| \to 0 \tag{S.51}$$

for $a, b = \text{e}, \text{i}$. In other words, the spatial Fourier series of feedforward connection profiles must decay to zero faster than the Fourier series of recurrent connection profiles. Compare to previous findings for the existence of a balanced firing rate solution [6].

For Gaussian-shaped connectivity, Eq. (S.51) implies that $\alpha_{\text{rec}} < \alpha_{\text{ffwd}}$. More generally, a quickly decaying Fourier series implies that power is concentrated at low spatial frequencies, giving a broad, slowly varying profile in the spatial domain. Conversely, a slowly decaying Fourier series implies high frequency modes representing sharp changes in connectivity over short ranges. Therefore, recurrent connectivity profiles must display "sharper" or "narrower" features than feedforward if correlations are to be $\mathcal{O}(1/N)$ at all distances. It is easily checked that Eq. (S.51) is violated in panels a-g of Supplementary Figure 2 but satisfied in panel h.

When Eq. (S.51) is violated, Eq. (S.38) no longer gives a well-defined asymptotic correlation profile since its inverse transform, *i.e.* the sum of the Fourier series in Eq. (S.39), does not exist. Instead, the finite $N$ correction in Eq. (S.49) must be used in place of Eq. (S.38). To understand why this gives rise to negative correlations and a non-monotonic dependence of correlation on distance, first note that the zero spatial Fourier mode ($\mathbf{n} = (0, 0)$) represents the average over the entire network,

$$\langle \mathbf{S}, \mathbf{S} \rangle(0, 0) = \iint_{\Gamma} \{\mathbf{S}, \mathbf{S}\}(\mathbf{u}) d\mathbf{u}$$

As long as the balance conditions in Eq. (S.6) are satisfied, $W(0,0)$ is invertible and therefore

$$\langle \mathbf{S}, \mathbf{S} \rangle(0,0) \propto \mathcal{O}(1/N)$$

by Eq. (S.49). Therefore, the average cross-spectral density over all neuron pairs is nearly zero for all of the networks we considered (see, *e.g.*, Figures 3e and 4e dashed gray). This can also be seen by applying the homogeneous mean-field theory of correlations from Section S.1.4 to a spatially extended network, *i.e.* by ignoring the spatial dependence and averaging over pairs at all distanced in the calculations.

Therefore, when Eq. (S.51) is not satisfied but $W(0,0)$ is invertible, correlations cannot be $\mathcal{O}(1/N)$ at *every* distance, but they are $\mathcal{O}(1/N)$ when *averaged* over all neuron pairs. Hence, positive and negative correlations at various distances must cancel to obtain the $\mathcal{O}(1/N)$ average. This explains the combination of positive and negative correlations in Figure 4 and Supplementary Figure 2a-g. For all examples we considered, this cancellation was realized by a non-monotonic dependence of correlation on distance. This shape determined by the precise dependence of $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ on $\mathbf{n}$ as determined by Eq. (S.49).

Note that, despite the fact that the average pairwise correlation is $\mathcal{O}(1/N)$, the average value of the correlation curves in Figure 4e and Supplementary Figure 2 is not close to zero. This is because, in the square-shaped network, there are more neuron pairs at larger distances than there are nearby neuron pairs. When the average is taken over all neuron pairs, the negative correlations at moderate distances contribute more to the average than the positive correlations at smaller distances. Therefore, the positive average correlation between nearby neuron pairs must be larger than the negative average correlation between more distant (but more numerous) pairs.

We have shown why correlations are not $\mathcal{O}(1/N)$ at every distance when $\alpha_{\text{rec}} > \alpha_{\text{ffwd}}$ and also why there is a combination of positive and negative correlations in this case. To understand the source of the non-monotonic dependence of correlation on distance, we must inspect the dependence of $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ on $\mathbf{n}$. Consider, for simplicity, Gaussian-shaped connection probabilities in which recurrent excitatory and inhibitory connections have the same width, $\alpha_{\text{rec}} = \alpha_{\text{e}}, \alpha_{\text{i}}$ and where the feedforward connection profile is narrower, $\alpha_{\text{ffwd}} = \alpha_{FF} < \alpha_{\text{rec}}$ as in Figure 4. In this case, we can write

$$\langle \mathbf{F}, \mathbf{F} \rangle(\mathbf{n}) = \langle \mathbf{F}, \mathbf{F} \rangle(0,0) e^{-4\alpha_{\text{ffwd}}^2 \pi^2 |\mathbf{n}|^2},$$

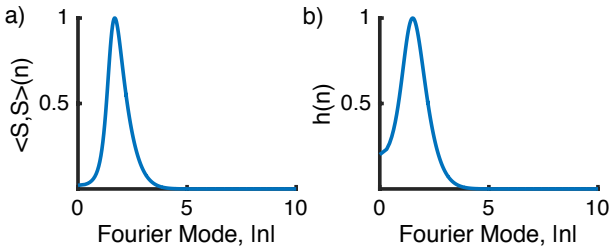
$$W(\mathbf{n}) = W(0,0) e^{-2\alpha_{\text{rec}}^2 \pi^2 |\mathbf{n}|^2}.$$

Note also that $G$ does not depend on $\mathbf{n}$, so effectively $G(\mathbf{n}) = G(0,0)$. The shape of the correlation profiles seen in Figure 4e and Supplementary Figure 2 is inherited by the dependence of $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ on $\mathbf{n}$. As a rough approximation to the dependence of $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ on $\mathbf{n}$, we therefore ignore the contribution of $\langle \mathbf{F}, \mathbf{F} \rangle(0,0)$ and $W(0,0)$ and $G(0,0)$ by setting them equal to 1 in Eq. (S.49), which gives the rough approximation

$$\langle \mathbf{S}, \mathbf{S} \rangle \sim h(\mathbf{n}) := \epsilon \left( \frac{e^{-2\alpha_{\text{ffwd}}^2 \pi^2 |\mathbf{n}|^2}}{\sqrt{\epsilon} + e^{-2\alpha_{\text{rec}}^2 \pi^2 |\mathbf{n}|^2}} \right)^2 \tag{S.52}$$

where $\epsilon \propto 1/N$ and recall that this only captures the *shape* of $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ as $|\mathbf{n}|$ changes, not the precise value. When $\alpha_{\text{rec}} < \alpha_{\text{ffwd}}$, this becomes

$$h(\mathbf{n}) = \epsilon e^{-4(\alpha_{\text{ffwd}}^2 - \alpha_{\text{rec}}^2)\pi^2 |\mathbf{n}|^2} + o(\epsilon) \tag{S.53}$$




Supplementary Figure 3: **Dependence of cross-spectra on Fourier mode. a)** *The cross-spectrum,* $\langle S_e, S_e \rangle(\mathbf{n})$*, between excitatory neurons as a function of the magnitude of the spatial Fourier mode,* $|\mathbf{n}|$*, computed using Eq.* (S.49) *with parameters from the example in Figure 4. The curve is normalized by its peak.* **b)** *The approximation,* $h(\mathbf{n})$*, to the cross-spectrum computed using Eq.* (S.53) *and plotted as a function of spatial Fourier mode,* $|\mathbf{n}|$*, with parameters from the simulation in Figure 4.*

and inverting the Fourier transform gives

$$\{\mathbf{S}, \mathbf{S}\}(\mathbf{u}) \sim \epsilon G(\mathbf{u}; 2\alpha_{\text{ffwd}}^2 - 2\alpha_{\text{rec}}^2) + o(\epsilon)$$

in agreement with the solution from Eq. (S.41) for the asynchronous state.

When $\alpha_{\text{rec}} > \alpha_{\text{ffwd}}$, $h(\mathbf{n})$ does not have a well-defined inverse Fourier transform at $\epsilon = 0$, so the $o(\epsilon)$ term in Eq. (S.53) needs to be accounted for. At low spatial frequencies ($|\mathbf{n}|$ small), $e^{-2\alpha_{\text{ffwd}}^2 \pi^2 |\mathbf{n}|^2}$ and $e^{-2\alpha_{\text{rec}}^2 \pi^2 |\mathbf{n}|^2}$ are moderate in magnitude, so the $\sqrt{\epsilon}$ in Eq. (S.52) can be ignored and Eq. (S.53) is accurate, so $h(\mathbf{n})$ is small in magnitude (on account of the $\epsilon = 1/N$ coefficient). When $|\mathbf{n}|$ is large enough that $e^{-2\alpha_{\text{rec}}^2 \pi^2 |\mathbf{n}|^2} \ll \sqrt{\epsilon}$, but not large enough that $e^{-2\alpha_{\text{ffwd}}^2 \pi^2 |\mathbf{n}|^2} \ll \sqrt{\epsilon}$, we can ignore the $e^{-2\alpha_{\text{rec}}^2 \pi^2 |\mathbf{n}|^2}$ in the denominator of Eq. (S.52) to get

$$h(\mathbf{n}) \approx e^{-4\alpha_{\text{ffwd}}^2 \pi^2 |\mathbf{n}|^2}.$$

Thus, $h(\mathbf{n})$ initially increases with $|\mathbf{n}|$. For sufficiently large $|\mathbf{n}|$,

$$e^{-2\alpha_{\text{ffwd}}^2 \pi^2 |\mathbf{n}|^2} \ll \sqrt{\epsilon},$$

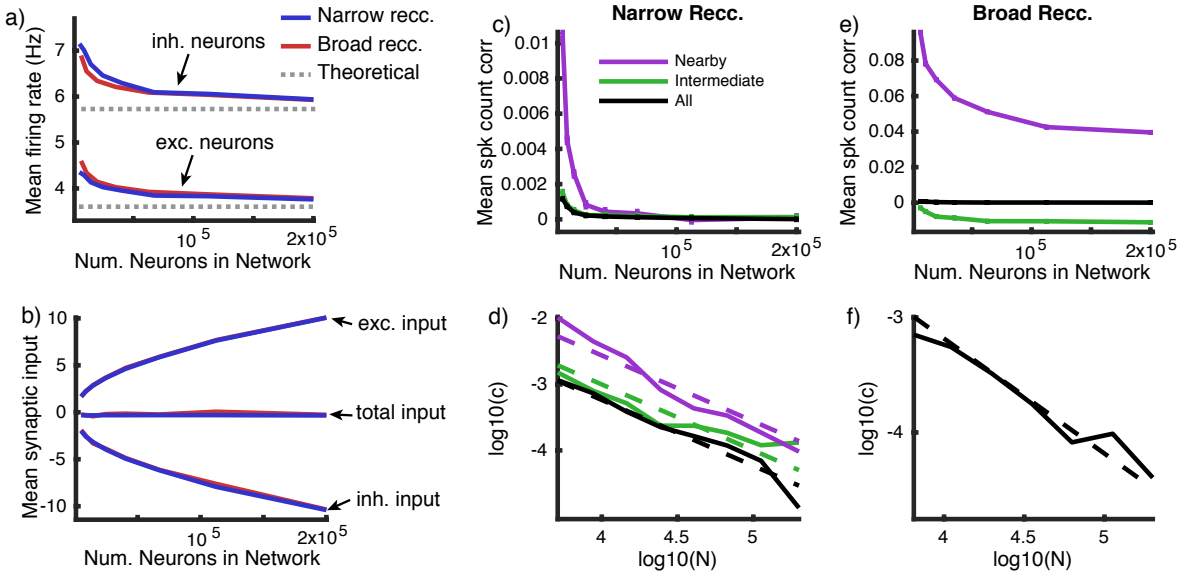
so

$$h(\mathbf{n}) \approx 0.$$

Thus, $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ initially increases from a $\mathcal{O}(1/N)$ value toward a $\mathcal{O}(1)$ value, then decreases toward zero as $|\mathbf{n}|$ grows. This is demonstrated in Supplementary Figure 3. Hence, $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ has a peak value at some $|\mathbf{n}| > 0$. Now note that, as long as connectivity is symmetric, $p_{ab}(u_1, u_2) = p_{ab}(-u_1, u_2)$ and $p_{ab}(u_1, u_2) = p_{ab}(u_1, -u_2)$, then the same holds for $\{\mathbf{S}, \mathbf{S}\}(\mathbf{u})$, so Fourier series in Eq. (S.39) can be re-written as

$$\{\mathbf{S}, \mathbf{S}\}(\mathbf{u}) = \sum_{n_1, n_2 = -\infty}^{\infty} \langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n}) \cos(2\pi n_1 u_1) \cos(2\pi n_2 u_2). \tag{S.54}$$

If the peak value of $\langle \mathbf{S}, \mathbf{S} \rangle(\mathbf{n})$ occurs at some $|\mathbf{n}| > 1$ (as in the example from Figure 4, demonstrated in Supplementary Figure 3), then the series in Eq. (S.39) has a non-monotonic dependence on $u_1$ and $u_2$ since a higher Fourier mode dominates.




Supplementary Figure 4: **Spike count correlations, firing rates, and input currents at increasing network size. a)** *Average excitatory (bottom) and inhibitory (top) firing rates from simulations with increasing numbers of neurons (solid curves) and from from the theoretical predictions in Eqs.* (S.5) *(dotted gray lines). Blue (red) curves are for simulations identical to the one in Figure 3 (Figure 4), except that the number of neurons was varied, with connection probability held fixed, and simulation time was increased from 22 s to 42 s.* **b)** *Average excitatory (top), inhibitory (bottom) and total (middle) synaptic input current to 200 randomly sampled excitatory neurons from the simulations in (a). Synaptic input currents were normalized by capacitance and are therefore reported in units $V/s$.* **c)** *Mean spike count correlation between neuron pairs whose distance is between 0 and 0.1 (purple); between pairs with distance between 0.2 and 0.3 (green); and between pairs chosen randomly at all distances (black). Computed for simulations from the blue curves in (a). Correlations computed from a sample of 5000 randomly selected excitatory neurons.* **d)** *Same as (c), but on a log-log scale. Dashed lines are best fit lines of slope -1.* **e,f)** *Same as (c,d), but for the simulations from the red curves in (a,b).*

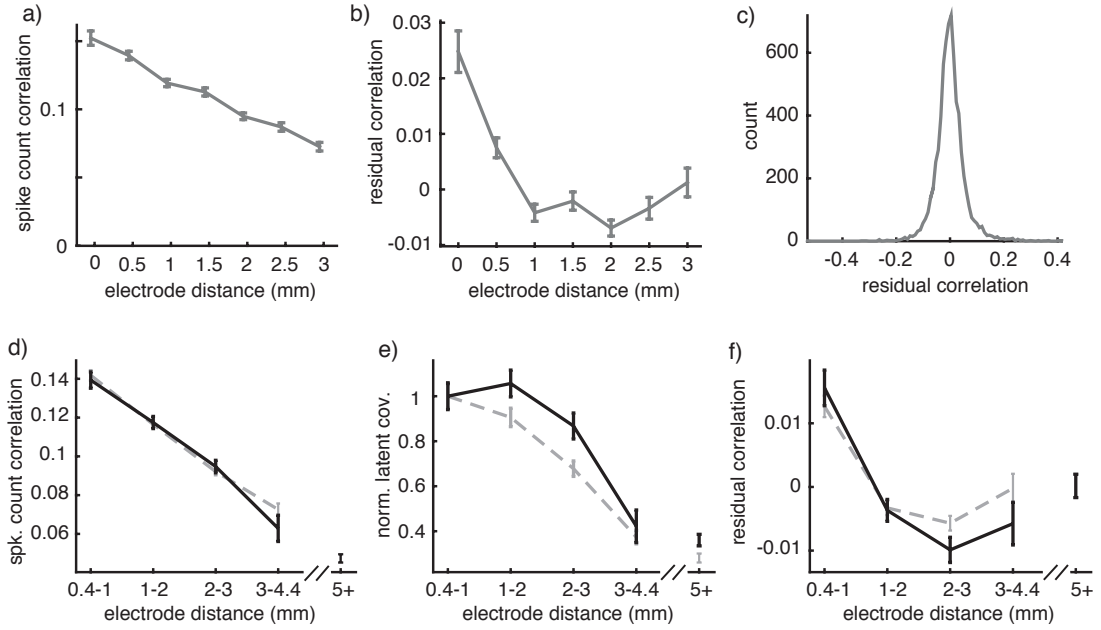
## S.2 Correlations, firing rates and input currents from network simulations at increasing network size

Our mathematical analysis of balanced networks (see above and Results) is asymptotically valid in the limit of large network size (large number of neurons, $N$, and number of synaptic inputs, $K$). To test the convergence of the simulations as $N$ increases, we performed simulations with increasing values of $N$, where connection probabilities are fixed so that $K$ increases proportionally (Supplementary Figure 4).

When recurrent projections are narrower in space than feedforward projections ($\alpha_{\text{rec}} < \alpha_{\text{ffwd}}$, as in Figure 3), the average correlation between nearby or more distant neuron pairs decays to zero as $N$ increases (Supplementary Figure 4a). The decay rate is approximately $\mathcal{O}(1/N)$ (Supplementary Figure 4b). When recurrent projections are broader in space than feedforward projections ($\alpha_{\text{rec}} > \alpha_{\text{ffwd}}$, as in Figure 4), the average correlation between nearby neuron pairs converges toward a positive number, the average correlation between neurons pairs at intermediate distances converges to a negative number, and the average correlation between neurons at all distances converges toward

zero (Supplementary Figure 4c). The decay of the average over all paris is approximately $\mathcal{O}(1/N)$ (Supplementary Figure 4d).

Despite the differences in correlations, both networks produce similar firing rates within a reasonable range that converge to a finite limit for large $N$ (Supplementary Figure 4e; see Section S.1.1 for a derivation of the theoretical rates). The asymptotic balance between excitation and inhibition is demonstrating by plotting the mean excitatory, inhibitory and total synaptic input as a function of $N$ (Supplementary Figure 4f). Note that "excitatory" synaptic inputs include both recurrent and feedforward excitation.




Supplementary Figure 5: **Further statistical analysis of correlations in macaque V1. a,b)** *Same as Figures 5b and 7b respectively, but distances are binned more finely. Correlations decreased from the first to the second bin, from the first to the third bin and from the second to the third bin (all three with $p<10^{-5}$; one-sided unpaired t-test). Correlations increased from the fifth to the seventh bin (p=0.008; one-sided unpaired t-test; t-value=2.4; df=1,704). The increase of correlations over the last three bins is significant under a Bonferroni correction for multiple comparisons over the last three bins (corrected p-value = 0.024). The apparent increase in correlation from the third to fourth bin is not significant (p=0.18; one-sided unpaired t-test; t-value=0.9; df=2,249). Distances 3.5-5mm are not shown since there were few such pairs and distances 5+mm are not shown because the precise distances for such pairs are not known, so they cannot be resolved into 0.5 mm bins (see Experimental Procedures and Figure 7b). Including these data does not alter significance (p-values remain <0.025).* **c)** *Histogram of all residual correlations.* **d,e,f)** *Solid black curves are the same as Figures 5b, 5c and 7b respectively, except we only included data from the two recording sessions in which linear electrodes were present. The increase from the third to the fifth bin was still significant when only including data from these two recording sessions ($p=10^{-4}$; one-sided, unpaired t-test; t-value=3.7; df=1,064). The dashed gray curves are the same as Figures 5b, 5c and 7b respectively.*

## S.3 Further statistical analysis of the dependence of residual correlation on electrode distance in macaque V1

In Figures 5b and 7b, we considered spike count correlations and residual correlations as a function of neuron distance, where distances were binned with a bin with of 1 mm. In Supplementary Figure 5a,b, we show the same data with a finer bin size of 500 $\mu$m. Bin sizes cannot be made much smaller because electrodes were spaced at 400 $\mu$m in the recordings. The non-monotonic dependence of correlations on distance remains statistically significant under this finer spatial binning, even though the increase between consecutive bins is not significant (see figure caption). Finer binning of distance necessarily reduces significance as the change in means between nearby consecutive bins becomes small. This issue points to a weakness in using binned data to test for an increase in residual correlations with distance.

Where t-tests were used, data distribution was assumed to be normal but this was not formally tested. The distribution of residual correlations is plotted in Supplementary Figure 5c.

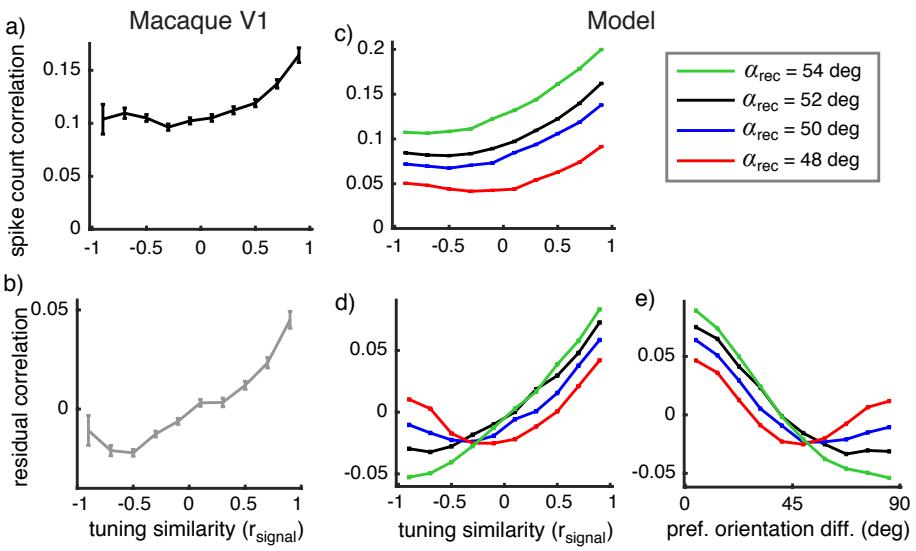
To further test the significance of the increase of residual correlations above 2 mm, while avoiding any potential issues arising from binning, we performed linear regression on all of the unbinned residual correlations with distances greater than or equal to 2 mm. The best fit line had a positive slope (p=0.0043; F-test; ), indicating that residual correlations increase with distance after 2 mm. Performing the same regression analysis on the correlations with distances less than 2 mm yields a negative slope ($p < 10^{-10}$; F-test; F-value=8.2; df=2,966), supporting the prediction from our theoretical model that residual correlations depend non-monotonically on distance.

In Figures 5b, 5c and 7b of the main text, we used data from eight recording sessions. In only two of those sessions, linear electrodes were placed ∼5-10 mm away from the electrode array. Hence, Figures 5b, 5c and 7b contain data averaged over recording sessions with and without the linear electrodes present. To check the effects of this averaging, we repeated the data analysis while only including the data from the two recording sessions for which the linear electrodes were present. We found a similar overall trend in the spike count correlations, latent covariances and residual correlations (Supplementary Figure 5d-f)

## S.4 Correlations as a function of orientation tuning similarity

We focused on the dependence of correlation and connection probability on the physical distance between neurons, but both correlations and connection probabilities are also known to depend on the tuning properties of neurons [8, 3]. The correlations between L2/3 neurons in our data set increased with their tuning similarity (Supplementary Figure 6a), as measured by the Pearson correlation coefficient between their tuning curves, $r_{signal}$ [8]. We next measured the average residual correlation as a function of tuning similarity (Supplementary Figure 6b). While the mean residual correlation depended non-monotonically on tuning similarity, this non-monotonicity was not statistically significant ($p = 0.051$ for the decrease from first bin to the third bin in Supplementary Figure 6b).

Our theoretical results do not apply as directly to tuning similarity as they do to physical distance because many of the recorded neurons are complex cells with multi-modal tuning curves. However, a simplified model is obtained by assigning a preferred orientation between 0° and 180°




Supplementary Figure 6: **Correlations as a function of tuning similarity. a)** *Mean spike count correlation and* **b)** *residual correlation between putative L2/3 neurons in macaque primary cortex as a function of tuning similarity. Same as Figures 5b and 7b respectively, but correlations are partitioned by tuning similarity instead of distance.* **c,d)** *Same as (a,b), but from a computational model where* $\alpha_{ffwd} = 20°$ *and for different values of* $\alpha_{rec}$ *(see legend).* **e)** *Residual correlation as a function of the difference between preferred orientations in the model. All plots show mean* $\pm$ *SEM. The non-monotonicity in (b) is not significant* ($p = 0.051$*; one-sided unpaired t-test between first and third bins; t-value=1.6; df=872).*

to each neuron, then letting connection probability depend on the difference between preferred orientations. This model is identical to the spatial model considered above except the network is on a one-dimensional state space, instead of the two-dimensional space, $\Gamma$. All of the calculations and equations above are identical except that there is just one spatial Fourier mode, $n$, in place of the vector $\mathbf{n} = (n_1, n_2)$.

The decay of inter- and intra-laminar connection probability with orientation tuning difference has not been measured in macaque V1 to our knowledge. Therefore, in contrast to our spatial model of V1 where $\alpha_{\text{rec}}$ and $\alpha_{\text{ffwd}}$ were constrained by anatomical measurements, we manually chose values of $\alpha_{\text{rec}}$ and $\alpha_{\text{ffwd}}$ to capture the dependence of residual correlation on tuning similarity that we observed in our data (Supplementary Figure 6c-e). Since the correlation data were used to choose the parameters of the model, however, these simulations should not be interpreted as making a prediction about the dependence of residual correlations on tuning similarity.

The connection between residual correlation structure and the connectivity parameters, $\alpha_{\text{rec}}$ and $\alpha_{\text{ffwd}}$, is further complicated by the limited range of orientation tuning space. Specifically, since orientations are between $0°$ and $180°$, the difference between neurons' preferred orientations cannot be larger than $90°$. If $\alpha_{\text{rec}}$ is sufficiently large, any potential non-monotonicity will be pushed beyond this range and residual correlations will depend monotonically on tuning similarity as a result, even when $\alpha_{\text{rec}} > \alpha_{\text{ffwd}}$ (Supplementary Figure 6d,e). Hence, a non-monotonic dependence of residual correlation on tuning similarity could only be predicted from our model if $\alpha_{\text{rec}}$ and $\alpha_{\text{ffwd}}$ were measured to find that $\alpha_{\text{rec}} > \alpha_{\text{ffwd}}$ with $\alpha_{\text{rec}}$ not too large.

In summary, our findings concerning the dependence of residual correlation on tuning similarity

in macaque V1 are inconclusive since the non-monotonicity in Supplementary Figure 6b is not significant and since the relevant anatomical parameters are unknown. Further experiments are needed to clarify the dependence of correlations on tuning similarity.

## S.5 Description of algorithm to generate spatially extended network architecture

For the spatially extended network simulations in Figures 3, 4, 6 and 7, connections were formed randomly with connection probability that depends on distance, as described in Methods. Here, we provide a more detailed description of the algorithm used to generate the connectivity.

To select one postsynaptic target in population $a = \mathrm{e}, \mathrm{i}$ for a presynaptic neuron in population $b = \mathrm{e}, \mathrm{i}, F$ at coordinates $(y_1, y_2)$, we first generated a pair of numbers, $z_1$ and $z_2$, independently from an unbiased Gaussian distribution with standard deviation $\alpha_b$. These random numbers represent the distance in each direction from the presynaptic neuron to the postsynaptic target. The first and second coordinates of the postsynaptic target location were then set to $x_1 = \mathrm{mod}(z_1 + y_1, 1)$ and $x_2 = \mathrm{mod}(z_2 + y_2, 1)$. This assures that the coordinates were between 0 and 1 and that the distance from the presynaptic neuron, measured periodically on $\Gamma$, was $z_1$ and $z_2$ in each direction. This implies that the probability density function of $x_1$ and of $x_2$ is a wrapped Gaussian distribution,

$$g(x; \alpha_b^2) = \frac{1}{\sqrt{2\pi}\alpha_b} \sum_{k=-\infty}^{\infty} e^{-(x+k)^2/(2\alpha_b^2)}$$

To find a postsynaptic neuron from population $a$ near the target location, we set $j_1 = \mathrm{round}(N_a x_1)$ and $j_2 = \mathrm{round}(N_a x_2)$ then set the index of the postsynaptic neuron in population $a$ to $j = N_a j_1 + j_2$. We repeated this procedure to generate a fixed number, $K_{ab}^{\mathrm{out}}$, of postsynaptic targets in population $a$ for each presynaptic neuron in population $b$. Thus, $K_{ab}^{\mathrm{out}}$ is the number of outgoing synaptic projections from each neuron in population $b$ to all neurons in population $a$. We followed this procedure for all pairings of pre- and post-synaptic populations, $b = \mathrm{e}, \mathrm{i}, F$ and $a = \mathrm{e}, \mathrm{i}$. The values of $\alpha_a$ and $K_{ab}^{\mathrm{out}}$ used in each figure are given in Methods.

## References


[1] B Doiron, A Litwin-Kumar, R Rosenbaum, G K Ocker, and K Josić. The mechanics of state-dependent neural correlations. *Nature Neuroscience*, 19(3):383–393, 2016.

[2] C W Gardiner. *Handbook of stochastic methods*, volume 3. Springer, Berlin, 1985.

[3] 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(7345):87–91, May 2011.

[4] V Pernice, B Staude, S Cardanobile, and S Rotter. How structure determines correlations in neuronal networks. *PLoS Comput. Biol.*, 7(5):e1002059, may 2011.

[5] 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(5965):587–590, 2010.

[6] R Rosenbaum and B Doiron. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Phys Rev X*, 4(2):021039, 2014.

[7] 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(10):108102, March 2008.

[8] M A Smith and A Kohn. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci*, 28(48):12591–603, 2008.

[9] 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(9):2133–84, 2008.

[10] F G Tricomi. *Integral equations*. Interscience, New York, 1957.

[11] 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(3):e1002408, 2012.

[12] C van Vreeswijk and H Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Comput*, 10(6):1321–1371, 1998.

[13] A.M. Yaglom. *An introduction to the theory of stationary random functions*. Dover, New York, 2004.