PNAS RESEARCH ARTICLE | NEUROSCIENCE

The mechanics of correlated variability in segregated cortical excitatory subnetworks

Alex Negrón[a,b,1], Matthew P. Getz[b,c,d,1,2], Gregory Handy[b,c,d,3,4], and Brent Doiron[b,c,d,3,5]

Affiliations are included on p. 11.

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received April 25, 2023; accepted April 3, 2024

Understanding the genesis of shared trial-to-trial variability in neuronal population activity within the sensory cortex is critical to uncovering the biological basis of information processing in the brain. Shared variability is often a reflection of the structure of cortical connectivity since it likely arises, in part, from local circuit inputs. A series of experiments from segregated networks of (excitatory) pyramidal neurons in the mouse primary visual cortex challenge this view. Specifically, the across-network correlations were found to be larger than predicted given the known weak cross-network connectivity. We aim to uncover the circuit mechanisms responsible for these enhanced correlations through biologically motivated cortical circuit models. Our central finding is that coupling each excitatory subpopulation with a specific inhibitory subpopulation provides the most robust network-intrinsic solution in shaping these enhanced correlations. This result argues for the existence of excitatory–inhibitory functional assemblies in early sensory areas which mirror not just response properties but also connectivity between pyramidal cells. Furthermore, our findings provide theoretical support for recent experimental observations showing that cortical inhibition forms structural and functional subnetworks with excitatory cells, in contrast to the classical view that inhibition is a nonspecific blanket suppression of local excitation.

excitatory–inhibitory assemblies | trial-to-trial variability | inhibition-stabilized network | cortical connectivity

Significance

The structure of recurrent connectivity within cortical networks has important implications for their activity. Previous work has found neurons preferentially interconnect to form clustered assemblies. Traditionally, such assemblies of neurons showed strong, positive within-population correlations and strong, negative cross-population correlations. Our work is motivated by recent experimental results that stand in stark contrast to these observations. Specifically, it was found that neurons in the mouse visual cortex exhibited highly correlated activity but with a small probability of connection. Using theoretical analyses, we find that the most robust solution involves inhibitory neurons being equally segregated in their interactions with excitatory neurons. That is, inhibition should strongly cocluster with excitation, a result that aligns with recent experimental observations.

Determining a structure–function relationship in a cortical circuit is a central goal in many neuroscience research programs. While the trial-averaged responses of a network to a fixed stimulus or repeated behavior give some information about the underlying circuit, the dynamic or trial-to-trial fluctuations of neuronal activity provide another important glimpse into how network structure determines network activity (1). Indeed, any correlations in the trial-to-trial fluctuations of a pair of neurons are thought to reflect, in part, their common synaptic inputs (2–6). Because correlated response variability is a salient feature of cortical responses (7), incorporating it into cortical models can offer an important constraint when choosing model parameters that allow model responses to best match those from experiment. This framework has been typically used to explore local circuit structure where pyramidal neurons are only labeled by their stimulus tuning. In this study, we use established circuit-based theories to make targeted predictions about cortical circuits where pyramidal neurons are distinguished by the foci of their synaptic projections.

Our modeling is heavily motivated by a series of studies that measure the pairwise trial-to-trial covariability of pyramidal neuron activity in layer 2/3 of the mouse primary visual cortex (V1). In a heroic set of combined in vivo and in vitro experiments, Ko et al. (8) and Cossell et al. (9) report that the magnitude of the pairwise correlations between two pyramidal cells (both trial-averaged and trial-variable) increases with their probability of synaptic connection. This observation is consistent with a Hebbian network where synaptic wiring is functionally aligned; i.e., neuron pairs showing coordinated activity (through tuning or trial-to-trial fluctuations) have stronger connections. However, these same researchers later investigated the functional properties of two distinct subpopulations of pyramidal cells in mouse V1 that project to separate downstream higher visual areas (10). These subpopulations are interconnected with lower probability than that of randomly sampled pyramidal cells within V1. Despite this weak connectivity, it was found that the correlations between these distinct subpopulations were much higher than predicted by their sparse interconnectivity. In fact, the magnitude of the correlated variability across the two subpopulations was comparable to the correlation between any randomly chosen pair of excitatory neurons.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]A.N. and M.P.G. contributed equally to this work.

[2]Present address: School of Life Sciences, Technical University of Munich, Freising 85354, Germany.

[3]G.H. and B.D. contributed equally to this work.

[4]Present address: School of Mathematics, University of Minnesota, Minneapolis, MN 55455.

[5]To whom correspondence may be addressed. Email: bdoiron@uchicago.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2306800121/-/DCSupplemental.

Published July 3, 2024.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

In this same vein, another experiment examining callosal projection neurons in mouse V1 found that these cells also cluster and connect more strongly as a class (11), yet their correlated variability is similar when comparing within-class and out-of-class neuron pairs. In total, these data illustrate that significant, positive correlations can persist in the absence of direct strong connections.

In this work, we apply established circuit modeling techniques (3–5, 12–14) to characterize the neural circuit properties which could explain the significant positive shared variability across segregated cortical subpopulations observed in Kim et al. (10). With the use of mean field circuit models, we show that population variability depends on the dynamical regime of the circuit and relies on how recurrent inhibition aligns with the functional segregation of pyramidal neurons. In a weakly coupled regime, correlations can be characterized through inheritance from outside sources, or increased through shared inhibitory inputs. By contrast, in a strongly coupled regime, shared inhibition largely acts to anticorrelate activity across the populations. Critically, we show that this anticorrelation can be mitigated if inhibition is similarly clustered with excitation, forming instead functionally defined excitatory–inhibitory assemblies. This regime of strongly coupled dynamics with clustered inhibition provides the most robust solution space to explain the elevated correlations reported in Kim et al. (10), without a dependence on inherited positive correlations from outside sources.

Our model prediction claims that inhibitory interactions with excitation should be as precise and selective as excitatory–excitatory interactions. Experimental validation would likely involve a combination of careful tracing and patch experiments, however recent experimental results support a looser version of our conclusions, having observed that murine visual cortex exists in an inhibition-stabilized regime (15) and that excitation and inhibition form functional clusters (16). Our modeling results highlight the fact—which has been growing in recognition recently—that to understand neural processing inhibition must be considered as much a key component as excitation.

Results

Segregated Synaptic Wiring Does Not Produce Segregated Functional Responses. Our work is motivated by an apparent inconsistency in a series of experimental studies exploring the relation between the recurrent circuitry and functional responses of neuronal populations in the sensory neocortex. Ko et al. (8) and Cossell et al. (9) used a combination of in vivo population imaging and in vitro electrophysiology to show that the activity correlations between pairs of pyramidal neurons in the mouse primary V1 increase monotonically with the probability of there existing synaptic connections between them. Later work from the same group (10) investigated two excitatory populations in mouse V1: neurons that are either anterolateral (AL)- or posteromedial (PM)-projecting. Despite being in close spatial proximity to each other, these neuronal subpopulations exhibit high within-group connectivity (prob. AL $\leftrightarrow$ AL connection $\sim$ 0.21, prob. PM $\leftrightarrow$ PM connection $\sim$ 0.18) and low between-group connectivity (prob. AL $\rightarrow$ PM connection $\sim$ 0.04, prob. PM $\rightarrow$ AL connection $\sim$ 0.05). To streamline our presentation we will label these two populations E_1 and E_2 (Fig. 1*A*). Given the low connection probability between E_1 and E_2 and the established relation between connectivity and activity correlations shown in Ko et al. (8) and Cossell et al. (9), one would predict that the degree of correlations between the activities of E_1 and E_2 would be negative (Fig. 1*B*, held out light purple square; from ref. 10, we estimate this value to lie in an interval approximately [−0.05,0]). However, Kim et al. (10) reported substantially higher than predicted mean $E_1 - E_2$ correlations (Fig. 1*B*, darker green off-diagonal squares; Kim et al. (10)

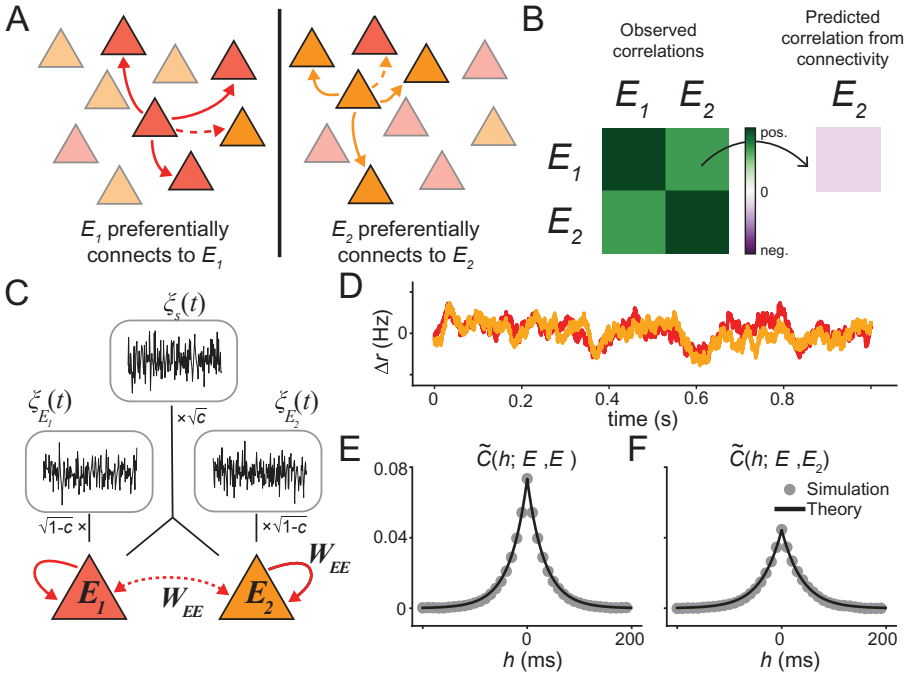

Fig. 1. Mean field model of segregated *E* populations. (*A*) Illustration of experimentally observed connectivity motif; the red (E_1) and orange (E_2) populations interconnect with lower probability than average. (*B*) Schematic of main experimental observations: $E_1 - E_2$ correlations were higher than would be predicted from their low connectivity. (*C*) Model schematic. Black traces and arrows denote noise sources. Red arrows indicate excitatory recurrent connections, where the dashed line connotes weakened connection strength. Feedforward stimulus drive omitted for clarity. (*D*) Example realization of network activity to a sustained, fixed stimulus. Colors as in (*A*). (*E*) E_1 autocorrelation function and (*F*) $E_1 - E_2$ cross-correlation function for the illustrated rate traces. For panels *D*–*F*: $c = 0.5$.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

measured it to be about 0.027). Additionally, these correlations were observed to lie close to the within-group correlation (e.g., $E_1 - E_1$) values (Fig. 1*B*, dark green diagonal squares; Kim et al. (10) reported approximately 0.035 to 0.04). In total, while pyramidal neurons in mouse V1 projecting to distinct targets show segregated synaptic connectivity, the degree of functional segregation between these subpopulations is below what is expected.

The central goal of our study is to put forth a circuit-based model framework that can robustly and self-consistently account for both of these experimental observations. It is important to note that Kim et al. (10) only considered total correlations (of the raw neural activity traces) in computing this expected correlation value. However, given the similarities observed in the signal and noise correlation structure in both this and previous studies (8, 10, 11), we focus here on noise correlations which relate more directly to the underlying structure of connectivity (3).

A Circuit Model of Fluctuations in Segregated Subpopulations. To study the structure of correlations in anatomically segregated networks and investigate the possible mechanisms responsible for the unexpectedly enhanced correlations between these subnetworks, we consider a phenomenological dynamic mean field model for the aggregate activity of each neural population (12, 17, 18). Assuming that the network has a steady-state solution ($\mathbf{r}_{ss}$), the linearized dynamics of population A around this equilibrium are given by (see *Materials and Methods* for additional details):

$$\tau_A \frac{d\Delta r_A}{dt} = -\Delta r_A + \sum_B W_{AB}\Delta r_B + \sigma_A\left[\sqrt{1-c}\cdot\xi_A(t) + \sqrt{c}\cdot\xi_S(t)\right], \quad [1]$$

where $\Delta r_A = r_A - r_{ss,A}$, τ_A is a time constant, and W_{AB} is the effective strength of connections from population B to A. For the purely excitatory network, A and B range over E_1 and E_2; when inhibitory connections are included in later sections, A and B will include those as well. The stochastic processes $\xi_A(t)$ and $\xi_S(t)$ represent private and shared global fluctuations, respectively, modeling stochastic inputs that are external to the network. $\xi_A(t)$ and $\xi_S(t)$ are taken to be independent Gaussian processes with $\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(t')\rangle = \delta(t-t')$. The parameter $c \in [0, 1]$ scales the proportion of shared noise relative to private noise (Fig. 1*C*) while $\sigma_A > 0$ represents the total intensity of the external fluctuations given to population A.

We make two assumptions: 1) the network has a stable solution r_{ss} about which the population dynamics fluctuate (Fig. 1*D*), and 2) connections within and inputs to the network are symmetric across the two E populations, with $W_{E_1E_1} = W_{E_2E_2} = W_{EE}$, $W_{E_1E_2} = W_{E_2E_1} = \alpha W_{EE}$, and $\sigma_{E_1} = \sigma_{E_2} = \sigma$. Note that parameter $0 < \alpha \ll 1$ represents the degree to which the interpopulation connections are weaker than the within-population connections (Fig. 1*C*). Since the system of recurrently coupled stochastic differential equations in Eq. **1** is a multidimensional Ornstein–Uhlenbeck (OU) process, one can derive (*Materials and Methods*) an analytical formula for its stationary autocovariance function

$$\widetilde{\mathbf{C}}(h) = \langle\Delta\mathbf{r}(t), \Delta\mathbf{r}(t+h)\rangle,$$

which agrees well with numerical simulations (Fig. 1 *E* and *F*). Further, and of particular interest in this work, is the long-time covariance matrix defined as

$$\mathbf{C} := \int_{-\infty}^{\infty} \widetilde{\mathbf{C}}(h)dh.$$

This may be expressed (*Materials and Methods*) as

$$\mathbf{C} = (\mathbf{I}-\mathbf{W})^{-1}\mathbf{D}\left[(\mathbf{I}-\mathbf{W})^{-1}\mathbf{D}\right]^{\top}, \quad [2]$$

where $\mathbf{W}$ is a matrix of effective connection strengths and $\mathbf{D}$ is a matrix that scales the input fluctuations. We define the correlations between E_1 and E_2 as

$$\mathrm{Corr}(E_1, E_2) := \frac{C_{E_1E_2}}{\sqrt{C_{E_1E_1}C_{E_2E_2}}} = \frac{C_{E_1E_2}}{C_{E_1E_1}}, \quad [3]$$

where C_{AB} is an element of $\mathbf{C}$ and the second equality follows by the assumed symmetry in the system. This framework enables us to formalize the motivating question of our study: what are the mechanisms that enable higher than expected correlations across anatomically segregated populations? For the sake of specificity, we choose the threshold $\mathrm{Corr}(E_1, E_2) > 0.6$ as an approximation of the ratio of mean across-population to within-population noise correlations in Kim et al. (10).

Inheritance Model of Correlations between Weakly Coupled Excitatory Populations. We begin by exploring how the strength of recurrent excitation (W_{EE}) and the proportion of fluctuations that are shared (c) shape correlations between the segregated E populations. In this section, to ensure that the network admits a stable activity solution we require $W_{EE} < 1$, else recurrent excitation would lead to runaway activity. Note that while we allow W_{EE} to vary, we maintain segregated populations by keeping α small and fixed. We find that while increasing W_{EE} leads to moderate increases in $\mathrm{Corr}(E_1, E_2)$ (Eq. **3**), a much more significant increase occurs by increasing c (Fig. 2*A*).

To better understand the underlying mechanisms responsible for these higher correlations within this parameter regime (i.e., to the right of the pink line in Fig. 2*A*), we perform a pathway expansion of the covariance matrix Eq. **2**. Since the steady state emitted by the system in Eq. **1** is stable, the term $(\mathbf{I}-\mathbf{W})^{-1}$ can be expanded as a series. This allows us to write Eq. **2** as (*Materials and Methods*)

$$\mathbf{C} = \sum_{n=0}^{\infty}\left[\sum_{i=0}^{n}\mathbf{W}^{n-i}\mathbf{D}\mathbf{D}^{\top}\left(\mathbf{W}^{\top}\right)^{i}\right], \quad [4]$$

where each term in the inner sum corresponds to an nth-order path through the network. Writing out the first three terms of this sum for the cross-covariance yields

$$C_{E_1E_2} = \sigma^2\left[c + (2c+2\alpha)W_{EE} + (3(1+\alpha^2)c + 6\alpha)W_{EE}^2\right] + \mathcal{O}(W_{EE}^3).$$

Rewriting this equation as

$$C_{E_1E_2} = \sigma^2[\underbrace{c\cdot(1+2W_{EE}+3(1+\alpha^2)W_{EE}^2)}_{(1)} + \underbrace{\alpha\cdot(2W_{EE}+6W_{EE}^2)}_{(2)}] + \mathcal{O}(W_{EE}^3), \quad [5]$$

reveals that each term contributing to this cross-covariance can be thought of as arising from one of two sources: 1) inherited from

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

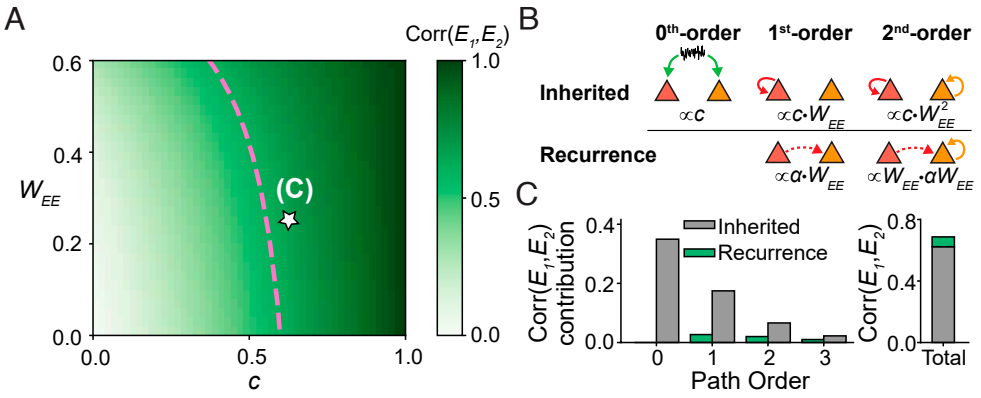

Fig. 2. Highly correlated regime in weakly coupled excitatory network relies on correlated feedforward inputs. (*A*) Corr(E_1, E_2) as a function of W_{EE} and the magnitude of shared input noise c. The dashed pink line indicates Corr$(E_1, E_2) = 0.6$, approximating the value reported in ref. 10. (*B*) Schematic of example synaptic paths through the network, along with their contribution to the cross-covariance, relating to the path expansion Eq. **4**. The inherited row refers to correlated paths stemming from correlations in the feedforward input, while the recurrence row arises from the recurrent connections across the populations. (*C*) Contributions of paths of given order to networks (*Left*) and the total correlation (*Right*) for the parameters $W_{EE} = 0.25$ and $c = 0.65$ (star from panel *A*). All panels: $\alpha = 0.1$.

the shared correlated input and dependent on the parameter c (Fig. 2 *B*, *Top*), and 2) purely arising from the recurrent connections and dependent on the parameter α (Fig. 2 *B*, *Bottom*). We note that the "propagation" of the inherited contribution to higher-order paths does not only rely on the $E_1 \leftrightarrow E_2$ connections (proportional to $\alpha^n c$). This is because the correlated activity is fed directly into each subpopulation at the 0th order, from which it can propagate into higher-order paths via self-loops contained within each population. We emphasize that eliminating the 0th order term (i.e., setting $c = 0$) eliminates all contributions from the inherited global source.

We now utilize this pathway expansion to compare the contributions from feedforward and recurrent mechanisms to the net cross-covariance for an example point lying in the highly correlated regime (Fig. 2*A*, star; $W_{EE} = 0.25$ and $c = 0.65$). We first note that this series converges quickly and only a few paths significantly contribute to the total correlation (Fig. 2*C*). The convergence of this series depends directly on the largest eigenvalue of $\mathbf{W}$ (*Materials and Methods*), namely

$$\lambda_{\max} = W_{EE} \cdot (1 + \alpha),$$

which is small for our choice of parameters. Our numerical results also illustrate that the contribution from the inherited source largely dominates at each order (Fig. 2 *C*, *Left*), and contributes ~90% of the total cross-correlation (Fig. 2 *C*, *Right*). These results hold qualitatively across this parameter regime, and lead us to conclude that it corresponds to a model in which large shared input fluctuations explain the heightened correlations between the separate E populations. Taken together, we characterize this solution which exhibits enhanced $E_1 - E_2$ correlations as a feedforward inheritance model.

However, under the condition where the shared input fluctuations are small, we still lack a potential mechanism for significant positive correlations. To surmount this shortcoming, we first need to extend our model to also include inhibitory populations.

Weak Recurrent Excitation with Global Inhibition. Parsimoniously, we begin by modeling inhibition as a single global population, consistent with observations that inhibition simply connects densely and nonspecifically within the cortex (19, 20) (Fig. 3*A*). To understand the effect of inhibition in this circuit, we explore how the strength of recurrent inhibitory connections ($W_{EI} < 0$ and $W_{IE} > 0$) shape correlations between the excitatory populations in the case when $c = 0$. Assuming W_{EE} remains weak (i.e., $W_{EE} < 1$), we find a large portion of the parameter regime yields negative cross-correlations (Fig. 3*B*; purple region). However, there is a region that satisfies our correlation condition, namely the dark green region that corresponds to strong $I \to E$ and weak $E \to I$ connections.

We again make use of a pathway expansion of Eq. **2** to help decipher this observation, this time accounting for the new inhibitory pathways (Fig. 3*C*). Writing out the expansion to second order in W yields

$$C_{E_1E_2} = \sigma^2[\underbrace{2\alpha W_{EE} + 6\alpha W_{EE}^2}_{\text{exc. paths}} + \underbrace{2W_{EI}W_{IE} + W_{EI}^2}_{\text{inh. paths}}] + \mathcal{O}(W^3), \quad [6]$$

where we have noted the terms involving only the excitatory components and terms which involve paths through the inhibitory population. We first observe that contributions to the cross-covariance due to the excitatory subnetwork at each order are the same as the previous network without the inhibitory connections (Eq. **5** for $c = 0$). This leads us to decompose the total covariance into an excitatory component and an inhibitory component (neglecting the $\mathcal{O}(W^3)$ terms in Eq. **6**)

$$C_{E_1E_2} = C^{\text{exc}}_{E_1E_2} + C^{\text{inh}}_{E_1E_2}. \quad [7]$$

As Eq. **6** suggests, depending on the strength of the underlying inhibitory connections, $C^{\text{inh}}_{E_1E_2}$ can either be positive (positively correlating the excitatory subpopulations; Fig. 3*B*, along W_{EI} axis) or negative (anticorrelating the subpopulations; Fig. 3*B*, purple region). By contrast, $C^{\text{exc}}_{E_1E_2}$ is clearly bounded below by zero.

Specifically, Eq. **6** reveals a "tug of war" that can arise early on in the pathway expansion between the $E \to I \to E$ (i.e., $W_{EI}W_{IE} < 0$) and the $I \to E$ (i.e., $W_{EI}^2 > 0$) inhibitory pathways. Choosing $|W_{EI}| > W_{IE} \approx 0$, we find that the positive term dominates, and the inhibitory population acts as a strong correlator of excitatory activity (Fig. 3*D*). We term this an inhibitory inheritance model by analogy to the feedforward inheritance model described above.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

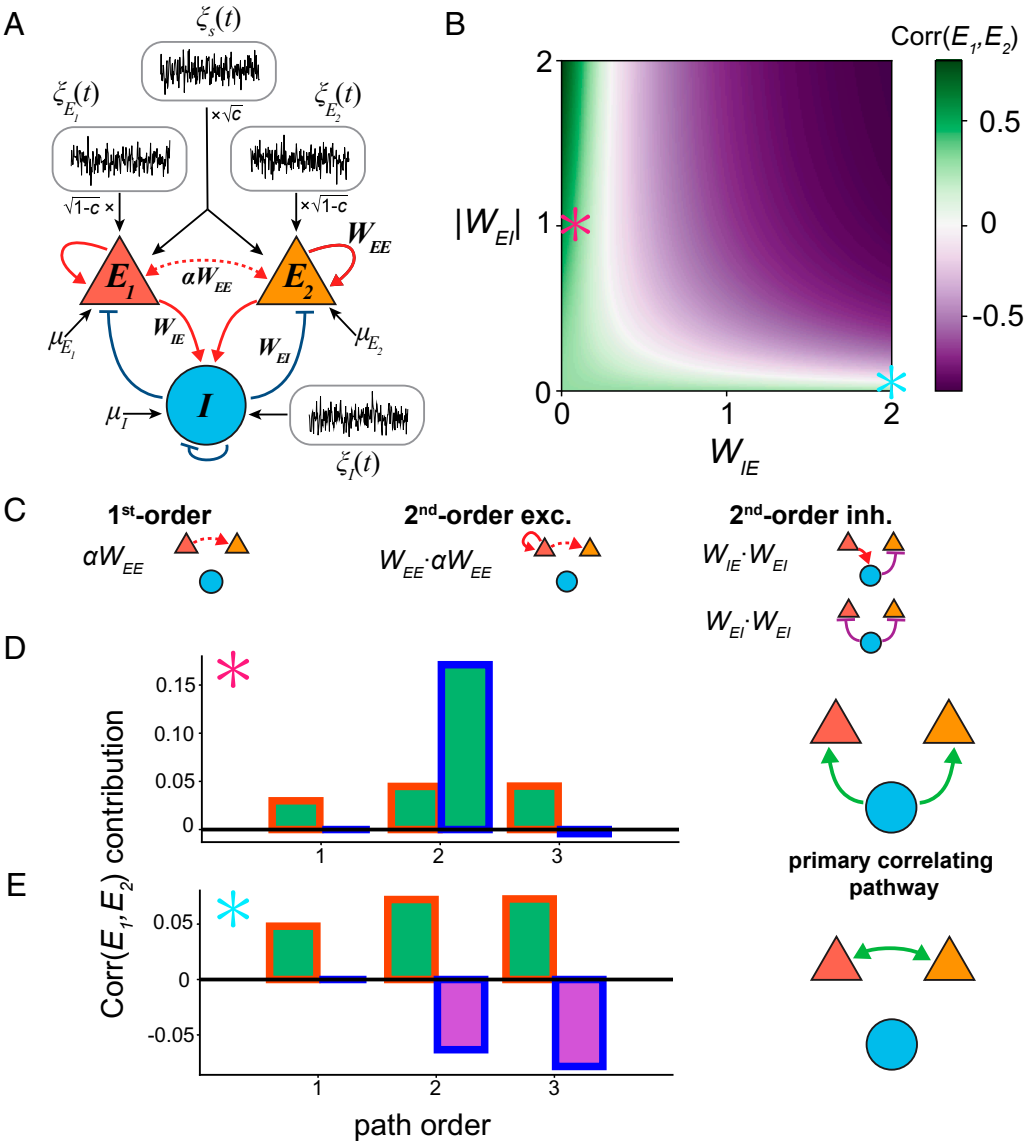

Fig. 3. Weakly coupled network. (*A*) Network model schematic as in Fig. 1*C*. Blue lines indicate recurrent inhibitory connections. (*B*) Corr(E_1, E_2) as a function of $|W_{EI}|$ and W_{IE}. (*C*) Illustrations of first- and second-order paths. (*D* and *E*) (*Left*) Contributions of *E* (red outlined bars) and *I* (blue outlined bars) to the net Corr(E_1, E_2). (*Right*) Schematic of dominant correlating pathway. Colored stars denote locations in *B*. Red star: $W_{EI} = -1, W_{IE} = 0.07$; blue star: $W_{EI} = -0.05, W_{IE} = 2$. For all panels $\alpha = 0.15$.

On the other hand, when $W_{IE} > |W_{EI}| \approx 0$, the negative term dominates, leading the inhibitory population to weaken the strength of cross-correlations. In this case, the primary correlating source across the excitatory populations is the weak $E_1 \leftrightarrow E_2$ connections (Fig. 3*E*). But as we noted previously (Fig. 2), this pathway alone is incapable of yielding high cross-correlations without strongly correlated feedforward input.

The regime of weakly coupled neural populations thus permits two solutions for correlating E_1 and E_2 to a sufficiently high degree, both of which can be characterized in terms of inheritance models. Namely, enhanced positive correlations can be inherited from outside sources or from local recurrent inhibition. Nevertheless, the ambiguity in the former solution and the fine-tuning required to achieve the latter solution push us to uncover a more robust mechanism.

Strong Recurrent Excitation with Global Inhibition. Up to this point, our assumption that the recurrent excitatory coupling is weak ensured that the stability of the equilibrium point was independent of the inhibitory currents. Such a network is commonly referred to as a non-inhibition-stabilized network (non-ISN) (21–23) (see *SI Appendix* for additional details). However, recent experimental evidence suggests that the mouse cortex operates in the ISN regime, where strong recurrent excitation is tracked and balanced by strong inhibitory feedback (15, 24). Since the ISN regime is known to exhibit sometimes perplexing dynamics, such as the well-studied paradoxical effect where a depolarizing input to inhibitory neurons results in a lowering of their firing rate (21, 22), it is initially unclear how shifting into this parameter regime will shape the $E_1 - E_2$ correlations.

We now strengthen the recurrent excitatory connections W_{EE} such that our model network lies in the ISN regime. Performing a similar analysis as before (i.e., fixing W_{EE} and W_{II}, while varying W_{EI} and W_{IE}) and assuming that the feedforward inputs are uncorrelated ($c = 0$), we find results that at first glance appear familiar (Fig. 4*A*). Namely, a portion of the parameter regime results in negative correlations (purple region), with a narrow parameter regime yielding positive correlations (green region).

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

However, unlike the previous network, these correlations are much larger across this band of parameter values, approaching unity as the system loses stability due to the inhibitory feedback becoming too weak to be able to balance out the strong excitation (gray and red-hatched region).

Unlike the non-ISN regime, where the weak recurrent excitatory connections corresponded with small eigenvalues and quick convergence in our path-expansion, here the eigenvalues of the system lie much closer to the boundary separating stability from instability, with a subset of parameter values producing a spectral radius of $\mathbf{W}$ greater than one. As a result, even when the series in Eq. **4** converges, it does so much more slowly and requires many more terms than before, complicating its interpretation. Instead, we seek to understand the mechanism driving these high correlations calculating directly them only with Eq. **3** and by exploring their apparent connection to the system's stability.

We start by considering the slice of the parameter space where $|W_{EI}| = W_{IE}$ that captures the system's transitions from negative correlations to positive correlations and then to instability (Fig. 4*A*, yellow line; Fig. 4 *B*, *Top*). Analysis of the eigenvalues of $\mathbf{W}$ reveals a pair of eigenvalues (λ_1 and λ_2) that depend on the strength of inhibitory connections and another eigenvalue that remains constant (and close to one) along this parameter slice ($\lambda_3 = W_{EE}(1-\alpha)$) (Fig. 4 *B*, *Bottom*). Interestingly, we find that decay for the stationary autocovariance function for the inhibitory population (Fig. 4 *D* and *E*, *Bottom*; see Eq. **9** in *Materials and Methods*) is well approximated by

$$\psi = \max(\mathrm{Re}(\lambda_1), \mathrm{Re}(\lambda_2)) .$$

From this link, we see that when $|W_{EI}| = W_{IE}$ is large, then ψ is small, meaning the timescale of inhibition is fast. This allows the inhibitory population to rapidly and effectively cancel the net excitatory inputs (Fig. 4 *C*, *Top*). We observe that in this parameter regime, Δr_I remains small while $\Delta r_{E_1} \approx -\Delta r_{E_2}$, leading to strong negative correlations between E_1 and E_2. As $|W_{EI}| = W_{IE}$ decreases, ψ increases toward one, which slows down the inhibitory timescale (Fig. 4 *D*, *Top*). This slower cancelation of the excitatory currents allows for larger deviations away from baseline for all neuronal populations. However, since the system is still stable, we observe that the populations covary together, leading to correlated excursions in the rates.

In total, the ISN regime yielded a more robust set of parameter values corresponding to high correlations across the segregated excitatory populations than the non-ISN regime observed previously. However, even in this improved scenario, the viable parameter regime is still limited to a relatively thin

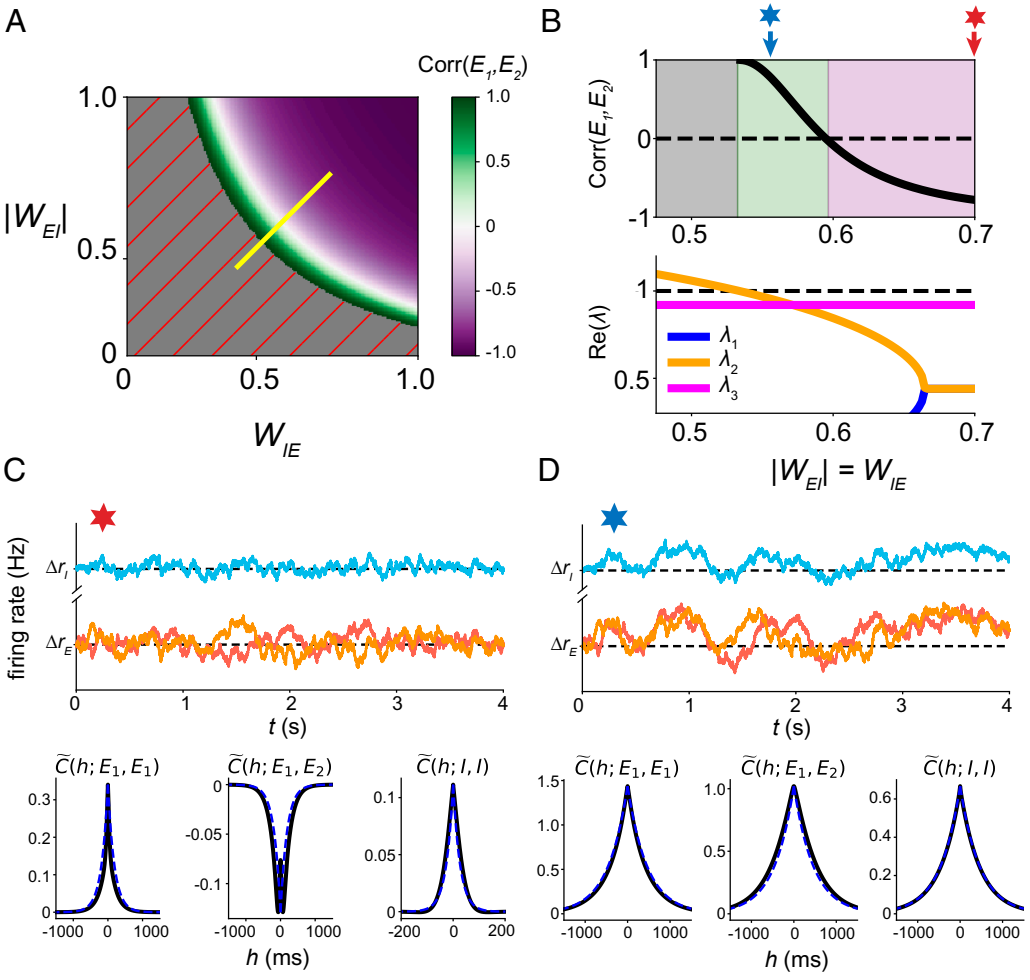

Fig. 4. Global inhibition in ISN regime. (*A*) Corr(E_1, E_2) as a function of W_{EI}, W_{IE} with $c = 0$. (*B*) *Top*: Corr(E_1, E_2) along the yellow path in *A*. Gray region: unstable; green region: positive correlations; purple region: negative correlations. *Bottom*: eigenvalues of the circuit along the yellow path in *A*. (*C* and *D*) *Top*: example rate traces (colors as in Fig. 3*B*). *Bottom*: auto- and cross-correlation functions computed numerically (black) and theoretically for the dominant timescale (blue dashed). Stars indicate parameter values shown in *B*. Here, $\alpha = 0.2$.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

band, and further, this band lies precariously close to regions of instability.

Strong Recurrent Excitation with Clustered Inhibition. The fine-tuning required to capture large $\mathrm{Corr}(E_1, E_2)$ despite having weak $E_1 \leftrightarrow E_2$ coupling ($\alpha \ll 1$) for both the purely excitatory and global inhibitory networks places doubt on these mechanisms being operative in real neuronal circuits. In this section, we hypothesize that a larger stable region of $\mathrm{Corr}(E_1, E_2) > 0$ may be permitted if the sources of inhibition for each excitatory subpopulation are similarly clustered. The intuition is that clustered inhibition will limit the effects of the anticorrelating $E_1 \to I \to E_2$ and $E_2 \to I \to E_1$ pathways.

We implemented inhibition that is coclustered with the excitatory subpopulations by separating the inhibitory population into two subpopulations with I_1 and I_2 corresponding to the respective excitatory populations E_1 and E_2 (Fig. 5*A*). In this case, each E_i/I_i cluster constitutes an ISN ($i = 1, 2$). At this point, the model contains no interpopulation connections except those between E_1 and E_2 and assumes no source of shared input correlations ($c = 0$). We have again assumed symmetry in the connection strengths such that the pairs (E_1, I_1) and (E_2, I_2) are identical in their connectivity and dynamics.

We fix W_{EE} and W_{II} and proceed by exploring the space of W_{EI} and W_{IE} connections. We find that this network structure now yields a robust region in which the network is stable and the $E_1 - E_2$ correlations are strong and positive (Fig. 5*B*, green). In fact, we can mathematically prove that with this given network configuration, network stability implies $\mathrm{Corr}(E_1, E_2) > 0$ (*SI Appendix*). This result emphasizes three important points. First, there exists a large space of connection parameters in which our criteria (large $\mathrm{Corr}(E_1, E_2)$ and $\alpha \ll 1$) may be met. Given the heterogeneity of neural circuits and plasticity of connections within the cortex, this parametric result is much more satisfying than a fine-tuned solution like that required in the model with global inhibition (Fig. 4). Second, this result does not depend on the presence of external correlated fluctuations. Third, this result is robust to the presence of external correlated input noise as it would only further increase $\mathrm{Corr}(E_1, E_2)$.

To supplement our mathematical proof with a clear intuition for the result we start by considering the single E_1/I_1 ISN unit unconnected from E_2/I_2 (i.e., $\alpha = 0$). In this parameter regime, despite receiving inputs from independent noisy processes, E_1 and I_1 are known to covary together (21, 22). The activity in E_1 will therefore correlate with an incoming excitatory "signal," as this signal will drive an increase in activity for both E_1 and I_1. This is exactly what happens for $\alpha > 0$, as E_2 sends an excitatory signal into E_1 (and vice versa for E_1 into E_2). We note that such an argument cannot be applied for the three-population model with a global inhibitory population we considered in the previous section; in that case, you do not have separate ISN units. There, I covaries with respect to the joint $E_1 + E_2$ activity, as opposed to each of them separately (as seen in Fig. 4*C*).

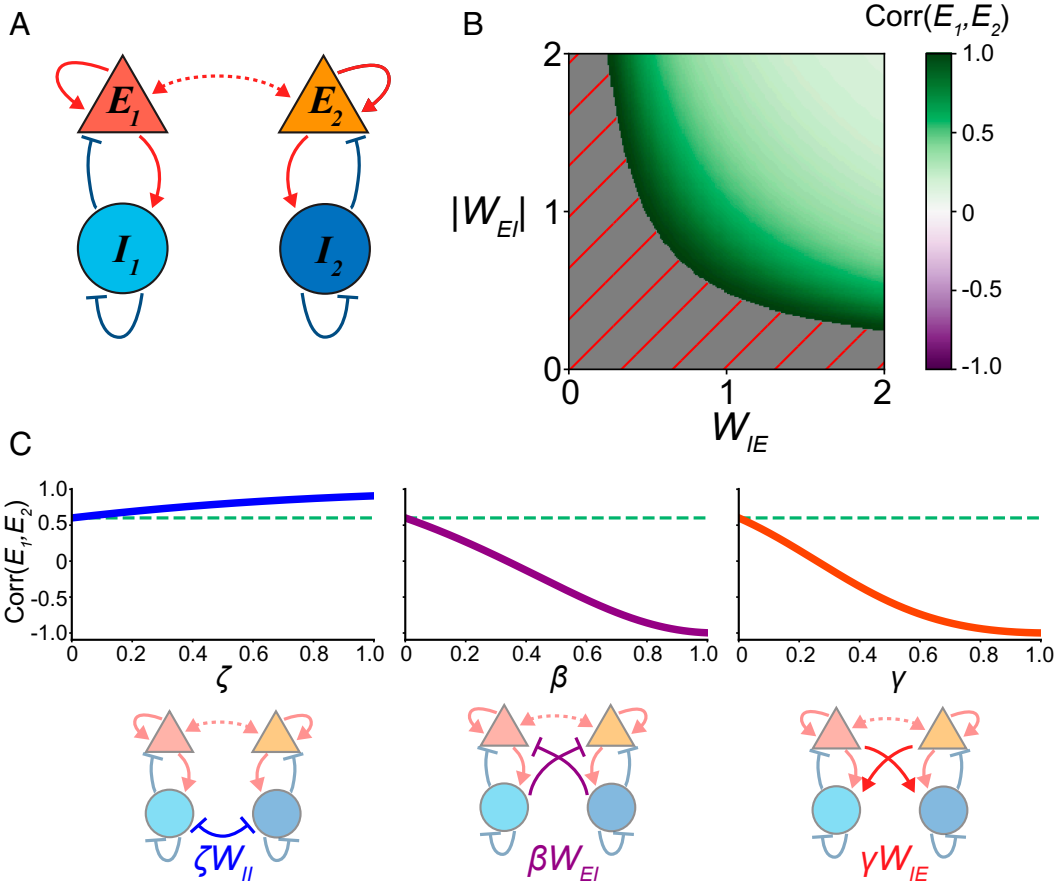

Fig. 5. Segregated *I* subpopulations produce robust positive correlations. (*A*) Model schematic. Input structure is consistent with Fig. 3*A* but omitted for clarity. (*B*) $\mathrm{Corr}(E_1, E_2)$ as a function of W_{EI}, W_{IE} with $c = 0$. (*C*) $\mathrm{Corr}(E_1, E_2)$ as a function of added connections between I_1, I_2 (*Left*); $I_1 \to E_2$ and $I_2 \to E_1$ (*Middle*); $E_1 \to I_2$ and $E_2 \to I_1$ (*Right*). Added connections W_{ij} are initialized to the same as elsewhere in the network, and scaled by ζ, $I \leftrightarrow I$; β, $I \to E$; γ, $E \to I$. The dashed turquoise line denotes $\zeta, \beta, \gamma = 0$, $W_{EI} = W_{IE} = 1$.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

Using this intuition, we then ask whether any of the other interpopulation connections (i.e., $E_i \leftrightarrow I_j$ or $I_i \leftrightarrow I_j$) would support large and positive $\mathrm{Corr}(E_1, E_2)$. Using the same logic as before but for inhibitory inputs, we anticipate that adding $I_j \rightarrow E_i$ connections would decrease correlations. More specifically, as the E_2/I_2 unit increases in activity, the inhibitory $I_2 \rightarrow E_1$ pathway would decrease activity in the E_1/I_1 unit, thus leading E_1 and E_2 to become anticorrelated. The same reasoning can almost be applied to the addition of $I_i \leftrightarrow I_j$ connections, except one must keep in mind the paradoxical effect that is a hallmark of the ISN regime (21, 22). Namely, an inhibitory current into the inhibitory population paradoxically leads to an increase in the inhibitory and excitatory populations. As a result, after following the same logic as before but with this effect in mind, we predict that adding $I_i \leftrightarrow I_j$ connections would increase E_1/E_2 correlations. Last, since the paradoxical effect also needs to be considered for the $E_i \rightarrow I_j$ connections, one predicts that this connection would decrease E_1/E_2 correlations.

To test these predictions, we now consider fixed values of W_{EI}, W_{IE}, and W_{II}, and introduce scaling parameters β, γ, ζ, respectively, to adjust the between-population strengths of each connection (see schematics in Fig. 5 *C*, *Bottom*). Consistent with our predictions, we find that only $\zeta > 0$ further enhances correlations above the value we found when $\zeta = 0$ (Fig. 5 *C*, *Left*), while any nonzero values of β, γ only reduce correlations (Fig. 5 *C*, *Middle* and *Right*). Further, this same relationship also held when β, γ, ζ were covaried. Specifically, only $I \leftrightarrow I$ connections served as a correlating force; all others induced a reduction in correlations (*SI Appendix*, Fig. S2). Hence, we conclude that while inhibition in mouse V1 can be promiscuously connected with other inhibitory units, it must be strongly coclustered with excitatory subpopulations and sparse in its connectivity with other excitatory subpopulations to yield the significant positive interpopulation excitatory correlations observed in Kim et al. (10).

Discussion

In this study, we sought to uncover possible neural circuit mechanisms underpinning the observation that pyramidal neurons in the primary visual cortex that project to different downstream targets connect with a much lower probability than random pairs of excitatory neurons, yet still exhibit correlated variability that is almost as large as the rest of mouse V1 (10). Notably, the magnitude of these correlations is much stronger than would be predicted given their weak connectivity. We found that a model with global inhibition resulted in highly constrained regions in which the data could be matched, encompassing two distinct solutions. In the case of weak network coupling, positive correlations resulted from two forms of an inheritance model: either $I \rightarrow E$ connections induced increased correlated activity through I affecting both excitatory populations in the same way or an unobserved external source of strong correlations fed these fluctuations across both E units. For the case of strong recurrent coupling where the circuit is in an inhibition-stabilized network (ISN) regime, the network connectivity needed to place the network dynamics right at the edge of an instability to produce positive correlations. By contrast, we found that a more generally robust solution in the ISN regime could be achieved by splitting the inhibitory population into two separate subpopulations coclustered with one of the excitatory subnetworks. It bears noting that only in the first model without inhibition did we explicitly explore the effects of correlations in the external inputs. In general, one could recapture the data simply by imposing significant positive correlations on the excitatory subpopulations in any of the models explored here. While possible, this solution is dissatisfying insofar as it is another form of fine-tuning, requiring some secondary source to generate the appropriate correlation structure.

We therefore argue that, on the basis of this robustness, our results predict that inhibition should cluster together with excitation in the mouse sensory cortex with a specificity that mirrors that of the excitatory connectivity. The other inferred models by contrast depend upon narrow parameter regimes to capture experimental observations. This fragility would require significant constraints on the properties of neural circuits. Yet, connections are plastic, connection strengths are heterogeneous, and neuron properties are affected by neuromodulation (25, 26). Given this stochasticity in the circuit structure itself, a fine-tuned solution is unlikely to best explain the data.

Rigorous experimental validation of our model predictions could be obtained through physiological or connectomics experiments which specifically target the relationship between excitatory projection neurons and local inhibitory neurons. In this case, we would expect that inhibitory interneurons which strongly connect to a given excitatory projection subclass would also exhibit a lower probability of connection with excitatory neurons of the other projection subclass. Given that our results are agnostic as to the promiscuity with which inhibition connects to other inhibitory cells, it would be interesting to learn the structure of these interactions. While it is well appreciated that inhibitory interneurons are very diverse in physiology and connectivity (27), we did not explicitly model this diversity in our study. Nevertheless, we anticipate that parvalbumin (PV)-positive cells may display the identified signatures of our I units, as they appear to play a critical role in stabilizing excitatory activity (28). Recent experimental evidence appears to support this claim from the perspective of stimulus tuning: While PV cells connect with most nearby pyramidal neurons, they were found to more strongly connect with those whose tuning properties they share (16).

Theoretical work has argued that E/PV assembly formation requires plasticity from both $E \rightarrow$ PV and PV $\rightarrow E$ connections (29). This bidirectionality could result in local, winner-take-all effects in $E \leftrightarrow I$ connectivity as any discrepancies in functional response properties between nearby pyramidal cells will bias the PV connectivity. This could result in the more specific coclustering of inhibition we predict. Motivated by these results, a potential indirect way to differentiate between the global and clustered inhibition models would be to record activity of AL- and PM-projecting neurons together with inhibitory interneurons. Comparison of their respective tuning functions could suggest whether the inhibitory cell is biased in its connectivity (by extension of ref. 16). Indeed, Najafi et al. (30) recently argued for coclustered excitation–inhibition in the context of posterior parietal cortex decision circuitry on the basis of neural response properties. Furthermore, in the mouse visual cortex, it has been shown that PM and AL exhibit distinct functional representations with some overlap (31), consistent with the tuning properties of V1 projection neurons (10).

Of course, it is possible that inhibitory–excitatory interactions may span a continuum between the global and clustered motifs identified here. This raises the possibility that heterogeneity in inhibitory connectivity motifs at small spatial scales may explain heterogeneity in pairwise covariance between AL- and PM-projecting pyramidal cells. To this end, modern theoretical tools enable the calculation of not just the mean covariance but the distribution of pairwise covariability across networks with disordered connectivity (37, 45). An interesting avenue for future work would be to relate the distribution of correlations across

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

different excitatory projection neurons to heterogeneity in their connectivity to inhibitory interneurons, beyond simply the mean correlations as we explored here.

A central issue in the extension of our results concerns the dynamical regime of the cortex, a topic which has received a significant amount of attention lately (32–34). One question concerns whether intracortical interactions are strong enough to require inhibition as a key stabilizer of activity, that is, whether the sensory cortex is an ISN (21, 23). Theoretical work predicts that in this regime the ratio of excitatory to inhibitory input drive to a neuron decreases with increasing stimulus intensity (35). Recent experimental evidence from recordings of the mouse primary visual cortex supports this claim (15). Another study used optogenetic perturbation of inhibitory neurons across mouse cortex to test for inhibition-stabilization without sensory stimulation, finding evidence that all considered cortical regions operate as an ISN (24).

Given this evidence for an ISN regime, a second question regards whether the network dynamics are poised near a network instability. In our model, loss of stability would result in large positive correlations through a slowing down of the dynamics (Fig. 4). Analysis of array recordings in the primate motor cortex has suggested that cortical dynamics may in fact be positioned close to an instability (45). More recently, reexamination of large-scale recordings in mice has suggested that other cortical regions, including the primary visual cortex, exhibit dynamics with dominant eigenvalues close to one (33). This could suggest that either the global or clustered inhibition model in an ISN regime may explain the data. Together with the foregoing evidence that PV and E neurons sharing tuning properties connect more strongly, we argue that this further supports a model of coclustered inhibition.

Other mechanisms by which correlations can grow near a change in stability have been identified in previous studies. Ginzburg and Sompolinsky (36) observed that near a bifurcation—in their case, a saddle node or Hopf—correlations in a weakly connected network grow from $\mathcal{O}(1/N)$ to near $\mathcal{O}(1)$, where N is the network size, together with a slowing down in the dynamics. Darshan et al. (37) derived conditions on what they term the interaction matrix (similar to our $\mathbf{W}$ matrix) under which correlations are amplified without critical slowing down. These network models thus suggest distinct mechanisms by which our results could be extended to spatially distributed spiking network models. Additionally, Litwin-Kumar and Doiron (38) studied the effect of clustered connectivity in balanced spiking networks on the structure of correlations; however, this work did not compare across-cluster to within-cluster correlations. Rosenbaum et al. (39) did consider a structure similar to our three-population global inhibition motif, demonstrating that, consistent with our conclusions, a spatially distributed spiking neural network with distinct subpopulations would show close to zero correlations on average due to strong positive correlations within a cluster and large negative correlations between the two clusters. Yet it remains for future work to determine the precise parametric values to recapitulate our results in spiking neural network models.

Our work can be seen as a case study of a particular network structure in the context of the theoretical investigation of dynamics on graphs (that is, a collection of nodes and edges). In general, graphical analysis has been used in a wide range of neuroscientific applications, from the determination of fixed points of dynamics (40) to network controllability (41). In relating connectivity motifs (elements of $\mathbf{W}$ and their combinations) to correlation structure in the circuit, our approach relates to a more general mathematical concept of relating process motifs on networks to underlying structure motifs of the graph (42).

Ultimately, our work demonstrates how ostensibly straightforward observations of connectivity and response properties from cortical cells have the capacity to lend fruitful insight into the structural and dynamical regimes of the cortex, which are critical to further understanding of information processing in the brain.

Materials and Methods

Firing Rate Model. As done previously (12, 18), we consider the firing rate dynamics of neuronal populations A given by the following

$$\tau_A \frac{dr_A}{dt} = -r_A + f_A\Big(\mu_A + \sum_B J_{AB} r_B + \hat{\sigma}_A\big[\sqrt{1-c}\cdot x_A(t) + \sqrt{c}\cdot x_s(t)\big]\Big),$$

where τ_A is the time constant, μ_A is a constant stimulus drive, and J_{AB} is the strength of connections from population B to A. The stochastic processes $x_A(t)$ and $x_s(t)$ represent private and shared global fluctuations, respectively. Each is taken to be the limiting process from

$$\tau_x \frac{dx}{dt} = -x + \sqrt{\tau_x}\xi_x(t),$$

for $\tau_x \to 0$, with $\langle \xi_i(t)\rangle = 0$ and $\langle \xi_i(t)\xi_i(t')\rangle = \delta(t-t')$. Intuitively, one may think of $x(t)$ as a "smoothed" white noise process (12). The parameter $c \in [0, 1]$ scales the proportion of shared noise relative to private noise, while $\hat{\sigma}_A$ represents the total intensity of the fluctuations.

We assume that the system of equations has an equilibrium point at r_{ss}, and that the noise is weak enough so that the fluctuations about this equilibrium ($\Delta r := r - r_{ss}$) can be approximated by

$$\tau_A \frac{d\Delta r_A}{dt} = -\Delta r_A + L_A \sum_B J_{AB}\Delta r_B + L_A\hat{\sigma}_A\Big[\sqrt{1-c}\cdot x_A(t) + \sqrt{c}\cdot x_s(t)\Big],$$

where $L_A = f_A'(r_{ss})$ is the gain of population A at the equilibrium point. We define the effective coupling as $W_{AB} := L_A J_{AB}$ and $\sigma_A := L_a\hat{\sigma}_A$, and approximate $x_A(t)$ and $x_s(t)$ as independent, zero-mean Gaussian processes $\xi_A(t)$ and $\xi_s(t)$ satisfying $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$. This yields Eq. **1**, which in matrix form can be written as

$$\mathbf{T}\frac{d\Delta\mathbf{r}}{dt} = (\mathbf{W} - \mathbf{I})\Delta\mathbf{r}(t) + \mathbf{D}\boldsymbol{\xi}(t). \qquad [8]$$

For notational simplicity, throughout we will assume unit time constants $\tau_A = 1$, so that $\mathbf{T} = \mathbf{I}$. For example, in the case of two excitatory populations and one inhibitory population $\{E_1, E_2, I\}$ the matrices are

$$\mathbf{W} = \begin{bmatrix} W_{E_1E_1} & W_{E_1E_2} & W_{E_1I} \\ W_{E_2E_1} & W_{E_2E_2} & W_{E_2I} \\ W_{IE_1} & W_{IE_2} & W_{II} \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} \sqrt{(1-c)}\cdot\sigma_{E_1} & 0 & 0 & \sqrt{c}\cdot\sigma_{E_1} \\ 0 & \sqrt{(1-c)}\cdot\sigma_{E_2} & 0 & \sqrt{c}\cdot\sigma_{E_2} \\ 0 & 0 & \sigma_I & 0 \end{bmatrix}.$$

The network structure is determined through the weight matrix $\mathbf{W}$. Since we are explicitly interested in segregated excitatory populations, we consider weak cross-population connections and set

$$W_{E_2E_1} = \alpha W_{E_1E_1}, \qquad W_{E_1E_2} = \alpha W_{E_2E_2},$$

for $\alpha \in (0, 1)$. The two excitatory populations, E_1 and E_2, are increasingly disconnected as $\alpha \to 0$. To obtain analytical expressions and constrain the searchable parameter space, we assume various symmetries in the network

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

connectivity. Specifically, we consider the following forms for connectivity matrices for the two (Fig. 1*A*), three (Fig. 3*A*), and four (Fig. 4*A*) population models:

$$\mathbf{W} = \begin{bmatrix} W_{EE} & \alpha W_{EE} \\ \alpha W_{EE} & W_{EE} \end{bmatrix},$$

$$\mathbf{W} = \begin{bmatrix} W_{EE} & \alpha W_{EE} & W_{EI} \\ \alpha W_{EE} & W_{EE} & W_{EI} \\ W_{IE} & W_{IE} & W_{II} \end{bmatrix},$$

and

$$\mathbf{W} = \begin{bmatrix} W_{EE} & \alpha W_{EE} & W_{EI} & \beta W_{EI} \\ \alpha W_{EE} & W_{EE} & \beta W_{EI} & W_{EI} \\ W_{IE} & \gamma W_{IE} & W_{II} & \zeta W_{II} \\ \gamma W_{IE} & W_{IE} & \zeta W_{II} & W_{II} \end{bmatrix},$$

where $\beta, \gamma, \zeta \in (0, 1)$.

Covariance Calculation. The autocovariance function for the OU process defined in Eq. **8** is given by

$$\widetilde{\mathbf{C}}(h) = \langle \Delta\mathbf{r}(t)\Delta\mathbf{r}(t+h)\rangle.$$

Let $\mathbf{M} = \mathbf{W} - \mathbf{I}$ and define $\mathbf{\Sigma} := \widetilde{\mathbf{C}}(0) = \langle \Delta\mathbf{r}(t)\Delta\mathbf{r}(t)\rangle$ as the stationary covariance matrix. Then, $\mathbf{\Sigma}$ is obtained as the solution to the Lyapunov equation $-\mathbf{M\Sigma} + \mathbf{\Sigma}(-\mathbf{M})^\top = \mathbf{DD}^\top$ (43). It follows that

$$\widetilde{\mathbf{C}}(h) = \begin{cases} e^{-\mathbf{M}h} \cdot \mathbf{\Sigma}, & h < 0, \\ \mathbf{\Sigma} \cdot e^{\mathbf{M}^\top h}, & h \geq 0. \end{cases} \quad [9]$$

Integrating $\widetilde{\mathbf{C}}(h)$ in each element over long times h yields the following compressed form for the long-time covariance matrix $\mathbf{C}$

$$\begin{aligned} \mathbf{C} &= \int_{-\infty}^{\infty} \widetilde{\mathbf{C}}(h)\, dh \\ &= \mathbf{M}^{-1}\mathbf{D}\left(\mathbf{M}^{-1}\mathbf{D}\right)^\top. \end{aligned}$$

If $\mathbf{C}_V = \sqrt{\mathrm{diag}(\mathbf{C})}$, then the correlation matrix is obtained

$$\boldsymbol{\rho} = \mathbf{C}_V^{-1}\mathbf{C}\mathbf{C}_V^{-1}. \quad [10]$$

Path Expansion. If the spectral radius $s(\mathbf{W}) = \max\{|\lambda_i|\colon \lambda_i \text{ is an eigenvalue of } \mathbf{W}\} < 1$, then $\mathbf{M}^{-1}$ has a convergent series representation

$$-\mathbf{M}^{-1} = (\mathbf{I} - \mathbf{W})^{-1} = \sum_{k=0}^{\infty} \mathbf{W}^k$$

known as a Neumann series (44). Intuitively, one may think of the Neumann series as a matrix analogue of the familiar geometric series. Under this representation, the long-time covariance matrix is

$$\mathbf{C} = \left(\sum_{j=0}^{\infty} \mathbf{W}^j\right) \mathbf{DD}^\top \left(\sum_{k=0}^{\infty} \mathbf{W}^k\right)^\top.$$

It is useful to rewrite this expansion as

$$\mathbf{C} = \sum_{n=0}^{\infty} \left[\sum_{i=0}^{n} \mathbf{W}^{n-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i\right],$$

where the terms in the inner sum can be interpreted as contributions due to nth-order paths through the network (5, 14).

If the outer sum converges quickly, the covariance matrix can be approximated as

$$\mathbf{C} \approx \sum_{n=0}^{N} \left[\sum_{i=0}^{n} \mathbf{W}^{n-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i\right].$$

The rate of convergence of this approximation depends on the magnitude of $s(\mathbf{W})$. In particular, the closer $s(\mathbf{W})$ is to 0, the faster the terms shrink. Consider the N-th order terms of this approximation,

$$\sum_{i=0}^{N} \mathbf{W}^{N-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i.$$

If $||\cdot||$ is the operator norm, then

$$\left\| \sum_{i=0}^{N} \mathbf{W}^{N-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i \right\| \leq \sum_{i=0}^{N} \left\| \mathbf{W}^{N-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i \right\|,$$

and that each term in this sum can be bounded above by

$$\left\| \mathbf{W}^{N-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i \right\| \leq \left\| \mathbf{W}^{N-i} \right\| \cdot \|\mathbf{D}\|^2 \cdot \left\| \left(\mathbf{W}^\top\right)^i \right\|.$$

After diagonalizing $\mathbf{W}$ and writing $\mathbf{W} = \mathbf{P\Lambda P}^{-1}$, where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues of $\mathbf{W}$, it follows that

$$\begin{aligned} \left\| \mathbf{W}^{N-i} \right\| \left\| \left(\mathbf{W}^\top\right)^i \right\| &= \left\| \left(\mathbf{P\Lambda P}^{-1}\right)^{N-i} \right\| \cdot \left\| \left((\mathbf{P}^{-1})^\top \mathbf{\Lambda P}^\top\right)^i \right\| \\ &= \left\| \mathbf{P\Lambda}^{N-i}\mathbf{P}^{-1} \right\| \cdot \left\| (\mathbf{P}^{-1})^\top \mathbf{\Lambda}^i \mathbf{P}^\top \right\| \\ &\leq \|\mathbf{P}\|^2 \left\| \mathbf{P}^{-1} \right\|^2 \|\mathbf{\Lambda}\|^N. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{i=0}^{N} \left\| \mathbf{W}^{N-i}\mathbf{DD}^\top \left(\mathbf{W}^\top\right)^i \right\| &\leq N \left(\|\mathbf{P}\| \left\| \mathbf{P}^{-1} \right\| \|\mathbf{D}\| \right)^2 \|\mathbf{\Lambda}\|^N \\ &\leq N \left(\|\mathbf{P}\| \left\| \mathbf{P}^{-1} \right\| \|\mathbf{D}\| \right)^2 s(\mathbf{W})^N. \end{aligned}$$

This bound shrinks quickly as $N \to \infty$ if $s(\mathbf{W})$ is small ($\ll 1$), as is the case when the system is in the weakly coupled regime.

Path expansion for weakly coupled $E_1 \leftrightarrow E_2$. In Fig. 2*B* we illustrate this quick convergence by showing the first three terms of this sum, namely

$$\begin{aligned} &\text{0th-order: } \mathbf{DD}^\top, \\ &\text{1st-order: } \mathbf{WDD}^\top + \mathbf{DD}^\top\mathbf{W}^\top, \\ &\text{2nd-order: } \mathbf{W}^2\mathbf{DD}^\top + \mathbf{WDD}^\top\mathbf{W}^\top + \mathbf{DD}^\top\left(\mathbf{W}^\top\right)^2. \end{aligned}$$

Using these terms, the cross-population covariance can be approximated by Eq. **5** in the main text.

We note that for a nth-order path, we multiply on the left and right by $\mathbf{C}_V^{-1}$ to obtain path contributions to the correlation matrix. In particular, we are interested in the contributions to $\rho_{E_1E_2}$ (that is, the element $\boldsymbol{\rho}_{1,2}$ of Eq. **10**).

Table 1. Strength of connections from pop

W_{AB}	E	I
E	0.5 (1.15)	0.5 (0.8)
I	0.5 (0.8)	0.5 (0.5)

B (columns) to *A* (rows) for the weakly (strongly) coupled model.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.

Table 2. Default parameter values

Parameter	Default value	Description
α	0.15	Interexcitatory population strength
σ_A	1	Total intensity of outside fluctuations
c	0	Scales the proportion of shared noise relative to private noise

Changes to any parameter are indicated in the figure caption.

Parameters and Simulations. Simulations were performed using an Euler–Maruyama scheme with time constants $\tau_E = \tau_I = 15$ ms, $dt = 0.01$ ms. Remaining default parameters values can be found in Tables 1 and 2.

Data, Materials, and Software Availability. Jupyter notebooks that run relevant simulations and reproduce the main figures of the paper can be accessed via a publicly available Zenodo repository (https://doi.org/10.5281/zenodo.11398126) (46).

ACKNOWLEDGMENTS. A.N., M.P.G., and G.H. thank the Simons Foundation SCGB Undergraduate Research Fellowship (SURF) for fostering this collaboration. A.N. was funded by Simons Foundation SURF. G.H. was supported by the Burroughs Wellcome Fund's Career Award at the Scientific Interface. B.D. is supported by the NIH (Grant Nos. 1U19NS107613-01 and R01EB026953), Vannevar Bush faculty fellowship (No. N00014-18-1-2002), and the Simons Foundation Collaboration on the Global Brain. B.D. acknowledges support from the University of Chicago's Center for Living Systems funded through NSF-PHY-2317138.

Author affiliations: [a]Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616; [b]Grossman Center for Quantitative Biology and Human Behavior, University of Chicago, Chicago, IL 60637; [c]Department of Neurobiology, University of Chicago, Chicago, IL 60637; and [d]Department of Statistics, University of Chicago, Chicago, IL 60637

Author contributions: A.N., M.P.G., G.H., and B.D. designed research; A.N., M.P.G., and G.H. performed research; A.N., M.P.G., and G.H. analyzed data; and A.N., M.P.G., G.H., and B.D. wrote the paper.

1. A. E. Urai, B. Doiron, A. M. Leifer, A. K. Churchland, Large-scale neural recordings call for new insights to link brain and behavior. *Nat. Neurosci.* **25**, 11–19 (2022).
2. B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker, K. Josić, The mechanics of state-dependent neural correlations. *Nat. Neurosci.* **19**, 383–393 (2016).
3. G. K. Ocker *et al.*, From the statistics of connectivity to the statistics of spike times in neuronal networks. *Curr. Opin. Neurobiol.* **46**, 109–119 (2017).
4. M. Helias, T. Tetzlaff, M. Diesmann, The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol.* **10**, e1003428 (2014).
5. J. Trousdale, Y. Hu, E. Shea-Brown, K. Josić, Impact of network structure and cellular response on spike time correlations. *PLoS Comput. Biol.* **8**, e1002408 (2012).
6. R. Perin, T. K. Berger, H. Markram, A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5419–5424 (2011).
7. A. A. Faisal, L. P. Selen, D. M. Wolpert, Noise in the nervous system. *Nat. Rev. Neurosci.* **9**, 292–303 (2008).
8. H. Ko *et al.*, Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
9. L. Cossell *et al.*, Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518**, 399–403 (2015).
10. M. H. Kim, P. Znamenskiy, M. F. Iacaruso, T. D. Mrsic-Flogel, Segregated subnetworks of intracortical projection neurons in primary visual cortex. *Neuron* **100**, 1313–1321 (2018).
11. K. M. Hagihara, A. W. Ishikawa, Y. Yoshimura, Y. Tagawa, K. Ohki, Long-range interhemispheric projection neurons show biased response properties and fine-scale local subnetworks in mouse visual cortex. *Cereb. Cortex* **31**, 1307–1315 (2021).
12. T. Kanashiro, G. K. Ocker, M. R. Cohen, B. Doiron, Attentional modulation of neuronal variability in circuit models of cortex. *eLife* **6**, e23978 (2017).
13. T. Tetzlaff, M. Helias, G. T. Einevoll, M. Diesmann, Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol.* **8**, e1002596 (2012).
14. V. Pernice, B. Staude, S. Cardanobile, S. Rotter, How structure determines correlations in neuronal networks. *PLoS Comput. Biol.* **7**, e1002059 (2011).
15. H. Adesnik, Synaptic mechanisms of feature coding in the visual cortex of awake mice. *Neuron* **95**, 1147–1159 (2017).
16. P. Znamenskiy *et al.*, Functional specificity of recurrent inhibition in visual cortex. *Neuron* **112**, 991–1000 (2024).
17. A. Renart, N. Brunel, X. J. Wang, "Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks" in *Computational Neuroscience*, J. Feng, Ed. (Chapman and Hall/CRC, 2004), pp. 431–490.
18. M. P. Getz, C. Huang, B. Doiron, Subpopulation codes permit information modulation across cortical states. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.09.28.509815 (Accessed 30 September 2022).
19. S. B. Hofer *et al.*, Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14**, 1045–1052 (2011).
20. A. M. Packer, R. Yuste, Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: A canonical microcircuit for inhibition? *J. Neurosci.* **31**, 13260–13271 (2011).
21. M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, B. L. McNaughton, Paradoxical effects of external modulation of inhibitory interneurons. *J. Neurosci.* **17**, 4382–4388 (1997).
22. H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, D. Ferster, Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**, 578–592 (2009).
23. S. Sadeh, C. Clopath, Inhibitory stabilization and cortical computation. *Nat. Rev. Neurosci.* **22**, 21–37 (2021).
24. A. Sanzeni *et al.*, Inhibition stabilization is a widespread property of cortical networks. *eLife* **9**, e54875 (2020).
25. G. G. Turrigiano, The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell* **135**, 422–435 (2008).
26. E. Marder, Neuromodulation of neuronal circuits: Back to the future. *Neuron* **76**, 1–11 (2012).
27. R. Tremblay, S. Lee, B. Rudy, Gabaergic interneurons in the neocortex: From cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
28. H. Bos, A. M. Oswald, B. Doiron, Untangling stability and gain modulation in cortical circuits with multiple interneuron classes. bioRxiv [Preprint] (2020). https://doi.org/10.1101/2020.06.15.148114 (Accessed 17 June 2020).
29. O. Mackwood, L. B. Naumann, H. Sprekeler, Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *eLife* **10**, e59715 (2021).
30. F. Najafi *et al.*, Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning. *Neuron* **105**, 165–179 (2020).
31. M. L. Andermann, A. M. Kerlin, D. K. Roumis, L. L. Glickfeld, R. C. Reid, Functional specialization of mouse higher visual cortical areas. *Neuron* **72**, 1025–1039 (2011).
32. Y. Ahmadian, K. D. Miller, What is the dynamical regime of cerebral cortex? *Neuron* **109**, 3373–3391 (2021).
33. G. B. Morales, S. Di Santo, M. A. Muñoz, Quasiuniversal scaling in mouse-brain neuronal activity stems from edge-of-instability critical dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2208998120 (2023).
34. C. Huang, Modulation of the dynamical state in cortical network models. *Curr. Opin. Neurobiol.* **70**, 43–50 (2021).
35. D. B. Rubin, S. D. Van Hooser, K. D. Miller, The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**, 402–417 (2015).
36. I. Ginzburg, H. Sompolinsky, Theory of correlations in stochastic neural networks. *Phys. Rev. E* **50**, 3171 (1994).
37. R. Darshan, C. Van Vreeswijk, D. Hansel, Strength of correlations in strongly recurrent neuronal networks. *Phys. Rev. X* **8**, 031072 (2018).
38. A. Litwin-Kumar, B. Doiron, Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**, 1498–1505 (2012).
39. R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, B. Doiron, The spatial structure of correlated neuronal variability. *Nat. Neurosci.* **20**, 107–114 (2017).
40. K. Morrison, C. Curto, "Predicting neural network dynamics via graphical analysis" in *Algebraic and Combinatorial Computational Biology*, R. Robeva, M. Macauley, Eds. (Elsevier, 2019), pp. 241–277.
41. J. Z. Kim *et al.*, Role of graph architecture in controlling dynamical networks with applications to neural systems. *Nat. Phys.* **14**, 91–98 (2018).
42. A. C. Schwarze, M. A. Porter, Motifs for processes on networks. *SIAM J. Appl. Dyn. Syst.* **20**, 2516–2557 (2021).
43. C. Gardiner, *Stochastic Methods* (Springer, Berlin, Germany, 2009), vol. 4.
44. M. Einsiedler *et al.*, *Functional Analysis, Spectral Theory, and Applications* (Springer, 2017), vol. 104.
45. D. Dahmen, S. Grün, M. Diesmann, M. Helias, Second type of criticality in the brain uncovers rich multiple-neuron dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13051–13060 (2019).
46. A. Negrón, M. Getz, G. Handy, The-mechanics-of-correlated-variability-in-segregated-cortical-excitatory-subnetworks. Zenodo. https://doi.org/10.5281/zenodo.11398126. Deposited 30 May 2024.

Downloaded from https://www.pnas.org by Kirsten Vallee on July 23, 2024 from IP address 99.7.2.48.