Article

Neuron

Circuit Models of Low-Dimensional Shared Variability in Cortical Networks

Highlights

- Low-dimensional shared variability can be generated in spatial network models
- Synaptic spatial and temporal scales determine the dimensions of shared variability
- Depolarizing inhibitory neurons suppresses the population-wide fluctuations
- Modeling the attentional modulation of variability within and between brain areas

Authors

Chengcheng Huang, Douglas A. Ruff, Ryan Pyle, Robert Rosenbaum, Marlene R. Cohen, Brent Doiron

Correspondence

bdoiron@pitt.edu

In Brief

Population-wide fluctuations of neural population activity are widely observed in cortical recordings. Huang et al. show that turbulent dynamics in spatially ordered recurrent networks give rise to low-dimensional shared variability, which can be suppressed by depolarizing inhibitory neurons.

Huang et al., 2019, Neuron *101*, 337–348
January 16, 2019
https://doi.org/10.1016/j.neuron.2018.11.034

CellPress

Neuron
Article

Circuit Models of Low-Dimensional Shared Variability in Cortical Networks

Chengcheng Huang,[1,2] Douglas A. Ruff,[2,3] Ryan Pyle,[4] Robert Rosenbaum,[4,5] Marlene R. Cohen,[2,3] and Brent Doiron[1,2,6,*]
[1]Department of Mathematics, University of Pittsburgh, Pittsburgh, PA, USA
[2]Center for the Neural Basis of Cognition, Pittsburgh, PA, USA
[3]Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA, USA
[4]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN, USA
[5]Interdisciplinary Center for Network Science and Applications, University of Notre Dame, Notre Dame, IN, USA
[6]Lead Contact
*Correspondence: bdoiron@pitt.edu
https://doi.org/10.1016/j.neuron.2018.11.034

SUMMARY

Trial-to-trial variability is a reflection of the circuitry and cellular physiology that make up a neuronal network. A pervasive yet puzzling feature of cortical circuits is that despite their complex wiring, population-wide shared spiking variability is low dimensional. Previous model cortical networks cannot explain this global variability, and rather assume it is from external sources. We show that if the spatial and temporal scales of inhibitory coupling match known physiology, networks of model spiking neurons internally generate low-dimensional shared variability that captures population activity recorded *in vivo*. Shifting spatial attention into the receptive field of visual neurons has been shown to differentially modulate shared variability within and between brain areas. A top-down modulation of inhibitory neurons in our network provides a parsimonious mechanism for this attentional modulation. Our work provides a critical link between observed cortical circuit structure and realistic shared neuronal variability and its modulation.

INTRODUCTION

The trial-to-trial variability of neuronal responses gives a critical window into how the circuit structure connecting neurons determines brain activity (Kass et al., 2018; Shadlen and Newsome, 1998; Doiron et al., 2016). This idea, combined with the widespread use of population recordings, has prompted deep interest in how variability is distributed over a population (Cohen and Kohn, 2011; Kohn et al., 2016). There has been a proliferation of datasets in which the shared variability over a population is low dimensional (Lin et al., 2015; Rabinowitz et al., 2015; Ecker et al., 2014; Williamson et al., 2016; Schölvinck et al., 2015), meaning that neuronal activity waxes and wanes as a group. In accord, one-dimensional measures such as local field potentials (Kelly et al., 2010; Middleton et al., 2012) and summed population firing rates (Okun et al., 2015; Schölvinck et al., 2015) can predict a majority of pairwise correlations. Further, the synthesis of diverse population datasets paints a picture in which low-dimensional shared variability is a signature of cognitive state, such as overall arousal, task engagement, and attention (Doiron et al., 2016; Schmitz and Duncan, 2018), as well as predictive of behavioral performance (Ni et al., 2018). Such low-dimensional dynamics portend a theory for the genesis and modulation of shared population variability in recurrent cortical networks.

Theories of cortical variability can be broadly separated into two categories: ones in which variability is internally generated through recurrent network interactions and ones in which variability originates external to the network. Networks of spiking neuron models where strong excitation is balanced by opposing recurrent inhibition produce high single-neuron variability through internal mechanisms (Shadlen and Newsome, 1998; van Vreeswijk and Sompolinsky, 1996; Amit and Brunel, 1997). However, these networks famously enforce an asynchronous state and as such fail to explain population-wide shared variability (Renart et al., 2010). This lack of success is contrasted with the ease of producing arbitrary correlation structure from external sources. Indeed, many past cortical models assume a global fluctuation from an external source and accurately capture the structure of population data (Doiron et al., 2016; Ponce-Alvarez et al., 2013; Wimmer et al., 2015; Kanashiro et al., 2017; Hennequin et al., 2018). However, these phenomenological models begin with an assumption of low-dimensional variability from an unobserved source to explain the variability in a recorded population. In this way, these models are somewhat circular, begging the question of what are the mechanisms underlying the assumed external variability. Thus, while neuronal variability has a rich history of study, there remains an impoverished mechanistic understanding of the low-dimensional structure of population-wide variability (Latham, 2016).

Determining whether output variability is internally generated through network interactions or externally imposed upon a network is a difficult problem, in which single-area population recordings may preclude any definitive solution. In this study we consider attention-mediated shifts in population variability obtained from simultaneous recordings of neuron pairs both within and between visual areas. In particular, attention reduces within area correlations (area V4; Cohen and Maunsell, 2009) while

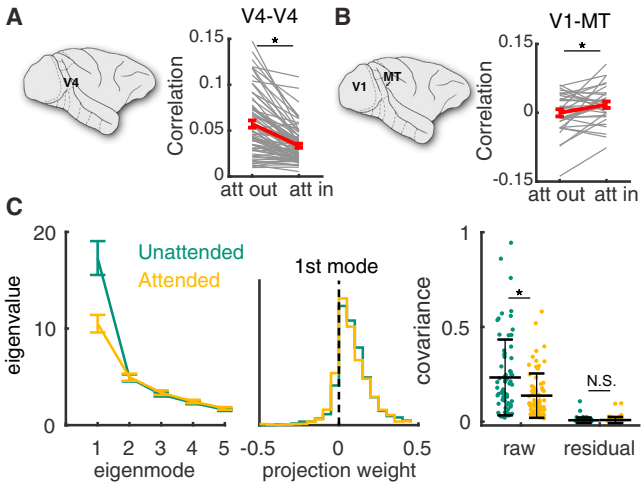

Figure 1. Attentional Modulation of Population Variability within and between Cortical Areas

(A) Mean spike count correlation r_{SC} per session obtained from multi-electrode array recording from V4 was smaller when attention was directed into the receptive fields of recorded neurons (n = 74 sessions, two-sided Wilcoxon rank-sum test between attentional states, p = 3.3 × 10^{-6}; reproduced from Cohen and Maunsell, 2009). Gray lines are individual session comparisons and the red line is the mean comparison across all sessions (error bars represent the SEM).

(B) Same as (A) for the mean spike count correlation r_{SC} between V1 units and MT units per session (n = 32 sessions, paired-sample t test, p = 0.0222; data reproduced from Ruff and Cohen, 2016a).

(C) Left: the first five largest eigenvalues of the shared component of the spike count covariance matrix from the V4 data (Cohen and Maunsell, 2009). Green, unattended; orange, attended; data from n = 72 sessions with 43 ± 15 neurons. Error bars are SEM. Middle: the vector elements for the first (dominant) eigenmode. Right: the mean covariance from each session in attended and unattended states before (raw) and after (residual) subtracting the first eigenmode (mean ± SD in black). Two-sided Wilcoxon rank-sum test (attended versus unattended), mean covariance, p = 1.3 × 10^{-3}; residual, p = 0.75.

simultaneously increasing between area correlations (areas V1 and MT; Ruff and Cohen, 2016a). We show that such differential correlation modulation is a difficult constraint to satisfy with a model in which fluctuations are strictly external to the network. Nevertheless, as discussed above, contemporary recurrent network models are at a loss to explain population-wide variability. A central goal of this study is to put forth a new circuit-based theory of low-dimensional variability in recurrent networks and explore how plausible modulation schemes can differentially control within and between area correlations.

The asynchronous solution of classical balanced networks necessitates that inhibition dynamically tracks and cancels any correlations stemming from recurrent excitation (Renart et al., 2010). This requirement has forced theorists to assume that the time course of inhibitory synapses is faster than that of excitatory synapses (Renart et al., 2010; Rosenbaum and Doiron, 2014; Rosenbaum et al., 2017; Monteforte and Wolf, 2012; van Vreeswijk and Sompolinsky, 1996). However, this is at odds with recorded synaptic physiology, in which excitatory conductances rise and decay faster than inhibitory ones (Geiger et al., 1997; Salin and Prince, 1996; Xiang et al., 1998; Angulo et al., 1999). Recently, we have extended the theory of balanced networks to include a spatial component to network architecture (Rosenbaum et al., 2017; Pyle and Rosenbaum, 2017; Rosenbaum and Doiron, 2014) and found network solutions in which firing rate balance and asynchronous dynamics are decoupled from one another (Rosenbaum et al., 2017). In this study, we consider multi-area models of spatially distributed balanced networks and show that when inhibition has slower kinetics than excitation in these networks, matching physiology, they internally produce low-dimensional population-wide variability. Unlike networks that lack spatial structure, these networks produce spiking activity that robustly captures the rich diversity of firing rate and correlation structure of real population recordings. Further, attention-mediated top-down modulation of inhibitory neurons in our model provides a parsimonious mechanism that controls population-wide variability in agreement with the within and between area experimental results.

There is a long-standing research program aimed at providing a circuit-based understanding for cortical variability (Shadlen and Newsome, 1998; van Vreeswijk and Sompolinsky, 1996; Amit and Brunel, 1997; Rosenbaum et al., 2017; Kass et al., 2018). Our work is a critical advance through providing a mechanistic theory for the genesis, propagation, and modulation of realistic low-dimensional population-wide shared variability based on established circuit structure and synaptic physiology.

RESULTS

Attentional Modulation of Shared Variability within and between Cortical Areas

Multi-electrode recordings from visual area V4 during an orientation change detection task show that the mean spike count correlation coefficient between neuron pairs in V4 is largely reduced when the monkeys were cued to pay attention to the neurons' spatial receptive field (Figure 1A; Cohen and Maunsell, 2009). Recently, simultaneous recordings from two visual areas, MT and V1, during a similar attention task (Ruff and Cohen, 2016a), show that in addition to a reduction of mean spike count correlations between neuron pairs within an area (mean pairwise attention-related MT correlation decrease was 0.019, Wilcoxon rank-sum test, p = 0.017; mean pairwise attention-related V1 correlation decrease was 0.008, Wilcoxon rank-sum test, p = 4.9 × 10^{-6}; Ruff and Cohen, 2016a), there is an attention-mediated *increase* of spike count correlations across areas V1 and MT (Figure 1B). This differential modulation of within and between area correlations offers a strong constraint from which to build circuit models of population-wide variability.

Before we explore population variability in circuit models of cortex, we first quantify how variability is structured across a recorded population. Using dimensionality reduction tools, we partition the V4 covariance matrix into the shared variability among the population and the private noise to each neuron (Cunningham and Yu, 2014; Williamson et al., 2016). The eigenvalues of the shared covariance matrix represent the variance along each dimension (or latent variable), while the corresponding eigenvectors represent the projection weights of the latent variables onto each neuron (STAR Methods). The V4 data show a

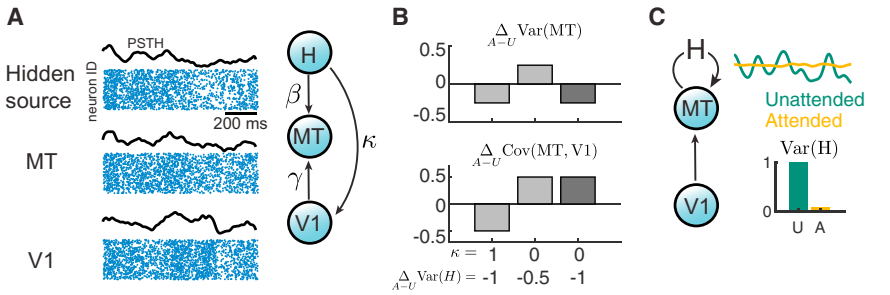

Figure 2. Model Constraints for Shared Variability within and between Areas

(A) Left: hidden variable model for connected cortical areas, V1 and MT, where the response variability of MT comes from both its upstream area V1 and a hidden source H. Due to the low-dimensional structure of shared variability in population activity (Figure 1C), we use the mean population rate (black curves) to represent the population spiking activity from each area (blue dot rasters). Right: the hidden source H projects to MT and V1 with strengths β and κ, respectively. The feedforward projection strength from V1 to MT is γ.

(B) Examples of attentional changes in the variance of MT, $\underset{A-U}{\Delta}\mathrm{Var}(\mathrm{MT})$, and the covariance between MT and V1, $\underset{A-U}{\Delta}\mathrm{Cov}(\mathrm{MT},\mathrm{V1})$. We consider combinations of shared H ($\kappa = 1$) versus private H ($\kappa = 0$) and a moderate reduction in hidden variability $\left(\underset{A-U}{\Delta}\mathrm{Var}(\mathrm{H}) = -0.5\right)$ versus a large reduction $\left(\underset{A-U}{\Delta}\mathrm{Var}(\mathrm{H}) = -1\right)$. Attention-mediated simultaneous decreases in $\mathrm{Var}(\mathrm{MT})$ and increase in $\mathrm{Cov}(\mathrm{MT},\mathrm{V1})$ occur for private variability with a large reduction in hidden variability (dark gray). The other combinations cause a shift in the same direction for within and between area variability (light gray). Other model parameters are $\gamma^U = 0.5$, $\gamma^A = 1$, $\mathrm{Var}^U(\mathrm{H}) = 1$, $\beta = 1$, and $\mathrm{Var}(\mathrm{V1}) = 1$, independent of attentional state. U, unattended; A, attended. For general analysis, see Methods S1.

(C) The differential modulation of shared variability within and between areas (Figures 1A and 1B) suggests the hidden variable H is internally generated within area MT and that attention should quench the variance of H substantially.

single dominant eigenmode (Figure 1C, left; for single-session results, see Figure S1), indicating a primarily one-dimensional latent structure in the population variability. The projection weight of the dominant eigenmode onto the individual neurons is primarily of the same sign (Figure 1C, middle, weights are dominant positive), meaning that the latent variable causes positive correlations across the population. Indeed, after subtracting the first eigenmode the mean residual covariances are very small (Figure 1C, right). Moreover, attention affects population-wide variability primarily by quenching this dominant eigenmode (Figure 1C, left, orange versus green) and the attentional modulation in the dominant eigenmode is highly correlated with the modulation in mean covariance (Figure S1C). The low-dimensional structure of shared variability in our data is consistent with similar analysis in other cortices (Williamson et al., 2016; Lin et al., 2015; Ecker et al., 2014), as well as alternative analysis of the same V4 data using generalized point process models (Rabinowitz et al., 2015).

Constraints for Circuit-Based Models of Shared Variability

Armed with the V4 and V1-MT population analysis, we next explore the constraints that circuit models must satisfy in order to capture attentional modulation that differentially modulates the shared variability within and between cortical areas. Since the population-wide fluctuations are well described by a single latent variable that influences all neurons (Figure 1C), we represent the aggregate population responses with scalar random variables: MT and V1 (Figure 2A).

To begin, we assume that population responses are linear in their inputs and that V1 projects to MT with strength γ. We suppose a hidden source of variability, H, which projects to MT and V1 with strength β and κ, respectively (Figure 2A); without loss of generality, we take $\beta = 1$. In total, we have $\mathrm{MT} = \gamma\mathrm{V1} + \mathrm{H}$ and $\mathrm{V1} = X_0 + \kappa\mathrm{H}$, where X_0 is independent from H. We assume that attention acts to reduce the variability of the hidden variable, $\mathrm{Var}(\mathrm{H})$, and to increase the coupling strength γ.

We first consider how attention affects the covariance between MT and V1 in our model; our linear system gives the following:

$$\mathrm{Cov}(\mathrm{MT},\mathrm{V1}) = \gamma\mathrm{Var}(\mathrm{V1}) + \kappa\mathrm{Var}(\mathrm{H}). \qquad \text{(Equation 1)}$$

For $\kappa > 0$, an attention-mediated reduction of $\mathrm{Var}(\mathrm{H})$ acts to reduce the covariance between V1 and MT. This is at odds with our cortical recordings (Figure 1B). We explore the case in which $\kappa = 0$ to circumnavigate the tension between an increase in γ and a decrease in $\mathrm{Var}(\mathrm{H})$. In effect, this assumes that MT has a source of variability that is private from V1.

With $\kappa = 0$, the variance of the MT population obeys

$$\mathrm{Var}(\mathrm{MT}) = \gamma^2\mathrm{Var}(\mathrm{V1}) + \mathrm{Var}(\mathrm{H}). \qquad \text{(Equation 2)}$$

The contributions to MT variability from the upstream area V1 and the hidden source H are clear. Further, attention drives opposing influences on $\mathrm{Var}(\mathrm{MT})$ through an increase in γ being countered by a decrease in $\mathrm{Var}(\mathrm{H})$. However, unlike the case of $\mathrm{Cov}(\mathrm{MT},\mathrm{V1})$, we cannot simply choose γ to be zero to mitigate this competition (because V1 would then not drive MT). Indeed, if the decrease in $\mathrm{Var}(\mathrm{H})$ is only moderate, then attention will increase both $\mathrm{Cov}(\mathrm{MT},\mathrm{V1})$ and $\mathrm{Var}(\mathrm{MT})$ (Figure 2B, middle), again at odds with experiments (Figures 1A and 1B). In total, with an assumption of H being private to MT ($\kappa = 0$), then to have a reduction in $\mathrm{Var}(\mathrm{MT})$ combined with an increase in $\mathrm{Cov}(\mathrm{MT},\mathrm{V1})$, we require that the attention-mediated suppression in the variability of H be large (Figure 2B, right). These arguments can be generalized over a range of parameters (Methods S1; Figure S2).

In sum, we have exposed three constraints that, if satisfied, will cause cortical circuit models of population-wide variability to capture the differential modulation of within and between area variability.

(1) The shared variability across a neuronal population is low dimensional.

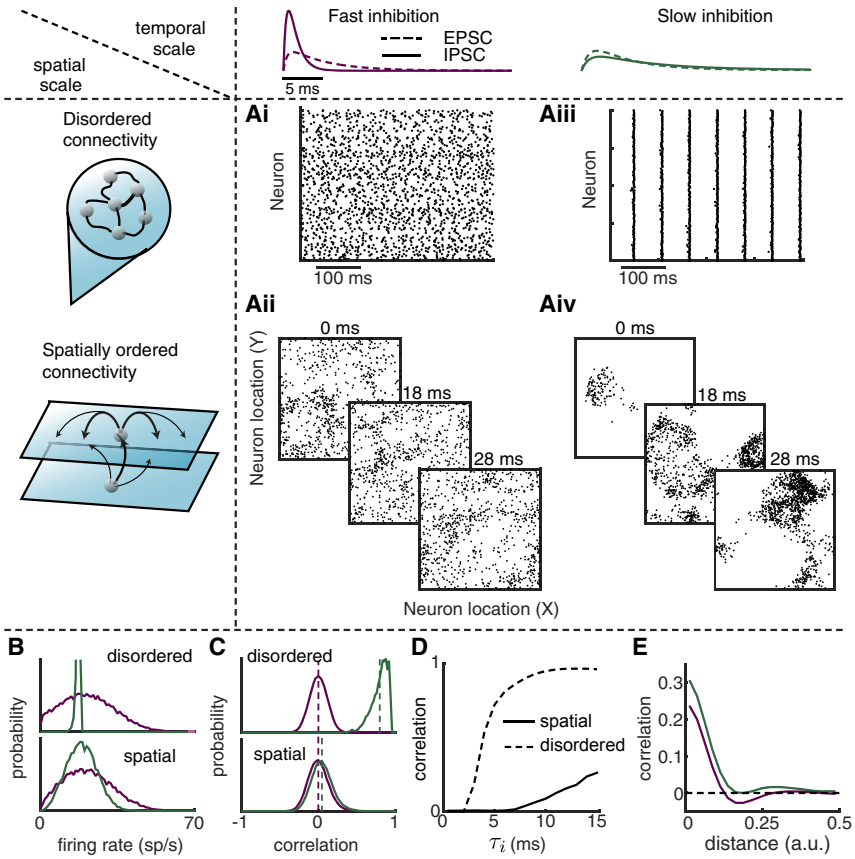

Figure 3. The Spatial and Temporal Scales of Synaptic Coupling Determine Internally Generated Variability

(A) Networks of excitatory and inhibitory neuron models were simulated with either disordered connectivity (Ai and Aiii) or spatially ordered connectivity (Aii and Aiv), and with either fast inhibition ($\tau_i = 1$ ms; Ai and Aii) or slow inhibition ($\tau_i = 8$ ms; Aiii and Aiv). The integral of inhibitory postsynaptic current over time is conserved as we change τ_i. In all models the timescale of excitation was $\tau_e = 5$ ms. In the disordered networks, spike train rasters assume no particular neuron ordering. In the spatially ordered networks, three consecutive spike raster snapshots are shown with a dot indicating that the neuron at spatial position (x, y) fired within 1 ms of the time stamp.

(B) Distributions of firing rates of excitatory neurons in the disordered (top) and spatially ordered (bottom) models, with faster inhibitory kinetics (purple) compared to slower inhibitory kinetics (green).

(C) Same as (B) for the distributions of pairwise correlations among the excitatory population.

(D) Mean correlation among the excitatory population as a function of the inhibitory decay time constant (τ_i).

(E) Pairwise correlation as a function of distance between neuron pairs for spatially ordered models with slower inhibitory kinetics (green) compared to faster inhibitory kinetics (purple).

(2) There is a source of attention-mediated population-wide variability in downstream areas that is private from upstream areas.
(3) The attention-mediated suppression of the private variability needs to be substantial.

The second constraint could be produced by each cortical population being paired with an external variability source that projects exclusively to that area. This solution requires strong assumptions about how the cortex devotes and organizes biological resources to drive neuronal variability. In contrast, we explore the more parsimonious hypothesis whereby recurrently coupled networks produce low-dimensional variability through internal interactions, and hence variability is private to the population by construction (Figure 2C). Further, if the variability is internally generated, then the third constraint requires a strong nonlinearity to fully suppress variability in the attended state. In the next sections, we investigate how a physiologically realistic network of spiking neuron models can satisfy these three constraints.

Population-wide Correlations with Slow Inhibition in Spatially Ordered Networks

Networks of spiking neuron models where strong excitation is balanced by opposing recurrent inhibition internally produce high single-neuron variability (Figure 3Ai) with a broad distribution of firing rates (Figure 3B, top purple curve; van Vreeswijk and Sompolinsky, 1996; Amit and Brunel, 1997; Renart et al., 2010). However, these networks enforce an asynchronous solution (Figure 3C, top purple), and as such fail to explain population-wide shared variability (Renart et al., 2010; Williamson et al., 2016). Typically, balanced networks have disordered connectivity, where connection probability is uniform between all neuron pairs. This approximation ignores the abundant evidence that cortical connectivity is spatially ordered with a connection probability falling off with the physical distance between neuron pairs (Levy and Reyes, 2012; Horvát et al., 2016; Mariño et al., 2005). Recently we and others have extended the theory of balanced networks to include such spatially dependent connectivity (Rosenbaum et al., 2017; Rosenbaum and Doiron, 2014; Darshan et al., 2018; Pyle and Rosenbaum, 2017). Briefly, we model a two-dimensional lattice of integrate-and-fire neurons, meaning neuron locations tile a space with x and y coordinates. Each neuron receives both feedforward projections from a layer of external Poisson processes and recurrent projections from within the network (STAR Methods); the connection probability of all projections decays like a Gaussian with distance. If the spatial scale of feedforward inputs is narrower than the scale of recurrent projections, the asynchronous state no longer exists (Rosenbaum et al., 2017), giving way to a solution with spatially structured correlations (Figures 3Aii and 3D, purple; Video S1). Nevertheless, the mean correlation across all neuron pairs vanishes for large network size (Figure 3C, bottom purple curve), in stark disagreement with a majority of experimental studies

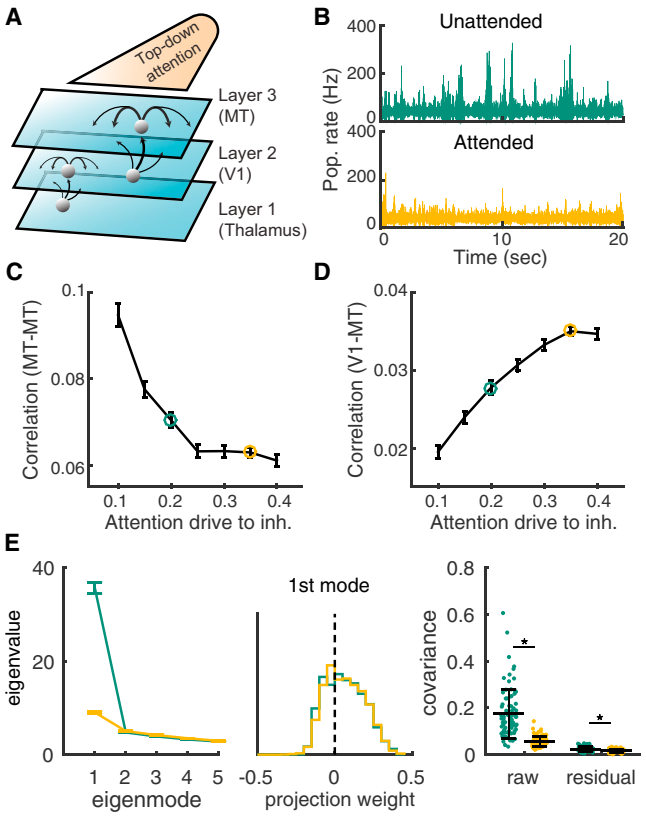

Figure 4. Top-Down Depolarization of MT Inhibitory Neurons Captures the Differential Attentional Modulation of Shared Variability within and across V1 and MT

(A) Thalamus, V1, and MT are modeled in a three-layer hierarchy of spatially ordered balanced networks. Top-down attentional modulation is modeled as a static depolarizing current, μ_I, to MT inhibitory neurons. In both V1 and MT the recurrent projections are broader than feedforward projections (V1, $\alpha^{(2)}_{\text{ffwd}} = 0.05$, $\alpha^{(2)}_{\text{rec}} = 0.1$; MT, $\alpha^{(3)}_{\text{ffwd}} = 0.1$, $\alpha^{(3)}_{\text{rec}} = 0.2$) and recurrent inhibition is slower than excitation ($\tau_I = 8$ ms, $\tau_e = 5$ ms).

(B) Population averaged firing rate fluctuations from MT in the unattended state ($\mu_I = 0.2$, green) and the attended state ($\mu_I = 0.35$, orange).

(C) Mean spike count correlation (r_{SC}) of excitatory neuron pairs in MT decreases with attentional modulation.

(D) Mean r_{SC} between the excitatory neurons in MT and the excitatory neurons in V1 increases with attention. Error bars are SEM.

(E) Left: the first five largest eigenvalues of the shared component of the spike count covariance matrix. Green, unattended; orange, attended; n = 80 sessions with 50 neurons each. Error bars are SEM. Middle: the vector elements for the first (dominant) eigenmode. Right: the mean covariance from each session in attended and unattended states before (raw) and after (residual) subtracting the first eigenmode (mean ± SD in black). Two-sided Wilcoxon rank-sum test (attended versus unattended), mean covariance, $p = 1.3 \times 10^{-21}$; residual, $p = 3.5 \times 10^{-8}$.

(Cohen and Kohn, 2011; Doiron et al., 2016) as well as with our motivating population data (Figures 1A and 1B).

Many previous balanced network models assume that the kinetics of inhibitory synaptic currents are *faster* (or at least not slower) than those of excitatory currents (Renart et al., 2010; van Vreeswijk and Sompolinsky, 1996; Lim and Goldman, 2014; Amit and Brunel, 1997), including our past work (Rosenbaum et al., 2017; Rosenbaum and Doiron, 2014; Pyle and Rosenbaum, 2017). However, this assumption is at odds with physiology in which excitatory α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors have faster kinetics than those of the inhibitory γ-aminobutyric acid (GABAa) receptors (Geiger et al., 1997; Xiang et al., 1998; Salin and Prince, 1996; Angulo et al., 1999). When the timescales of excitatory and inhibitory synaptic currents match experimental values in networks with disordered connectivity, the activity becomes pathologic, with homogeneous firing rates (Figure 3B, top green) and excessive synchrony (Figures 3Aiii and 3C, top green), as has been previously remarked (Börgers and Kopell, 2005). This consequence is likely the *ad hoc* justification for the faster inhibitory kinetics in disordered balanced model networks.

When the spatially ordered model has synaptic kinetics that match physiology, a population-wide turbulent dynamic emerges (Figure 3Aiv; Video S2). This dynamic produces a small, but non-zero, mean pairwise spike count correlation across the population ($r_{SC} = 0.04$), comparable to experiment (Figure 1A, right). Further, both firing rates and pairwise correlations are broadly distributed (Figures 3B and 3C, bottom green curves). Low (but significant) correlation is a robust feature of spatially ordered networks. This is clear from the gradual rise in mean correlation with inhibitory timescale for the two-dimensional spatially ordered network (Figure 3D, solid curve), in contrast to the rapid rise of correlation to pathologic correlation in the disordered network (Figure 3D, dashed curve). Further, this weak sensitivity of correlations on inhibitory timescale is restricted to networks with two spatial scales (i.e., x and y), since networks constrained to one spatial dimension (i.e., neurons are arranged on a ring) also show excessive synchrony when the inhibitory timescale is large (Figure S3).

Our previous work that studied a spatially ordered model with fast inhibition identified a signature spatial organization for correlation, namely positive for nearby neuron pairs and negative for farther away neuron pairs, so that the overall correlation was small (Rosenbaum et al., 2017). In contrast, the spatial model with slow inhibition has positive net correlation across all pair distances (Figure 3E, green)—this is critical for the mean r_{SC} to be positive. In sum, when realistic spatial synaptic connectivity is paired with realistic temporal synaptic kinetics in balanced networks, internally generated population dynamics produce spiking dynamics whose marginal and pairwise variability conform to experimental results. The spatially ordered model with slow inhibition is thus well positioned to satisfy the three constraints required to match how attention modulates within and between area correlations.

Attentional Modulation of Low-Dimensional Population-wide Variability

We model the V1 and MT network by extending our spatially ordered balanced networks with slow inhibition to include three layers: a bottom layer of independent Poisson processes modeling thalamus, and middle and top layers of integrate-and-fire neurons modeling V1 and MT, respectively (Figure 4A; STAR Methods). We follow our past work with simplified firing rate networks (Kanashiro et al., 2017) and model a top-down attentional signal as an overall static depolarization to inhibitory

neurons in the MT layer (Figure 4A). This mimics cholinergic pathways that primarily affect interneurons (Kuchibhotla et al., 2017; Kim et al., 2016) and are thought to be engaged during attention (Schmitz and Duncan, 2018). The increased recruitment of inhibition during attention reduces the population-wide fluctuations in the MT layer (Figure 4B) and decreases pairwise spike count correlations of MT-MT neuron pairs (Figure 4C), while simultaneously increasing the correlation of V1-MT neuron pairs (Figure 4D). Further, neuron pairs with larger firing rate increases also show larger correlation reductions (Figure S4), in agreement with population recordings during both spatial and feature attention (Cohen and Maunsell, 2011). Finally, there is a slight attention-mediated decrease of both the average firing rates of MT neurons (~3%) and MT neuron spike count Fano factor (~20%) (Figure S5). In total, our model and its simple implementation of attentional modulation capture the main aspects of the pairwise co-variability in the V1-MT dataset (Figure 1B). However, it remains to show that our model does this by satisfying the three constraints identified with our heuristic model (Figure 2).

The first model constraint is that the population-wide variability must be low dimensional. We analyzed the spike count covariance matrix constructed from a subsampling of the spike trains in the third layer of our network model (n = 50 neurons). The network with slow inhibition produces shared variability with a clear dominant eigenmode that mimicked many of the core features observed in the V4 data (Figure 4E compared with Figure 1C). The projection weights of the dominant mode are mostly positive (Figure 4E, middle) and the distribution of mean covariance across sessions is skewed to the positive side (Figure 4E, right). Removing the first eigenmode also results in small residual covariance (Figure 4E, right). Further, the top-down attentional modulation of inhibition also suppresses this dominant eigenmode (Figure 4E, right, orange versus green).

The agreement between model and data breaks down when inhibitory temporal kinetics and the spatial wiring structure are changed. When the model has fast inhibition, the shared variability does not have a dominant eigenmode (Figure S6A), the raw mean correlation coefficient is near zero (Figure S6C), and attentional modulation has a negligible effect on population variability (Figures S6A–S6C, orange versus green). Experimental measurements of local cortical circuitry show that excitation and inhibition project on similar spatial scales (Levy and Reyes, 2012; Mariño et al., 2005). Inhibitory projections that are broader than excitatory produce strong positive and negative correlations within the network, owing to a competitive dynamic across the network. The resultant population-wide correlations are not low dimensional (Figure S6F), while still being high in magnitude, as has been noted in past work from spiking networks with lateral inhibition (Keane and Gong, 2015; Pyle and Rosenbaum, 2017; Williamson et al., 2016). Nonetheless, as in the case with fast inhibition, the mean correlation coefficient is near zero (Figure S6H), and attentional modulation has only a negligible effect (Figures S6F–S6H, orange versus green).

In sum, satisfying our first constraint of shared variability having low-dimensional structure over the population requires inhibition that is neither faster nor anatomically broader than excitation—both features of real cortical circuits (Levy and Reyes, 2012; Mariño et al., 2005; Salin and Prince, 1996; Geiger et al., 1997; Xiang et al., 1998; Angulo et al., 1999). Further, a simple recruitment of inhibition through top-down drive can restore stability and quench low-dimensional population variability.

Relating Low-Dimensional Variability to Spatiotemporal Pattern Formation

To provide intuition about how recurrent circuitry shapes low-dimensional shared variability, we considered a firing rate model that incorporates both the spatial architecture and synaptic dynamics that are central to our spiking model (STAR Methods). While firing rate models lack a principled connection to spiking network models, they do produce qualitatively similar dynamics in recurrent networks and their simplicity makes them amenable to analysis techniques from dynamical systems theory (Ermentrout, 1998).

Solutions in which firing rates are constant over time are interpreted as asynchrony within the network, since only dynamical co-fluctuations in firing rates would model correlated spiking. We focus on how the stability of the asynchronous firing rate solution depends upon the temporal (τ_i) and spatial (σ_i) scales of inhibition. A firing rate solution is stable if the linearized dynamics are such that every eigenmode has eigenvalues with strictly negative real part. Since our network is spatially ordered, the eigenmodes are also organized in space, each with their own distinct wave number (spatial frequency). If the solution loses stability at a particular eigenmode, then the spatiotemporal dynamics of the resulting network firing rates will inherit the spatial scale of that eigenmode—this process is termed spatiotemporal pattern formation (Cross and Hohenberg, 1993).

If τ_i and σ_i are near those of recurrent excitation, then a stable firing rate solution exists (Figure 5A, gray region; Figure 5B, top left, black curve with τ_i = 5ms). Our past work explored activity within this regime (Rosenbaum et al., 2017). When τ_i increases and excitation and inhibition project with the same spatial scale $(\sigma_i = \sigma_e)$, firing rate stability is first lost at an eigenmode with zero wave number (Figure 5B, left). This creates population dynamics with a broad spatial pattern, allowing variability to be shared over the entire network. Simulations of the three-layered spiking network model in this regime show turbulent dynamics that extend across the entire network (Figure 5B, right; Video S3). In contrast to this case, when τ_i increases yet inhibition projects lateral to excitation $(\sigma_i > \sigma_e)$, stability is first lost at a non-zero wave number (Figure 5C, left). This creates population dynamics with coherence over a band of higher spatial frequencies, producing higher dimensional shared variability, as evident in the spatially patchy spiking dynamics of the three-layered spiking network in this regime (Figure 5C, right; Video S4). Thus, the spatial and temporal scales of inhibition determine in large part the spatiotemporal patterns of network activity.

In the firing rate network, we can also model attention as a depolarization to the inhibitory neurons, as was done in the network of spiking neuron models. In the firing rate network, attentional modulation expanded the stable region in the bifurcation diagram (Figure 5A, dashed black line). In other words, attention increased the domain of firing rate stability. Thus, with $\tau_i > \tau_e$ chosen so that in the unattended state the network was unstable at a low spatial frequency yet with attention the network was in

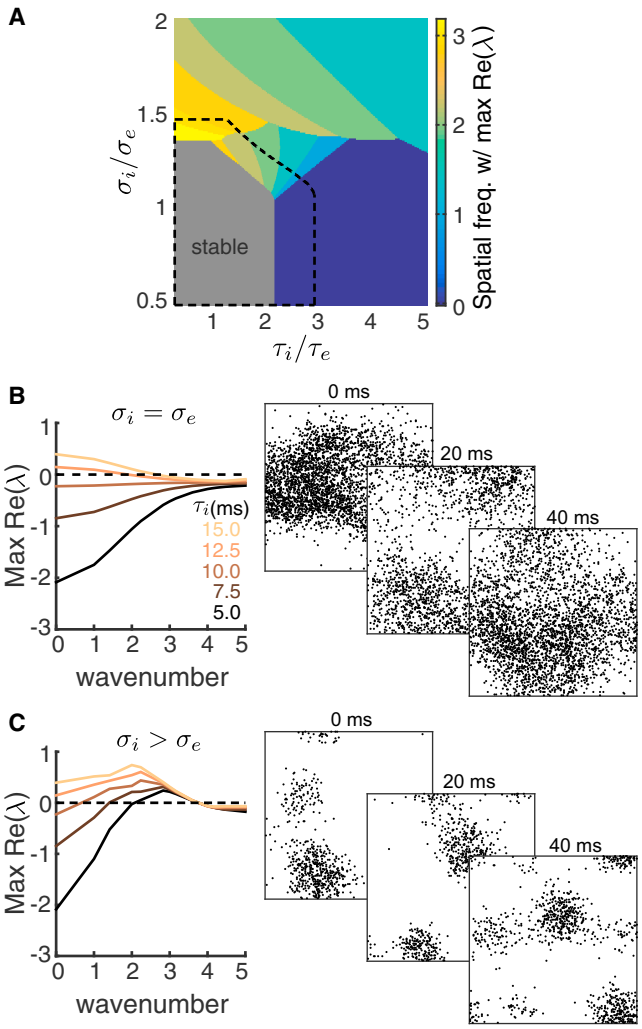

Figure 5. Stability Analysis of a Two-Dimensional Firing Rate Model

(A) Bifurcation diagram of a firing rate model as a function of the inhibitory decay timescale τ_i and inhibitory projection width σ_i. The excitatory projection width and time constant are fixed at $\sigma_e = 0.1$ and $\tau_e = 5$ ms, respectively. Color represents the wavenumber with the largest real part of eigenvalue and the gray region is stable. Top-down modulation of inhibitory neurons modeling attention expands the stable region (black dashed).

(B) Left: the real part of eigenvalues as a function of wavenumber for increasing τ_i when $\sigma_i = \sigma_e$. Right: three consecutive spike raster snapshots of a spiking neuron network with $\sigma_i = \sigma_e$ and slow inhibition (same network as in Figure 4 in the unattended state).

(C) Same as (B) for σ_i larger than σ_e. Right: spike raster snapshots of a spiking neuron network with broad inhibitory projections; the excitatory and the inhibitory projection widths of layer 3 were $\alpha_e^{(3)} = 0.1$ and $\alpha_i^{(3)} = 0.2$, respectively. Other parameters were the same as in (B).

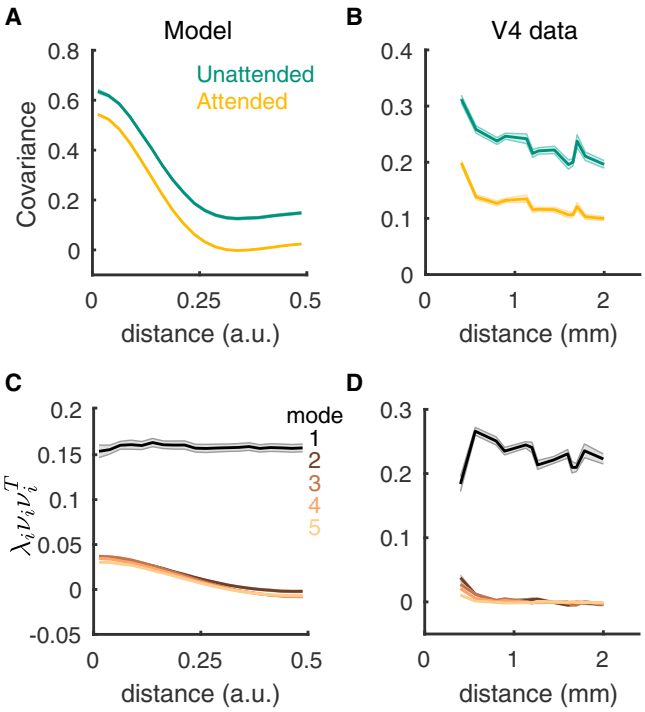

Figure 6. Distance Dependence of Pairwise and Population-wide Variability

(A) Pairwise covariance of spike counts from our spiking model as a function of the distance between the neurons.

(B) Same as (A) but for the V4 data. Here the distance is between the electrodes that recorded the neuron pair.

(C) The distance dependence functions of the first five covariance components computed from factor analysis of the model spiking activity in the unattended state. For mode *i* the product of the eigenmode loading onto a pair of neurons is plotted as a function of the distance between the neurons. To properly compare the modes, we scaled each curve by the eigenvalue λ_i for that mode.

(D) Same as (C) but for the V4 data in the unattended state. Shaded regions are SEM. See Table S2 for the number of pairs at each distance value for the V4 data; for the model we used n = 80 sessions of 500 neurons each.

the stable regime, our model captures the large attention-mediated quenching of population-wide shared variability reported in the population recordings (Figure 1C) and network of spiking neuron models (Figure 4E).

In general, dynamical systems can transition through an instability point by changing any one of many model parameters. This suggests that tonic drive to inhibitory neurons is not the only mechanism that can capture the neuronal correlates of attentional modulation. For example, by reducing the strength of recurrent excitation, providing direct hyperpolarization to excitatory neurons, among other cellular and circuit modulations, our model can mimic the shift in network stability achieved through top-down drive to inhibition (Figure S7). However, despite these differing biophysical models of attention, the mechanisms all share an attention-mediated shift toward inhibition stabilizing runaway excitation.

Finally, our model predicts that population-wide variability is due to a dynamical instability that propagates spiking activity broadly over space. The sparse sampling of spiking activity in our neuronal recordings (<100 neurons spanning a few square millimeters of cortical tissue) makes a direct test of this prediction difficult. The pairwise covariance from both our model simulations (Figure 6A) and V4 data (Figure 6B) decreases with the distance between neurons in both the attended and unattended states. The large covariance for nearby neurons in the model is for neuron pairs that are within one spatial footprint of

the excitatory and inhibitory coupling (distances < 0.25 in Figure 6A); the 400 μm electrode spacing in the V4 data does not permit a sampling of small distances between neurons. For the larger distances the decrease in covariance is gradual, consistent with a previous V4 population dataset (Smith and Sommer, 2013). Further, there is a near spatially uniform reduction of covariance by attention in both the model and V4 data (Figures 6A and 6B, compare orange and green). While these agreements between model and data are promising, a simple decay of covariance with distance can be replicated with many different models, notably our model with fast (Figure S6D) or broad (Figure S6I) inhibition.

A more stringent test of our model is to compare how the spatial dependence of pairwise covariance decomposes over the low-dimensional latent variable space. The loss of stability at the zero spatial Fourier mode in our model produces global fluctuations over the network (Figure 5B). In the factor analysis of the shared covariance, this is reflected by the dominant eigenmode being uniform across neuronal space (Figure 6C, black curve). In other words, this dominant latent variable projects to all neurons irrespective of their location and drives global correlations across the network. By contrast, the higher eigenmodes contain covariance structure that is spatially localized (Figure 6C, colored curves). This feature is specific to our model, since there is no spatial invariance of the dominant mode in the spiking network when inhibition is either fast (Figure S6E) or spatially broad (Figure S6J). Analysis of the V4 data clearly identifies a spatial invariance of the dominant eigenmode (Figure 5D, black curve), validating our model prediction.

Chaotic Population-wide Dynamics Reflect Internally Generated Variability

The attention-mediated differential modulation of within and between area correlations (Figure 1) lead us to propose our second and third model constraints—that shared variability has a sizable internally generated component and that attention must quench this variability. The third constraint requires that the mechanisms that produce internally generated variability sensitively depend on top-down modulations. The firing rate model captured this sensitivity through a spatiotemporal pattern-forming transition in network activity. However, the firing rate model does not internally produce trial-to-trial variability that can be compared to experiment, and we thus return to analysis of the network of spiking neuron models to probe how trial-to-trial variability is internally generated through recurrent coupling.

To isolate the sources of externally and internally generated fluctuations in the third layer of our network, we fixed the spike train realizations from the first layer (thalamic) neurons as well as the membrane potential states of the second layer (V1) neurons, and only the initial membrane potentials of the third layer (MT) neurons were randomized across trials (Figure 7A). This produced deterministic network dynamics when conditioned on activity from the first two layers, and consequently any trial-to-trial variability is due to mechanics internal to the third layer.

The spike trains from third layer neurons in both the unattended and attended states have significant trial-to-trial variability despite the frozen layer one and two inputs. This is reflective of a well-studied chaotic network dynamic in balanced networks in which the spike times from individual neurons are very sensitive to perturbations that affect the spiking of other neurons (Monteforte and Wolf, 2012; London et al., 2010). To investigate how this microscopic (single neuron) variability possibly manifests as macroscopic population activity, we considered the trial-to-trial variability of the population-averaged instantaneous firing rate. While the population firing rate is dynamic in the attended state, there is very little variability from trial to trial (Figure 7B, left; Figure 7D, orange). A consequence of this low population-wide variability is the faithful tracking of the spatiotemporal structure of layer two outputs by layer three responses (Figure 7B, right; Figure 7E, orange). This tracking reflects the higher correlation between layer two and three spiking in the attended state (Figure 4D), and represents the attention-mediated increase in V1 to MT coupling (γ) in our linear model (Figure 2).

In contrast, in the unattended state there is significant population-wide recruited activity. The periods of spiking coherence across the network are not trial locked and rather contribute to sizable trial-to-trial variability of population activity (Figure 7C, left; Figure 7D, green). This degrades the tracking of layer two outputs (Figure 7C, right; Figure 7E, green) and ultimately lowers the correlation between layer two and three spiking (Figure 4D). Taken together, while the network model is chaotic in both the attended and unattended states, the chaos is population-wide only in the inhibition-deprived unattended state. Furthermore, since the population-wide variability is internally generated in the MT layer, our framework satisfies our second model constraint of private variability.

The nonlinear pattern-forming dynamics of the spatially distributed recurrent network impart extreme sensitivity to the population-wide internally generated variability. Indeed, in our model the trial-to-trial population rate variability is almost extinguished with attention (Figure 7D, right). In our heuristic model with hidden variable H this amounts to $\mathrm{Var}(H)$ reducing drastically with attention, which is precisely what is needed to account for the differential modulation of within and between area correlations (compare Figure 2B with Figure 7D, right). Thus, our model satisfies the third constraint we derived from our hidden variable model, namely that the shared variability in MT should be substantially quenched by attention.

DISCUSSION

There is a long-standing research program aimed at understanding how variability is an emergent property of recurrent networks (van Vreeswijk and Sompolinsky, 1996; Amit and Brunel, 1997; Monteforte and Wolf, 2012; London et al., 2010; Rosenbaum et al., 2017; Rosenbaum and Doiron, 2014). However, models are often restricted to simple networks with disordered connectivity. Consequently, in these networks population-wide activity is asynchronous, at odds with many experimental findings (Cohen and Kohn, 2011; Doiron et al., 2016). A parallel stream of research focuses on spatiotemporal pattern formation in neuronal populations, with a rich history in both theoretical (Ermentrout, 1998) and experimental contexts (Sato et al., 2012). Yet a majority of these studies consider only trial-averaged activity, with tacit assumptions about how spiking

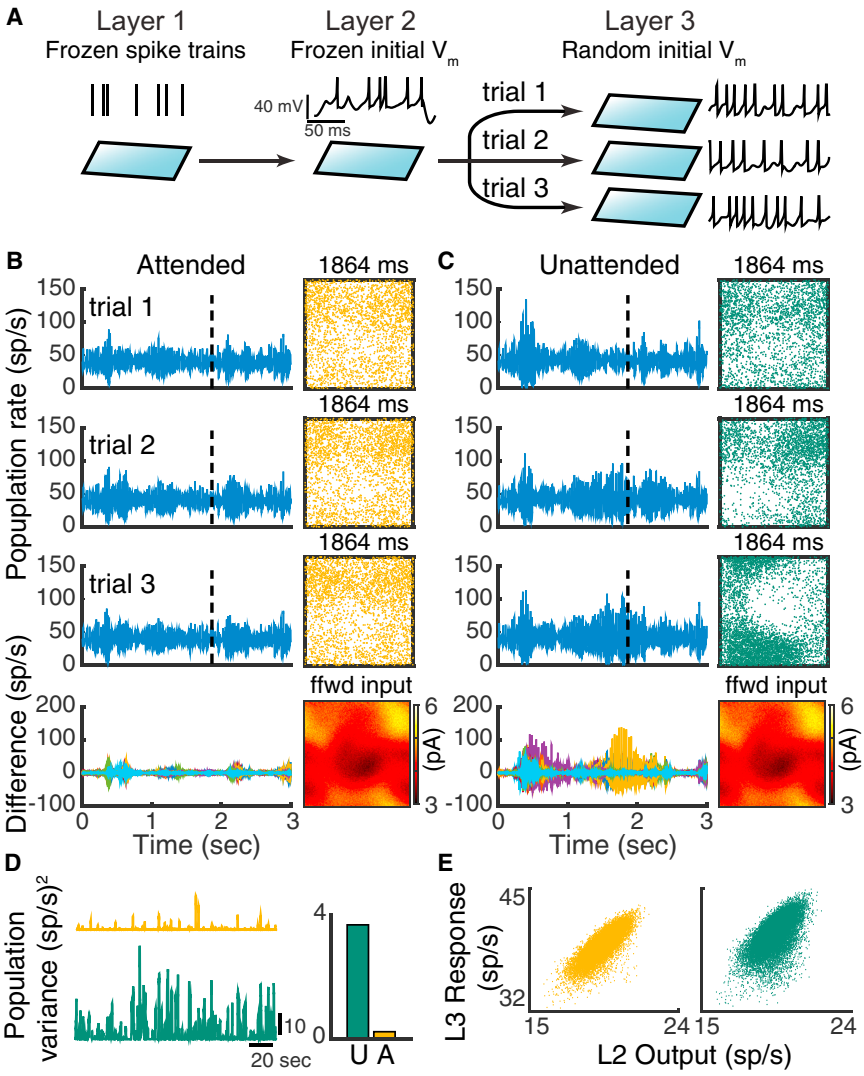

Figure 7. Chaotic Population Firing Rate Dynamics Are Quenched by Attention

(A) Schematic of the numerical experiment. The spike train realizations in layer one and the initial states of the membrane potential of layer two neurons are identical across trials, while in each trial we randomized the initial states of the layer three neuron's membrane potentials.

(B) Three representative trials of the layer three excitatory population rates in the attended state (left, rows 1–3). Bottom row: difference of the population rates across 20 trials. Right (rows 1–3): snapshots of the neuron activity at time point 1,864 ms. Each dot is a spike within 2 ms window from the neuron at that location. Right bottom: the synaptic current each layer three neuron receives from layer two at time 1,864 ms.

(C) Same as (B) for the network in the unattended state.

(D) Trial-to-trial variance of layer three population rates as a function of time. Right: mean variance across time.

(E) The layer three population rate tracks the layer two population rate better in the attended state. Both outputs and responses are smoothed with a 200 ms window.

variability emerges (but see Keane and Gong, 2015 and Rosenbaum et al., 2017). In this study we combined these modeling traditions with the goal of circuit-based understanding of the genesis and modulation of low-dimensional internally generated shared cortical variability.

Population-wide Variability in Balanced Networks

Our model extends classical work in balanced cortical networks (van Vreeswijk and Sompolinsky, 1996; Renart et al., 2010) to include two well-accepted experimental observations. First, cortical connectivity has a wiring rule that depends upon the distance between neuron pairs (Horvát et al., 2016; Levy and Reyes, 2012). Theoretical studies that model distance-dependent coupling commonly assume that inhibition projects more broadly than excitation (Ermentrout, 1998; Compte et al., 2003; Keane and Gong, 2015). However, measurements of local cortical circuitry show that excitation and inhibition project on similar spatial scales (Levy and Reyes, 2012; Mariño et al., 2005), and long-range excitation is known to project more broadly than inhibition (Bosking et al., 1997). Our work shows that local inhibitory projections are required for internally generated population variability to be low dimensional (Figure S6F).

The second observation is that inhibition has temporal kinetics that are slower than excitation (Salin and Prince, 1996; Geiger et al., 1997; Angulo et al., 1999; Xiang et al., 1998). Past theoretical models of recurrent cortical circuits have assumed that inhibition is not slower than excitation (Renart et al., 2010; van Vreeswijk and Sompolinsky, 1996; Stringer et al., 2016; Lim and Goldman, 2014), including past work from our group (Rosenbaum et al., 2017; Rosenbaum and Doiron, 2014; Pyle and Rosenbaum, 2017). Consequently, these studies could only capture the residual correlation structure of population recordings once the dominant eigenmode was subtracted (Rosenbaum et al., 2017; Williamson et al., 2016). When inhibition has kinetics that are slower than excitation, the asynchronous solution is unstable. In disordered networks with strong coupling, this causes pathologic levels of rhythmic synchrony (Figure 3Aiii), often requiring sources of external variability that are independent over neurons to tame network activity (Börgers and Kopell, 2005). In contrast, we have shown that networks with slow inhibition and neuronal coupling that depend upon two spatial dimensions produce spiking dynamics that are only weakly correlated, with firing rate and correlation values that match experiment (Figures 3Aiv and 4). In total, by including accepted features of cortical anatomy and physiology, long ignored by theorists, our model network recapitulates low-dimensional

population-wide variability to a much larger extent than previous models.

The above narrative is somewhat revisionist; there are several well-known theoretical studies in disordered networks in which one-dimensional population-wide correlations do emerge, notably in networks where rhythmic (Amit and Brunel, 1997) or "up-down" (Compte et al., 2003; Stringer et al., 2016) dynamics are prominent. Networks with dense yet disordered connectivity ensure that all neuron pairs receive some shared inputs from overlapping presynaptic projections. In such a network, if the asynchronous state becomes unstable then this shared wiring will correlate spiking activity across the entire network. In other words, any shared variability will be one dimensional (scalar) by construction. In contrast, the ordered connectivity in our network is such that neuron pairs that are distant from one another have no directly shared presynaptic connections. Consequently, when asynchrony is unstable, one-dimensional population dynamics are not preordained; rather, the spatial network can support higher dimensional shared variability depending on the temporal and spatial scales of recurrent coupling (Figures S6 and 5). From the vantage of this model, we discovered the conditions for recurrent architecture and synaptic physiology for low-dimensional shared variability.

Recently, several studies have shown that networks of firing rate models with a low-rank perturbation of the recurrent connectivity structure can exhibit low-dimensional coherent chaotic dynamics (Mastrogiuseppe and Ostojic, 2018; Landau and Sompolinsky, 2018). Such a low-rank connectivity effectively embeds a feedforward loop in the network and thus drives the population activity in one direction. However, such a low-rank connectivity requires each neuron projects a component of its activity to the whole population, a circuit assumption that lacks biological evidence. In contrast, our network simply incorporates spatially ordered connectivity, which has been commonly observed in most cortical areas (Horvát et al., 2016). The recurrent wiring structure within our network is high-rank due to the distance-dependent connections; cell pairs that are distant from one another do not share presynaptic inputs. The low-dimensional variability in our model emerges from a network instability at zero spatial Fourier mode that produces activity that propagates broadly across the network through polysynaptic connections.

Internal versus External Population Variability

Our circuit model assumed that the component of population-wide variability that is subject to attentional modulation was internally generated within the network. While our model is a parsimonious explanation of the data, it does not definitively exclude mechanisms in which variability is inherited from outside sources. Fluctuations from external sources are an often assumed and straightforward mechanism for population-wide spiking variability (Doiron et al., 2016; Hennequin et al., 2018; Ponce-Alvarez et al., 2013; Wimmer et al., 2015; Kanashiro et al., 2017; Bondy et al., 2018). For instance, pupil diameter is an indicator of overall brain state and arousal level, and fluctuations in pupil diameter are correlated with the fluctuations in cholinergic and noradrenergic projections to sensory cortex (Reimer et al., 2016). The reduction of population-wide variability reported in aroused states and during locomotion is likely a reflection of the quenching of these external fluctuations (McGinley et al., 2015). While it is tempting to extend this idea to variability modulation in selective attention, there are some key differences that complicate this interpretation. Whole-brain state is not changing when attention is directed into or out of the receptive field of a neuron, and thus the neuronal correlates of arousal are possibly distinct from those of attention. Further, the differential modulation of r_{SC} between and within cortical areas (Figures 1A and 1B) is difficult to explain with just a single "brain state" latent variable (Figure 2A). Rather, we expect that arousal would reduce both within and between area population-wide variability.

Nevertheless, it is popular to associate the variability within a lower area as inherited from top-down projections (Bondy et al., 2018; Wimmer et al., 2015). However, cooling experiments that inactivate top-down projections from visual areas V2 and V3 to area V1 produce only a slight reduction in single V1 neuron variability (Gómez-Laberge et al., 2016). In contrast, the bottom-up transfer of variability from lower to higher visual areas can be significant (Gómez-Laberge et al., 2016; Ruff and Cohen, 2016b). Additional multi-area population recordings between connected brain regions will be needed to probe how correlated variability flows along bottom-up and top-down pathways.

A second way to change population output variability is to keep input fluctuations fixed and shift the operating point of the network through an additional static top-down modulation. Here "operating point" designates the mean firing rates and neuronal gains about which the network will filter and transfer inputs. This shift in operating point allows the nonlinearities inherent in spiking dynamics to change the gain of how input variability transfers to output variability (Doiron et al., 2016). This mechanism has been suggested for how top-down or bottom-up modulation affects population variability in recurrent excitatory-inhibitory cortical networks (Kanashiro et al., 2017; Hennequin et al., 2018). Our model of variability modulation is similar, since the top-down attentional signal does shift the operating point of our nonlinear network; however, there are some key distinctions.

In disordered networks (Kanashiro et al., 2017) or networks with only one-dimensional structure (Hennequin et al., 2018), an external source of fluctuations is required; otherwise, the network is either in the asynchronous or pathologically synchronous solution depending upon parameter choices (Figures 3Ai, 3Aiii, and 3D). Since these networks do not produce variability internally, the operating point shift merely changes how the network filters the external fluctuations. However, in our model the two-dimensional spatial structure supports rich internal chaotic network dynamics outside the asynchronous state, yet with population-wide correlations that are a reasonable mimic of experiment (Figures 3Aiv, 3C, and 3E). There is no need to assume a source of external fluctuations.

Spatiotemporal chaos is a hallmark feature of systems that are far from equilibrium in physics, chemistry, and biology (Cross and Hohenberg, 1993). In particular, low-viscosity fluids produce a special brand of spatiotemporal chaotic behavior labeled turbulence, characterized by the presence of vortices and eddies in the fluid flow (Davidson, 2015). Like our network, the character of turbulent flow is very dependent upon the dimension of the

fluid, with one-dimensional fluids not showing turbulence, and two-dimensional turbulent flow having larger spatial scales than the flow in full three-dimensional fluids (Davidson, 2015). The dynamics within recurrent networks of neurons are certainly not equivalent to that of fluids; nevertheless, the fluid analogy to our work is tempting since the chaotic dynamics of our two-dimensional network have a macroscopic character that permits low, but non-vanishing, pairwise correlations that extend broadly over the network. The effect of top-down attention is to not only shift the operating point of the network but also dampen the macroscopic chaotic dynamics of the network. In other words, attention not only attenuates the transfer of population-wide variability, as in other models (Kanashiro et al., 2017; Hennequin et al., 2018), but also quenches the variability that is to be transferred. This permits a near-complete attention-mediated suppression of internally generated correlations (Figure 7D). This extreme sensitivity allows top-down inputs to easily control the processing state of a network.

State-dependent shifts in population-wide variability are widespread throughout cortex (Doiron et al., 2016) and are often a signature of cognitive control. The circuit structure of our network is not a special feature of the primate visual system, but rather a generic property of most cortices. We thus expect that the basic mechanisms for population-wide variability and its modulation exposed in our study will be operative in many regions of the cortex, and in many animal systems.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Spiking neuron network
  - Neural field model and stability analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Datasets
  - Noise correlation
  - Factor analysis
  - Measure internal variability
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures, two tables, four videos, and supplemental methods and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.11.034.

ACKNOWLEDGMENTS

Swartz Foundation Fellowship #2017-7 (C.H.); NIH grants CRCNS R01DC015139-01ZRG1 (B.D.), 1U19NS107613-01 (B.D.), R01EB026953 (B.D.), 1RF1MH114223-01 (B.D.), 4R00EY020844-03 (M.R.C.), R01 EY022930 (M.R.C.), 5T32NS7391-14 (D.A.R.), and Core Grant P30 EY008098; NSF grants DMS-1517828 (R.R.), DMS-1654268 (R.R.), and Neuronex DBI-1707400 (R.R.) and DMS-1517082 (B.D.); Vannevar Bush faculty fellowship N00014-18-1-2002 (B.D.); a Whitehall Foundation Grant (M.R.C.); a Klingenstein-Simons Fellowship (M.R.C.); grants from the Simons Foundation (B.D. and M.R.C.); a Sloan Research Fellowship (M.R.C.); and a McKnight Scholar Award (M.R.C.).

AUTHOR CONTRIBUTIONS

C.H., M.R.C., and B.D. conceived the project; C.H. performed the simulations and data analysis; R.P. and R.R. analyzed the firing rate model; D.A.R. and M.R.C. provided the experimental data; B.D. supervised the project; and all authors contributed to writing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 26, 2018
Revised: October 25, 2018
Accepted: November 19, 2018
Published: December 20, 2018

REFERENCES

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

Angulo, M.C., Rossier, J., and Audinat, E. (1999). Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex. J. Neurophysiol. *82*, 1295–1302.

Bondy, A.G., Haefner, R.M., and Cumming, B.G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. Nat. Neurosci. *21*, 598–606.

Börgers, C., and Kopell, N. (2005). Effects of noisy drive on rhythms in networks of excitatory and inhibitory neurons. Neural Comput. *17*, 557–608.

Bosking, W.H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J. Neurosci. *17*, 2112–2127.

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

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

Cohen, M.R., and Maunsell, J.H. (2011). Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron *70*, 1192–1204.

Compte, A., Sanchez-Vives, M.V., McCormick, D.A., and Wang, X.J. (2003). Cellular and network mechanisms of slow oscillatory activity (<1 Hz) and wave propagations in a cortical network model. J. Neurophysiol. *89*, 2707–2725.

Cross, M.C., and Hohenberg, P.C. (1993). Pattern formation outside of equilibrium. Rev. Mod. Phys. *65*, 851.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

Darshan, R., van Vreeswijk, C., and Hansel, D. (2018). Strength of correlations in strongly recurrent neuronal networks. Phys. Rev. X. *8*, 031072.

Davidson, P. (2015). Turbulence: An Introduction for Scientists and Engineers (Oxford University Press).

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

Ecker, A.S., Berens, P., Cotton, R.J., Subramaniyan, M., Denfield, G.H., Cadwell, C.R., Smirnakis, S.M., Bethge, M., and Tolias, A.S. (2014). State dependence of noise correlations in macaque primary visual cortex. Neuron *82*, 235–248.

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. Rep. Prog. Phys. *61*, 353.

Geiger, J.R., Lübke, J., Roth, A., Frotscher, M., and Jonas, P. (1997). Submillisecond AMPA receptor-mediated signaling at a principal neuron-interneuron synapse. Neuron *18*, 1009–1023.

Gómez-Laberge, C., Smolyanskaya, A., Nassi, J.J., Kreiman, G., and Born, R.T. (2016). Bottom-up and top-down input augment the variability of cortical neurons. Neuron *91*, 540–547.

Hennequin, G., Ahmadian, Y., Rubin, D.B., Lengyel, M., and Miller, K.D. (2018). The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. Neuron *98*, 846–860.e5.

Horvát, S., Gămănuţ, R., Ercsey-Ravasz, M., Magrou, L., Gămănuţ, B., Van Essen, D.C., Burkhalter, A., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2016). Spatial embedding and wiring cost constrain the functional layout of the cortical network of rodents and primates. PLoS Biol. *14*, e1002512.

Kanashiro, T., Ocker, G.K., Cohen, M.R., and Doiron, B. (2017). Attentional modulation of neuronal variability in circuit models of cortex. eLife *6*, e23978.

Kass, R.E., et al. (2018). Computational neuroscience: mathematical and statistical perspectives. Annu. Rev. Stat. Appl. *5*, 183–214.

Keane, A., and Gong, P. (2015). Propagating waves can explain irregular neural dynamics. J. Neurosci. *35*, 1591–1605.

Kelly, R.C., Smith, M.A., Kass, R.E., and Lee, T.S. (2010). Local field potentials indicate network state and account for neuronal response variability. J. Comput. Neurosci. *29*, 567–579.

Kim, H., Ährlund-Richter, S., Wang, X., Deisseroth, K., and Carlén, M. (2016). Prefrontal parvalbumin neurons in control of attention. Cell *164*, 208–218.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. (2016). Correlations and neuronal population information. Annu. Rev. Neurosci. *39*, 237–256.

Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A., Miller, K.D., and Froemke, R.C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci. *20*, 62–71.

Landau, I.D., and Sompolinsky, H. (2018). Coherent chaos in a recurrent neural network with structured connectivity. bioRxiv. https://doi.org/10.1101/350801.

Latham, P.E. (2016). Correlations demystified. Nat. Neurosci. *20*, 6–8.

Levy, R.B., and Reyes, A.D. (2012). Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. J. Neurosci. *32*, 5609–5619.

Lim, S., and Goldman, M.S. (2014). Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. J. Neurosci. *34*, 6790–6806.

Lin, I.C., Okun, M., Carandini, M., and Harris, K.D. (2015). The nature of shared cortical variability. Neuron *87*, 644–656.

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P.E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. Nature *466*, 123–127.

Mariño, J., Schummers, J., Lyon, D.C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., and Sur, M. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. Nat. Neurosci. *8*, 194–201.

Mastrogiuseppe, F., and Ostojic, S. (2018). Linking connectivity, dynamics and computations in recurrent neural networks. Neuron *99*, 609–623.e29.

McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015). Waking state: rapid variations modulate neural and behavioral responses. Neuron *87*, 1143–1161.

Middleton, J.W., Omar, C., Doiron, B., and Simons, D.J. (2012). Neural correlation is stimulus modulated by feedforward inhibitory circuitry. J. Neurosci. *32*, 506–518.

Monteforte, M., and Wolf, F. (2012). Dynamic flux tubes form reservoirs of stability in neuronal circuits. Phys. Rev. X *2*, 041007.

Ni, A.M., Ruff, D.A., Alberts, J.J., Symmonds, J., and Cohen, M.R. (2018). Learning and attention reveal a general relationship between population activity and behavior. Science *359*, 463–465.

Okun, M., Steinmetz, N., Cossell, L., Iacaruso, M.F., Ko, H., Barthó, P., Moore, T., Hofer, S.B., Mrsic-Flogel, T.D., Carandini, M., and Harris, K.D. (2015). Diverse coupling of neurons to populations in sensory cortex. Nature *521*, 511–515.

Ponce-Alvarez, A., Thiele, A., Albright, T.D., Stoner, G.R., and Deco, G. (2013). Stimulus-dependent variability and noise correlations in cortical MT neurons. Proc. Natl. Acad. Sci. USA *110*, 13162–13167.

Pyle, R., and Rosenbaum, R. (2017). Spatiotemporal dynamics and reliable computations in recurrent spiking neural networks. Phys. Rev. Lett. *118*, 018103.

Rabinowitz, N.C., Goris, R.L., Cohen, M., and Simoncelli, E.P. (2015). Attention stabilizes the shared gain of V4 populations. eLife *4*, e08998.

Reimer, J., McGinley, M.J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D.A., and Tolias, A.S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat. Commun. *7*, 13289.

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

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

Rosenbaum, R., Smith, M.A., Kohn, A., Rubin, J.E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. Nat. Neurosci. *20*, 107–114.

Ruff, D.A., and Cohen, M.R. (2016a). Attention increases spike count correlations between visual cortical areas. J. Neurosci. *36*, 7523–7534.

Ruff, D.A., and Cohen, M.R. (2016b). Stimulus dependence of correlated variability across cortical areas. J. Neurosci. *36*, 7546–7556.

Salin, P.A., and Prince, D.A. (1996). Spontaneous GABAA receptor-mediated inhibitory currents in adult rat somatosensory cortex. J. Neurophysiol. *75*, 1573–1588.

Sato, T.K., Nauhaus, I., and Carandini, M. (2012). Traveling waves in visual cortex. Neuron *75*, 218–229.

Schmitz, T.W., and Duncan, J. (2018). Normalization and the cholinergic microcircuit: a unified basis for attention. Trends Cogn. Sci. *22*, 422–437.

Schölvinck, M.L., Saleem, A.B., Benucci, A., Harris, K.D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. J. Neurosci. *35*, 170–178.

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

Smith, M.A., and Sommer, M.A. (2013). Spatial and temporal scales of neuronal correlation in visual area V4. J. Neurosci. *33*, 5422–5432.

Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., and Lesica, N.A. (2016). Inhibitory control of correlated intrinsic variability in cortical networks. eLife *5*, e19695.

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

Williamson, R.C., Cowley, B.R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M.A., and Yu, B.M. (2016). Scaling properties of dimensionality reduction for neural populations and network models. PLoS Comput. Biol. *12*, e1005141.

Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., and de la Rocha, J. (2015). Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. Nat. Commun. *6*, 6177.

Xiang, Z., Huguenard, J.R., and Prince, D.A. (1998). GABAA receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex. J. Physiol. *506*, 715–730.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Algorithms for simulation of spiking neuron networks	This paper	https://github.com/hcc11/SpatialNeuronNet
MATLAB	MathWorks	https://www.mathworks.com

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Brent Doiron (bdoiron@pitt.edu).

METHOD DETAILS

Spiking neuron network

The network consists of three layers. Layer 1 is modeled by a population of $N_1 = 2,500$ excitatory neurons, the spikes of which are taken as independent Poisson processes with a uniform rate $r_1 = 10$ Hz. Layer 2 and Layer 3 are recurrently coupled networks with excitatory $(\alpha = \mathrm{e})$ and inhibitory $(\alpha = \mathrm{i})$ populations of $N_\mathrm{e} = 40,000$ and $N_\mathrm{i} = 10,000$ neurons, respectively. Each neuron is modeled as an exponential integrate-and-fire (EIF) neuron whose membrane potential is described by:

$$C_m \frac{dV_j^\alpha}{dt} = -g_L\left(V_j^\alpha - E_L\right) + g_L \Delta_T e^{\left(V_j^\alpha - V_T\right)/\Delta_T} + I_j^\alpha(t). \quad \text{(Equation 3)}$$

Each time $V_j^\alpha(t)$ exceeds a threshold V_th, the neuron spikes and the membrane potential is held for a refractory period τ_ref then reset to a fixed value V_re. Neuron parameters for excitatory neurons are $\tau_m = C_m/g_L = 15$ ms, $E_L = -60$ mV, $V_T = -50$ mV, $V_\mathrm{th} = -10$ mV, $\Delta_T = 2$ mV, $V_\mathrm{re} = -65$ mV and $\tau_\mathrm{ref} = 1.5$ ms. Inhibitory neurons are the same except $\tau_m = 10$ ms, $\Delta_T = 0.5$ mV and $\tau_\mathrm{ref} = 0.5$ ms. The total current to the j$^\mathrm{th}$ neuron is:

$$\frac{I_j^\alpha(t)}{C_m} = \sum_{k=1}^{N_F} \frac{J_{jk}^{\alpha F}}{\sqrt{N}} \sum_n \eta_F\left(t - t_n^{F,k}\right) + \sum_{\beta = \mathrm{e,i}} \sum_{k=1}^{N_\beta} \frac{J_{jk}^{\alpha\beta}}{\sqrt{N}} \sum_n \eta_\beta\left(t - t_n^{\beta,k}\right) + \mu_\alpha, \quad \text{(Equation 4)}$$

where $N = N_e + N_i$ is the total number of the network population. The postsynaptic current is given by

$$\eta_\beta(t) = \frac{1}{\tau_{\beta d} - \tau_{\beta r}} \begin{cases} e^{-t/\tau_{\beta \mathrm{d}}} - e^{-t/\tau_{\beta r}}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad \text{(Equation 5)}$$

where $\tau_\mathrm{er} = 1$ ms, $\tau_\mathrm{ed} = 5$ ms and $\tau_\mathrm{ir} = 1$ ms, $\tau_\mathrm{id} = 8$ ms. The feedforward synapses from Layer 1 to Layer 2 have the same kinetics as the recurrent excitatory synapse, i.e., $\eta_F^{(2)}(t) = \eta_\mathrm{e}(t)$. The feedforward synapses from Layer 2 to Layer 3 have a fast and a slow component.

$$\eta_F^{(3)}(t) = p_\mathrm{f}\eta_\mathrm{e}(t) + p_\mathrm{s}\eta_\mathrm{s}(t) \quad \text{(Equation 6)}$$

with $p_\mathrm{f} = 0.2$, $p_\mathrm{s} = 0.8$. $\eta_\mathrm{s}(t)$ has the same form as Equation 5 with a rise time constant $\tau_r^s = 2$ ms and a decay time constant $\tau_d^s = 100$ ms. The excitatory and inhibitory neurons in Layer 3 receive static current μ_e and μ_i, respectively.

Neurons on the three layers are arranged on a uniform grid covering a unit square $\Gamma = [0,1] \times [0,1]$. The probability that two neurons, with coordinates $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ respectively, are connected depends on their distance from one another measured periodically on Γ:

$$p_{\alpha\beta}(\mathbf{x}, \mathbf{y}) = \overline{p}_{\alpha\beta} g(x_1 - y_1; \alpha_\beta) g(x_2 - y_2; \alpha_\beta). \quad \text{(Equation 7)}$$

Here $\overline{p}_{\alpha\beta}$ is the mean connection probability and

$$g(x; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} e^{-(x+k)^2/\left(2\sigma^2\right)} \quad \text{(Equation 8)}$$

is a wrapped Gaussian distribution. Excitatory and inhibitory recurrent connection widths of Layer 2 are $\alpha_\mathrm{rec}^{(2)} := \alpha_\mathrm{e}^{(2)} = \alpha_\mathrm{i}^{(2)} = 0.1$ and feedforward connection width from Layer 1 to Layer 2 is $\alpha_\mathrm{ffwd}^{(2)} = 0.05$. The recurrent connection width of Layer 3 is $\alpha_\mathrm{rec}^{(3)} = 0.2$ and the

feedforward connection width from Layer 2 to Layer 3 is $\alpha_{\text{ffwd}}^{(3)} = 0.1$. A presynaptic neuron is allowed to make more than one synaptic connection to a single postsynaptic neuron.

The recurrent connectivity of Layer 2 and Layer 3 have the same synaptic strengths and mean connection probabilities. The recurrent synaptic weights are $J_{ee} = 80$ mV, $J_{ei} = -240$ mV, $J_{ie} = 40$ mV and $J_{ii} = -300$ mV. Recall that individual synapses are scaled with $1/\sqrt{N}$ (Equation 4); so that, for instance, $J_{ee}/\sqrt{N} \approx 0.36$ mV. The mean connection probabilities are $\overline{p}_{ee} = 0.01$, $\overline{p}_{ei} = 0.04$, $\overline{p}_{ie} = 0.03$, $\overline{p}_{ii} = 0.04$. The out-degrees are $K_{ee}^{\text{out}} = 400$, $K_{ei}^{\text{out}} = 1600$, $K_{ie}^{\text{out}} = 300$ and $K_{ii}^{\text{out}} = 400$. The feedforward connection strengths from Layer 1 to Layer 2 are $J_{eF}^{(2)} = 140$ mV and $J_{iF}^{(2)} = 100$ mV with probabilities $\overline{p}_{eF}^{(2)} = 0.1$ and $\overline{p}_{iF}^{(2)} = 0.05$ (out-degrees $K_{eF2}^{\text{out}} = 4000$ and $K_{eF2}^{\text{out}} = 500$). The feedforward connection strengths from Layer 2 to Layer 3 are $J_{eF}^{(3)} = 25$ mV and $J_{iF}^{(3)} = 15$ mV with mean probabilities $\overline{p}_{eF}^{(3)} = 0.05$ and $\overline{p}_{iF}^{(3)} = 0.05$ (out-degrees are $K_{eF3}^{\text{out}} = 2000$ and $K_{iF3}^{\text{out}} = 500$). Only the excitatory neurons in Layer 2 project to Layer 3.

The spatial models in Figures 3Aii and 3Aiv contain only Layer 1 and Layer 2. In the model with disordered connectivity, the connection probability between a pair of neurons is $\overline{p}_{\alpha\beta}$, independent of distance. Other parameters are the same as the spatial model. The decay time constant of IPSC (τ_{id}) was varied from 1 to 15 ms (Figure 3E). The rise time constant of IPSC (τ_{ir}) is 1 ms when $\tau_{\text{id}} > 1$ ms and 0.5 ms when $\tau_{\text{id}} = 1$ ms.

The parameters used in Figures 4C and 4D are $\mu_I = [0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4]$ mV/ms and $\mu_E = 0$ mV/ms. The mean firing rates in Layer 2 are $r_{\text{e}}^{(2)} = 19$ Hz and $r_{\text{i}}^{(2)} = 9$ Hz. In the further analysis (Figures 4E, 5C, 5D, 6B, and 6C), we used $\mu_I = 0.2$ mV/ms for the unattended state and $\mu_I = 0.35$ mV/ms for the attended state. In simulations of the spatial model with broad inhibitory projection (Figure 6C), $\alpha_{\text{e}}^{(3)} = 0.1$, $\alpha_{\text{i}}^{(3)} = 0.2$. Other parameters were not changed.

The slow component of feedforward excitation (Equation 6) allows for large spike count Fano factors (Figure S5), but when replaced with the same fast kinetics of recurrent excitation the low dimensional population-wide variability and its attentional modulation are qualitatively unaffected (Figure S8).

All simulations were performed on the CNBC Cluster in the University of Pittsburgh. All simulations were written in a combination of C and MATLAB (MATLAB R 2015a, MathWorks). The differential equations of the neuron model were solved using forward Euler method with time step 0.01 ms.

Neural field model and stability analysis

We use a two dimensional neural field model to describe the dynamics of population rate (Figure 6). The neural field equations are

$$\tau_\alpha \frac{\partial r_\alpha(x,t)}{\partial t} = -r_\alpha + \phi_\alpha(w_{\alpha e} * r_e + w_{\alpha i} * r_i + \mu_\alpha), \qquad \text{(Equation 9)}$$

where $r_\alpha(x,t)$ is the firing rate of neurons in population $\alpha = \text{e, i}$ near spatial coordinates $x \in [0,1] \times [0,1]$. The symbol $*$ denotes convolution in space, μ_α is a constant external input and $w_{\alpha\beta}(x) = \overline{w}_{\alpha\beta} g(x;\sigma_\beta)$ where $g(x;\sigma_\beta)$ is a two-dimensional wrapped Gaussian with width parameter σ_β, $\beta = \text{e,i}$. The transfer function is a threshold-quadratic function, $\phi_\alpha(x) = k_\alpha [x^2]_+$. The timescale of synaptic and firing rate responses are implicitly combined into τ_α. In networks with approximate excitatory-inhibitory balance, rates closely track synaptic currents (Renart et al., 2010), so τ_α represents the synaptic time constant of population $\alpha = \text{e,i}$.

For constant inputs, μ_e and μ_i, there exists a spatially uniform fixed point, which was computed numerically using an iterative scheme (Rosenbaum and Doiron, 2014). Linearizing around this fixed point in Fourier domain gives a Jacobian matrix at each spatial Fourier mode (Rosenbaum and Doiron, 2014)

$$J(\vec{n}) = \begin{bmatrix} (-1 + g_e \tilde{w}_{ee}(\vec{n}))/\tau_e & g_e \tilde{w}_{ei}(\vec{n})/\tau_e \\ g_i \tilde{w}_{ie}(\vec{n})/\tau_i & (-1 + g_i \tilde{w}_{ii}(\vec{n}))/\tau_i \end{bmatrix}.$$

where $\vec{n} = (n_1, n_2)$ is the two-dimensional Fourier mode, $\tilde{w}_{\alpha\beta}(\vec{n}) = \overline{w}_{\alpha\beta} \exp(-2|\vec{n}|^2 \pi^2 \sigma_\beta^2)$ is the Fourier coefficient of $w_{\alpha\beta}(x)$ with $\|\vec{n}\|^2 = n_1^2 + n_2^2$ and g_a is the gain, which is equal to $\phi'_\alpha(r_\alpha)$ evaluated at the fixed point. The fixed point is stable at Fourier mode $\vec{n}$ if both eigenvalues of $J(\vec{n})$ have negative real part. Note that stability only depends on the wave number, $k = \|\vec{n}\|$, so Turing-Hopf instabilities always occur simultaneously at all Fourier modes with the same wave number (spatial frequency).

For the stability analysis in Figure 6A, τ_i varies from 2.5 ms to 25 ms, σ_i varies from 0.05 to 0.2, and $\tau_e = 5$ ms and $\sigma_e = 0.1$. The rest of the parameters were $\overline{w}_{ee} = 80$, $\overline{w}_{ei} = -160$, $\overline{w}_{ie} = 120$, $\overline{w}_{ii} = -200$, $k_e = 1$, $k_i = 1$, $\mu_e = 0.48$ and $\mu_i = 0.32$. Depolarizing the inhibitory population ($\mu_i = 0.5$) expands the stable region (Figure 6A, black dashed).

QUANTIFICATION AND STATISTICAL ANALYSIS

Datasets

Each of the two datasets (recordings from V4 and recordings from V1 and MT) was collected from two different rhesus monkeys as they performed an orientation-change detection task. All animal procedures were in accordance with the Institutional Animal Care and Use Committee of Harvard Medical School, University of Pittsburgh and Carnegie Mellon University.

For analysis in Figures 1A, 1C, 5A, and 5B, data was collected with two microelectrode arrays implanted bilaterally in area V4 (Cohen and Maunsell, 2009). In our analysis, we include stimulus presentations prior to the change stimulus from correct trials, excluding

the first stimulus in a trial to avoid adaptation effects. Spike counts during the sustained response (120 - 260 ms after stimulus onset) are considered for the correlation and factor analysis. Neurons recorded from either the left or right hemisphere in one session are treated separately. There are a total of 42,496 trials for 72,765 pairs from 74 recording sessions. Two sessions from the original study were excluded for factor analysis due to inadequate trials. The trial number and unit number of each session is summarized in Table S1.

For analysis in Figure 1B, data was collected with one microelectrode array implanted in area V1 and a single electrode or a 24-channel linear probe inserted into MT (Ruff and Cohen, 2016a). Again, our analysis includes full contrast stimulus presentations prior to the change stimulus from correct trials and excludes the first stimulus in a trial to avoid adaptation effects. Spike counts are measured 30 - 230 ms after stimulus onset for V1 and 50 - 250 ms after stimulus onset for MT to account for the average visual latencies of neurons in both areas. There are a total of 1,631 V1-MT pairs from 32 recording sessions.

Noise correlation

To compute the noise correlation of each simulation, 500 neurons were randomly sampled without replacement from the excitatory population of Layer 3 and Layer 2 within a [0, 0.5]x[0, 0.5] square (considering periodic boundary condition). Spike counts were computed using a sliding window of 200 ms with 1 ms step size and the Pearson correlation coefficients were computed between all pairs. Neurons of firing rates less than 2 Hz were excluded from the computation of correlations. In Figures 4C and 4D, for each μ_i there were 50 simulations and each simulation was 20 s long. Connectivity matrices and the initial states of each neuron's membrane potential were randomized in each simulation. The first 1 s of each simulation was excluded from the correlation analysis. Standard error was computed based on the mean correlations of each simulation. For simulations of Figure 3E, there was one simulation of 20 s per τ_{id} and the connectivity matrices were randomized for each simulation. To compute the noise correlation, 1000 neurons were randomly sampled without replacement in the excitatory population of Layer 2 within a [0, 0.5]x[0, 0.5] square. Correlations are computed between firing rates that are smoothed with a Gaussian window of width 10 ms.

Factor analysis

Factor analysis assumes spike counts of n simultaneously recorded neurons $x \in \mathcal{R}^{n\times 1}$ is a multi-variable Gaussian process

$$x \sim \mathcal{N}\left(\mu, LL^T + \Psi\right)$$

where $\mu \in \mathcal{R}^{n\times 1}$ is the mean spike counts, $L \in \mathcal{R}^{n\times m}$ is the loading matrix of the m latent variables and $\Psi \in \mathcal{R}^{n\times 1}$ is a diagonal matrix of independent variances for each neuron. We choose $m = 5$ and compute the eigenvalues of LL^T, λ_i $(i = 1, 2, \ldots, m)$, ranked in descending order. The corresponding eigenvectors are denoted as v_i and the covariance components are $\lambda_i v_i v_i^T$ $(i = 1, 2, \ldots, m)$. The covariance matrix is approximated as

$$\mathrm{Cov}(x, x) = \sum_{i=1}^{m} \lambda_i v_i v_i^T + \Psi. \qquad \text{(Equation 10)}$$

We compute the residual covariance after subtracting the first mode as

$$Q = \mathrm{Cov}(x, x) - L_1 \times L_1^T, \qquad \text{(Equation 11)}$$

where $\mathrm{Cov}(x, x)$ is the raw covariance matrix of x and L_1 is the loading matrix when fitting with $m = 1$. The mean raw covariance and residual (Figures 4E and 1C, right) are the mean of the off-diagonal elements of $\mathrm{Cov}(x, x)$ and Q, respectively.

When applying factor analysis on model simulations (Figure 4E), we randomly selected 50 excitatory neurons from Layer 3, whose firing rates were larger than 2 Hz in both the unattended and attended states. There were 15 simulations of 20 s each per connectivity matrices realization, and there were 8 realizations of connectivity matrices in total. Spike trains were truncated into 140 ms spike count window with a total of 2,025 counts per neuron. There were 80 non-overlapping sampling of neurons (10 sampling per realization of connectivity matrices) and we applied factor analysis on each sampling of neuron spike counts.

To compute the distance dependent functions of each covariance component $\lambda_i v_i v_i^T$ (Figures 5C and 5D), we randomly selected 500 excitatory neurons from Layer 3, whose firing rates were larger than 2 Hz in both the unattended and attended states. To compute the distance dependent functions of covariance as well as convariance components for the V4 data (Figures 5A and 5B), the pairs across sessions were pooled to compute the mean and SEM at each distance value. The distance values were discrete since neurons were recorded with an multi-electrode array with a distance between adjacent electrodes being 400 μm. The number of pairs at each distance value are shown in Table S2.

Measure internal variability

To study the chaotic population firing rate dynamics of Layer 3 (Figure 7), we fixed the spike trains realizations from Layer 1 neurons, the membrane potential states of the Layer 2 neurons and all connectivity matrices. Only the initial membrane potentials of Layer 3 neurons were randomized across trials. There were 10 realizations of Layer 1 and Layer 2, each of which was 20 s long. For each simulation of Layer 2, 20 repetitions with different initial conditions were simulated for Layer 3. The connectivity matrices in Layer 3 were the same across the 20 repetitions but different for each realization of Layer 1 and Layer 2. The realizations of Layer 1 and Layer 2

and the connectivity matrices were the same for the attended and unattended states. Trial-to-trial variance of Layer 3 population rates (Figure 7D) was the variance of the mean population rates of the Layer 3 excitatory population, smoothed by a 200 ms rectangular filter, across the 20 repetitions. The first second of each simulation was discarded.

DATA AND SOFTWARE AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon request. Computer code for all simulations and data analysis can be found at https://github.com/hcc11/SpatialNeuronNet.

Neuron, Volume *101*

Supplemental Information

Circuit Models of Low-Dimensional Shared Variability in Cortical Networks

Chengcheng Huang, Douglas A. Ruff, Ryan Pyle, Robert Rosenbaum, Marlene R. Cohen, and Brent Doiron

Supplemental Information

Circuit models of low dimensional shared variability in cortical networks

Chengcheng Huang, Douglas A. Ruff, Ryan Pyle, Robert Rosenbaum,
Marlene R. Cohen and Brent Doiron

This PDF file includes:

Supplemental Figures S1-S8
Table S1-S2
Method S1

Other Supplementary Materials for this manuscript includes the following:

Movies S1-S4

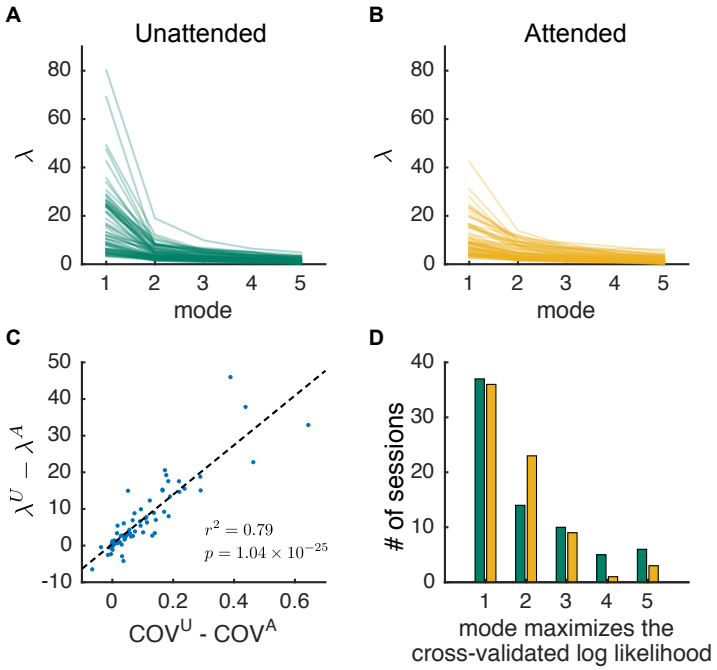

Figure S1: **Related to Figure 1. Factor analysis of the multi-electrode recordings from V4 (Cohen and Maunsell, 2009). A**, The first five largest eigenvalues of the shared component of the spike count covariance matrix. Each line is for data from each recording session (72 in total, trial number and unit number of each session see Table S1). **B**, Same as panel A for the attended state. **C**, The difference between the largest eigenvalues and the difference between the mean covariances from the unattended and attended states are correlated (U: unattended; A: attended) . **D**, Histogram of the modes that maximize the cross-validated log likelihood across sessions. More details see Experimental methods and Statistical methods.

Method S1: Hidden variable model (related to Figure 2 and section 'Constraints for circuit-based models of shared variability')

Consider a population of N neurons. Let r_i be the rates of neuron i ($i = 1, 2, \ldots, N$) and $R = \langle r_i \rangle_i$ be the mean population rate where $\langle \cdot \rangle_i$ means average over i. If N is large, then the variance of R is approximately the average of pairwise covariance, i.e.

$$\mathrm{Var}(R) \approx \langle \mathrm{Cov}(r_i, r_j) \rangle_{i,j}.$$

Since population activity has been shown to be low dimensional (Fig. 1C), we use a scalar response variable, R, to describe population activity and consider $\mathrm{Var}(R)$ as the mean pairwise covariance of the population.

Consider two hierarchically connected cortical areas, such as V1 and MT. Let R be the mean population rate of the neurons in the downstream area, MT, and X be the mean population rate of the upstream area, V1 (Fig. 2A). Suppose all the variability in R is from X and the transfer function of X to R can be linearly approximated as $\delta R = \gamma \delta X$, then

$$\mathrm{Var}(R) = \gamma^2 \mathrm{Var}(X), \tag{1}$$
$$\mathrm{cov}(R, X) = \gamma \mathrm{Var}(X). \tag{2}$$

which gives

$$\mathrm{Var}(R) = \mathrm{cov}(R, X)^2 / \mathrm{Var}(X).$$

Hence any decrease in $\mathrm{Var}(R)$ by attention predicts a decrease in $\mathrm{cov}(R, X)$, which is in contradiction with the electrophysiological recordings from visual areas V1 and MT (Fig. 1B; Ruff and Cohen, 2016a).

Suppose an hidden source of variability, H, projects to R and X with strength β and κ, respectively (Fig. 2A). Specifically, we have

$$\begin{aligned} R &= \gamma X + \beta H, \\ X &= X_0 + \kappa H, \end{aligned}$$

where X_0 is independent from H. Assume $\mathrm{Var}(X) = 1$ and β and κ are attention independent, then

$$\mathrm{Var}(R) = \gamma^2 + \left(\beta^2 + 2\beta\kappa\gamma\right) \mathrm{Var}(H), \tag{3}$$
$$\mathrm{cov}(R, X) = \gamma + \beta\kappa \mathrm{Var}(H). \tag{4}$$

Reduction in covariance within one area

We first consider the attentional effect on noise correlations in one cortical area. Here we take the variance of H, $\mathrm{Var}^{\alpha}(H)$, to be dependent on the attentional state, where $\alpha = U$ means value at the unattended state and $\alpha = A$ means value at the attended state. Assume that other parameters are independent on the attentional state. We denote

$$P_H = \frac{\beta^2 \mathrm{Var}^U(H)}{\mathrm{Var}^U(R)} = \frac{\beta^2 \mathrm{Var}^U(H)}{\mathrm{Var}(X) + \beta^2 \mathrm{Var}^U(H)} \tag{5}$$

as the relative influence of H on R ($0 < P_H < 1$). Then the relationship between attention mediated changes in H and R is

$$\frac{\underset{U-A}{\Delta}\text{Var}(H)}{\text{Var}^U(H)} = \frac{1}{P_H}\frac{\underset{U-A}{\Delta}\text{Var}(R)}{\text{Var}^U(R)}. \tag{6}$$

Here $\underset{U-A}{\Delta}\text{Var}(H) = \text{Var}^U(H) - \text{Var}^A(H)$ (same for $\underset{U-A}{\Delta}\text{Var}(R)$). The population recordings from V4 provide that the relative change in mean pairwise covariance is about 30% (Fig. 1A; Cohen and Maunsell, 2009). Hence we choose $\underset{U-A}{\Delta}\text{Var}(R)/\text{Var}^U(R) = 0.3$. Then the relative change in $\text{Var}(H)$, $\underset{U-A}{\Delta}\text{Var}(H)/\text{Var}^U(H)$, is inversely proportional to the influence of H on R, P_H (Eq. (6), Fig. S2A).

Certain parameter choices of our simplified model are unreasonable (denoted by the pink region in Fig. S2A). For instance, if β is overly large then the influence of X on the variability of R is diluted. This would imply that higher order visual areas do not inherit fluctuations from lower order areas, which is at odds with several experimental reports (Ruff and Cohen, 2016b; Gómez-Laberge et al., 2016). Alternatively, if $\text{Var}(H) \to 0$ in the attended state this would then require that the area that produces H to be silent with attention. There is no evidence that attention has that degree of influence on any brain area. Fortunately, there are moderate β and $\text{Var}(H)$ choices that capture the data (section of the blue curve that is not in the pink region in Fig. S2A), in effect mitigating the above issues. Thus, a latent variable model where the variability is external to the population can account for the attentional modulation reported in our V4 data, as has been previously remarked (Rabinowitz et al., 2015; Kanashiro et al., 2017).

Increase in covariance between two areas

We next consider the constraint for an attentional increase in the covariance between two connected areas (Fig. 1B, Ruff and Cohen, 2016a), i.e.

$$\text{cov}^A(R, X) > \text{cov}^U(R, X).$$

From Eq. (4), we can see that a reduction in $\text{Var}(H)$ by attention would decrease the $\text{cov}(R, X)$ if other parameters remain constant, since H is a common source of fluctuations between R and X. In order to have an attention-mediated increase in $\text{cov}(R, X)$, we need that γ increases. However, an increase in γ results in an increase in $\text{Var}(R)$ (Eq. 3) due to the strengthened transfer of input fluctuations. Therefore, an attention-mediated simultaneous reduction in $\text{Var}(R)$ and increase in $\text{cov}(R, X)$ tightens model constraints due to the trade-off between an reduction in the hidden source of fluctuations and an increase in variability transfer from X to R.

In the following, we derive the constraints to have $\text{Var}(R)$ decreases while $\text{cov}(R, X)$ increases with attention. The relative change in $\text{cov}(R, X)$ by attention is

$$\frac{\underset{A-U}{\Delta}\text{cov}(R, X)}{\text{cov}^U(E, X)} = \frac{\underset{A-U}{\Delta}\gamma - \beta\kappa\underset{U-A}{\Delta}\text{Var}(H)}{\gamma_U + \beta\kappa\text{Var}^U(H)}. \tag{7}$$

An attention-mediated increase of the covariability between X and R implies $\underset{A-U}{\Delta}\text{cov}(R,X) > 0$, which from Eq. (7) requires:

$$\underset{A-U}{\Delta}\gamma > \beta\kappa \underset{U-A}{\Delta}\text{Var}(H). \tag{8}$$

The reduction in $\text{Var}(R)$ by attention is

$$\begin{aligned}
\underset{U-A}{\Delta}\text{Var}(R) &= -\left(\gamma_A^2 - \gamma_U^2\right) + \beta^2 \underset{U-A}{\Delta}\text{Var}(H) + 2\beta\kappa\left(\gamma_U \text{Var}^U(H) - \gamma_A \text{Var}^A(H)\right) \\
&= -\left(2\gamma_U + \underset{A-U}{\Delta}\gamma\right)\underset{A-U}{\Delta}\gamma + \beta^2 \underset{U-A}{\Delta}\text{Var}(H) + 2\beta\kappa\gamma_U \underset{U-A}{\Delta}\text{Var}(H) - 2\beta\kappa\text{Var}^A(H)\underset{A-U}{\Delta}\gamma \\
&= \beta^2 \underset{U-A}{\Delta}\text{Var}(H) - 2\gamma_U \underset{A-U}{\Delta}\text{cov}(R,X) - \left(\underset{A-U}{\Delta}\gamma\right)^2 - 2\beta\kappa\text{Var}^A(H)\underset{A-U}{\Delta}\gamma
\end{aligned}$$

Hence the relative reduction in $\text{Var}(R)$ is

$$\begin{aligned}
\frac{\underset{U-A}{\Delta}\text{Var}(R)}{\text{Var}^U(R)} = \frac{\underset{U-A}{\Delta}\text{Var}(H)}{\text{Var}^U(H)}P_H - 2\gamma_U \frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)}\frac{\text{cov}^U(R,X)}{\text{Var}^U(R)} \\
- \frac{\left(\underset{A-U}{\Delta}\gamma\right)^2 + 2\beta\kappa\text{Var}^A(H)\underset{A-U}{\Delta}\gamma}{\text{Var}^U(R)}.
\end{aligned} \tag{9}$$

The second term from the RHS of Eq. (9) is

$$\begin{aligned}
2\gamma_U \frac{\text{cov}^U(R,X)}{\text{Var}^U(R)}\frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)} &= \frac{2\gamma_U^2 + 2\beta\kappa\gamma_U\text{Var}^U(H)}{\gamma_U^2 + (\beta^2 + 2\beta\kappa\gamma_U)\,\text{Var}^U(H)}\frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)} \\
&> (1 - P_H)\frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)}
\end{aligned}$$

With inequality (8), the third term from the RHS of Eq. (9) is

$$\begin{aligned}
\frac{\left(\underset{A-U}{\Delta}\gamma\right)^2 + 2\beta\kappa\text{Var}^A(H)\underset{A-U}{\Delta}\gamma}{\text{Var}^U(R)} &= \frac{\underset{A-U}{\Delta}\gamma\left(\underset{A-U}{\Delta}\gamma + 2\beta\kappa\text{Var}^A(H)\right)}{\beta^2\text{Var}^U(H)}P_H \\
&> \frac{\left(\beta\kappa\underset{U-A}{\Delta}\text{Var}(H)\right)\left(\beta\kappa\underset{U-A}{\Delta}\text{Var}(H) + 2\beta\kappa\text{Var}^A(H)\right)}{\beta^2\text{Var}^U(H)}P_H \\
&= \frac{\beta^2\kappa^2\underset{U-A}{\Delta}\text{Var}(H)\left(\text{Var}^U(H) + \text{Var}^A(H)\right)}{\beta^2\text{Var}^U(H)}P_H \\
&> \kappa^2\text{Var}^U(H)\frac{\underset{U-A}{\Delta}\text{Var}(H)}{\text{Var}^U(H)}P_H
\end{aligned}$$

Hence,

$$\frac{\underset{U-A}{\Delta}\text{Var}(R)}{\text{Var}^U(R)} < \left(1-\kappa^2\text{Var}^U(H)\right)\frac{\underset{U-A}{\Delta}\text{Var}(H)}{\text{Var}^U(H)}P_H - \frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)}\left(1-P_H\right)$$

which gives the constraint on H as

$$\frac{\underset{U-A}{\Delta}\text{Var}(H)}{\text{Var}^U(H)}P_H > \frac{\underset{U-A}{\Delta}\text{Var}(R)}{\text{Var}^U(R)}\frac{1}{1-\kappa^2\text{Var}^U(H)} + \frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)}\frac{1-P_H}{1-\kappa^2\text{Var}^U(H)}. \quad (10)$$

where $0 < 1-\kappa^2\text{Var}^U(H) = \frac{\text{Var}^U(X_0)}{\text{Var}(X)} < 1$, since $\text{Var}(X) = \text{Var}(X_0) + \kappa^2\text{Var}(H) = 1$.

Therefore, the lower bound on $\frac{\underset{U-A}{\Delta}\text{Var}(H)}{\text{Var}^U(H)}P_H$ increases with the relative increase in $\text{cov}(R,X)$, $\frac{\underset{A-U}{\Delta}\text{cov}(R,X)}{\text{cov}^U(R,X)}$, and the projection strength (κ) from H to X (Fig. S2B). These constraints become more restrictive than simply the reduction of within area variability, so that accounting for the data in this latent variable model becomes difficult (the blue constraint curves are pushed into the pink region of S2B). However, setting $\kappa = 0$ provides the least restrictive constraints for β and $\underset{U-A}{\Delta}\text{Var}(H)$.

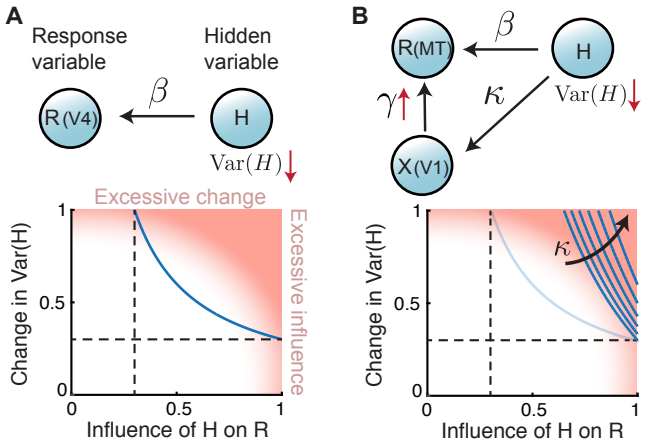

Figure S2: **Related to Figure 2. Hidden variable models of shared variability.** **A**, Top: hidden variable model where the response variability R (modeling V4) comes from a hidden variable H that projects to R with strength β. Bottom: the attention-mediated reduction in r_{SC} gives a constraint that is a trade-off between the reduction in $\mathrm{Var}(H)$ ($\frac{\Delta_{U-A}\mathrm{Var}(H)}{\mathrm{Var}^U(H)}$) and the influence of H on R (P_H, see Eq. 5). The blue curve matches the 30% reduction of r_{SC} reported in the V4 data (Fig. 1A; Eq. 6 with $\underset{U-A}{\Delta}\mathrm{Var}(R)/\mathrm{Var}^U(R) = 0.3$). Pink region: parameter choices that are considered unreasonable, such as β being overly large so that R is no longer driven by X, or $\mathrm{Var}(H) \to 0$ in the attended state, requiring the area that produces H to be silent. **B**, Top: hidden variable model for connected areas X (modeling V1) and R (modeling MT); H projects to X with strength κ and the transfer strength from X to R is γ. Bottom: the attention mediated changes in r_{SC} give further constraints on H, the blue curves are the lower bounds of permissible parameters (Eq. 10 with $\frac{\Delta_{U-A}\mathrm{Var}(R)}{\mathrm{cov}^U(R,X)} = 0.3$, $\frac{\Delta_{A-U}\mathrm{cov}(R,X)}{\mathrm{cov}^U(R,X)} = 1$ and $\kappa^2\mathrm{Var}(H)$ ranges from 0 to 0.5 for the different curves). Light blue curve is the same as that in panel A for comparison.

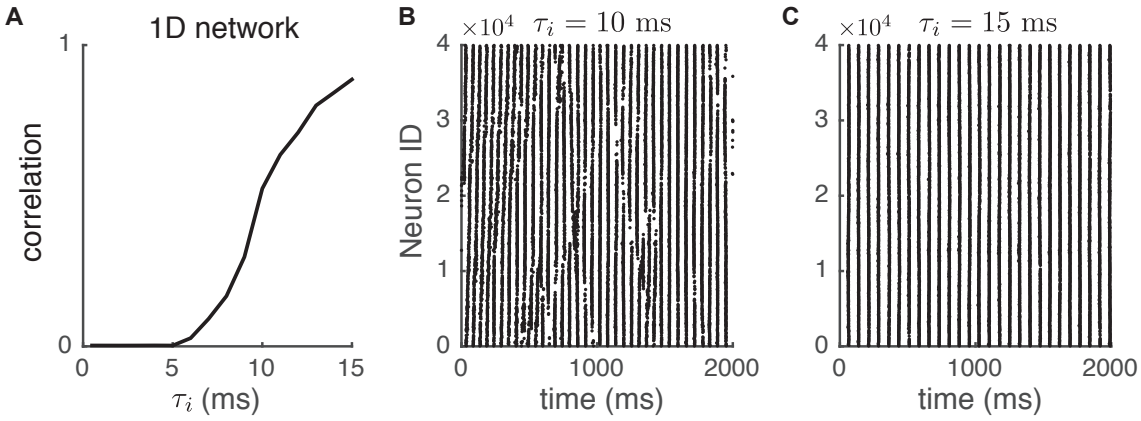

Figure S3: **Related to Figure 3D. The dynamics of one-dimensional spatial models with different time scales of inhibition.** **A**, One-dimensional spatial model shows a rapid increase in mean pairwise correlation with increasing time scale of inhibitory synaptic current. **B**, Example rasters of network activity when $\tau_i = 10$ ms. **C**, Same as panel B with $\tau_i = 15$ ms. Parameters of the one dimensional model are the same as those in the two-dimensional spatial model in Fig. 3Aii, Aiv, except that neurons are ordered on interval [0, 1].

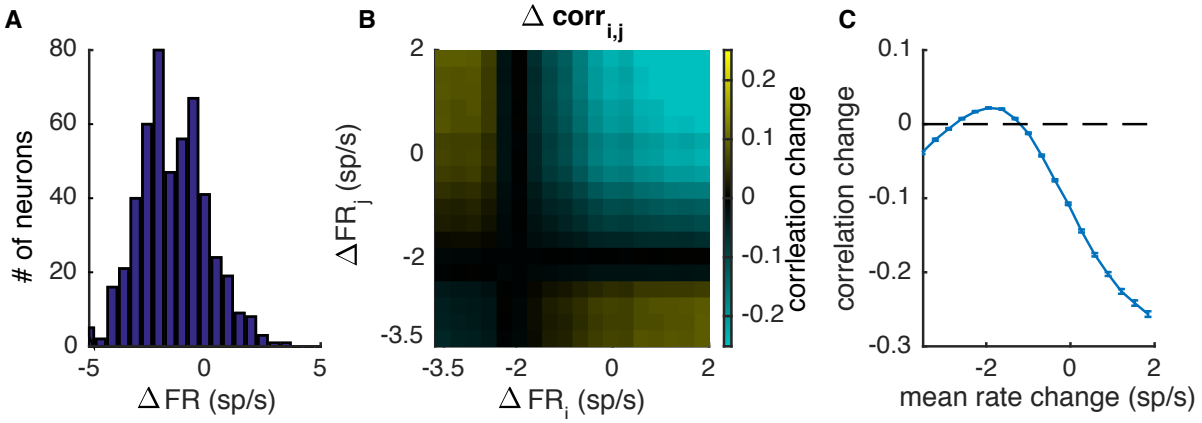

Figure S4: **Related to Figure 4. Relationship between correlation change and firing rate change of the pair by attention. A**, Histogram of firing rate change by attention. **B**, Correlation change of neuron pair i and j as a function for firing rate change of neuron i (x-axis) and firing rate change of neuron j (y-axis). **C**, Correlation change as a function of the mean firing rate change of the pair. Parameters of the network are the same as Fig. 4 with $\mu_I = 0.2$ mV/ms for unattended state and $\mu_I = 0.35$ mV/ms for attended state. A total of 500 excitatory neurons were sampled from MT layer and there was a total of 2025 spike counts per neuron to compute the correlation for each attentional state. Δ=Attended-Unattended.

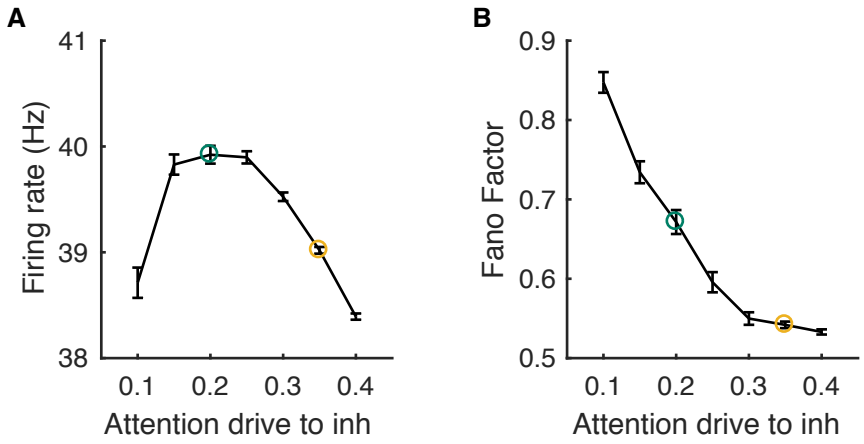

Figure S5: **Related to Figure 4. A**, Mean firing rates of the excitatory population change with attentional modulation in the network. Top-down attentional modulation is modeled as a depolarization to MT inhibitory neurons (μ_I). **B**, Same as panel A for mean Fano factors of the excitatory population. Error bars are SEM.

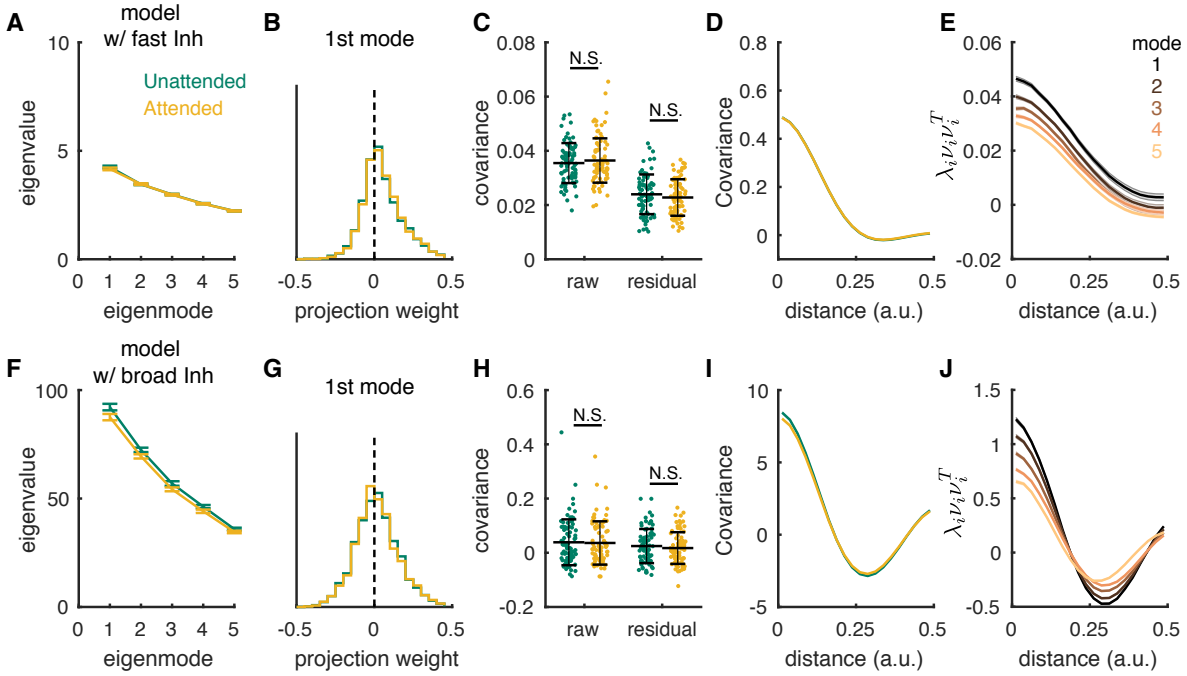

Figure S6: **Related to Figure 4E and Figure 6. Networks with either fast inhibitory synaptic current (A-E) or broad inhibitory projections (F-J) do not produce low-dimensional shared variability. A, F**, The first five largest eigenvalues of the shared component of the spike count covariance matrix. Green: unattended ($\mu_I = 0.2$); orange: attended ($\mu_I = 0.35$); Error bars are SEM. **B, G**, The vector elements for the first (dominant) eigenmode. **C, H**, The mean covariance from each session in attended and unattended states before (raw) and after (residual) subtracting the first eigenmode (mean $\pm$ SD in black). Two-sided Wilcoxon rank-sum test (attended vs unattended): mean covariance,**C**, $P = 0.62$, **H**, 0.90; residual: **C**, $P = 0.34$, **H**, $P = 0.53$. **D, I**, Pairwise covariance of spike counts as a function of the distance between the pair. **E, J**, The distance dependence functions of the first five covariance components computed from the Factor analysis of the unattended state, normalized by the magnitude at distance 0. Shaded regions are SEM. In networks with fast inhibition (**A-E**), the inhibitory synaptic constants in Layer 3 were $\tau_{\mathrm{ir}} = 0.5$ ms and $\tau_{\mathrm{id}} = 1$ ms. In networks with broad inhibitory projections (**F-J**), the excitatory and the inhibitory projection widths of Layer 3 were $\alpha_{\mathrm{e}}^{(3)} = 0.1$ and $\alpha_{\mathrm{i}}^{(3)} = 0.2$, respectively. Other parameters were the same as the original model. The feedforward connections from Layer 2 to Layer 3 were the same as those in simulations of the original model. The statistical methods were the same as that used in the original model (Fig. 4E and Fig. 5). **A-C, F-H**, $n = 80$ sessions of 50 neurons each. **D, I, E, J**, $n = 80$ sessions of 500 neurons each.

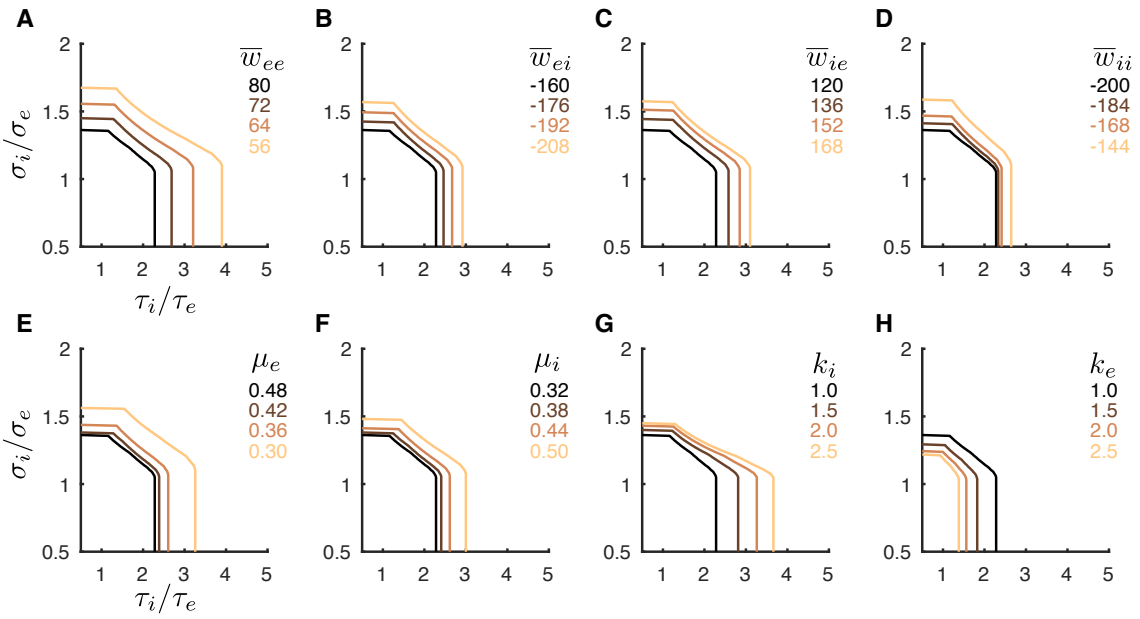

Figure S7: **Related to Figure 5A. The dependence of the stability region on various parameters of the spatio-temporal firing rate model. A**, The stability region expands with a decrease in the mean strength of recurrent excitation, $\overline{w}_{ee}$. Regions below the curves are (τ_i, σ_i) parameters that produce stable firing rates in the spatio-temporal firing rate model. The black curve is the boundary of the stable region in Figure 5A (gray). **B-H**, Same as panel A for: (**B**), the mean connection strength from the inhibitory neurons to the excitatory neurons, $\overline{w}_{ei}$; (**C**), the mean connection strength from the excitatory neurons to the inhibitory neurons, $\overline{w}_{ie}$; (**D**), the recurrent inhibition $\overline{w}_{ii}$; (**E**), the input current to the excitatory neurons μ_e; (**F**), the input current to the inhibitory neurons μ_e; (**G**), the slope of the inhibitory transfer function k_i; (**H**), the slope of the excitatory transfer function k_e. The lightest orange curve in panel F corresponds to the black dashed curves in Figure 5A.

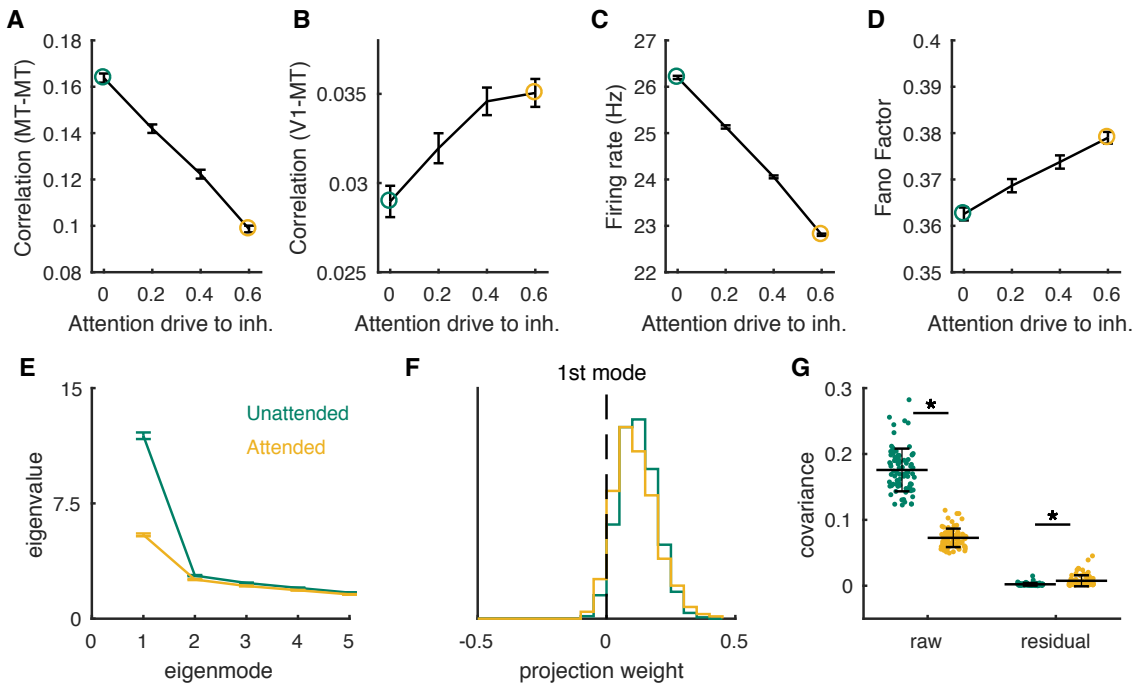

Figure S8: **Related to STAR Methods. Networks without the slow feedforward excitatory current also produces low- dimensional shared variability.** The feedforward synapses from Layer 2 to Layer 3 have the same kinetics as the recurrent excitatory synapses, i.e. $\eta_F^{(3)}(t) = \eta_{\rm e}(t)$ with $\tau_{\rm er} = 1$ ms and $\tau_{\rm ed} = 5$ ms. The recurrent connection width of Layer 3 is $\alpha_{\rm rec}^{(3)} = 0.2$ and the feedforward connection width from Layer 2 to Layer 3 is $\alpha_{\rm ffwd}^{(3)} = 0.05$. The feedforward connection strengths from Layer 2 to Layer 3 are $J_{\rm eF}^3 = 10$ mV and $J_{\rm iF}^3 = 1$ mV. Other parameters were the same as the original model. Attention is also modeled as a depolarization to MT inhibitory neurons (μ_I mV/ms). **A**, Mean spike count correlation ($r_{\rm SC}$) of excitatory neuron pairs in MT decreases with attentional modulation (μ_I mV/ms). Circles denote the parameters used for the unattended state ($\mu_I = 0$, green) and the attended state ($\mu_I = 0.6$, orange). Error bars are SEM ($n = 50$ simulations of 20 seconds each). **B**, Mean $r_{\rm SC}$ between the excitatory neurons in MT and the excitatory neurons in V1 increases with attention. **C** Same as panel A for firing rate change. **D** Same as panel A for Fano factor change. **E**, The first five largest eigenvalues of the shared component of the spike count covariance matrix ($n = 80$ sessions of 50 neurons each). Error bars are SEM. **F**, The vector elements for the first (dominant) eigenmode. **G**, The mean covariance from each session in attended and unattended states before (raw) and after (residual) subtracting the first eigenmode (mean $\pm$ SD in black). Two-sided Wilcoxon rank-sum test (attended vs unattended): mean covariance, $P = 9.4 \times 10^{-28}$; residual: $P = 2.2 \times 10^{-11}$.

Session #	Unit #	Trial # (Att.)	Trial # (Unatt.)	Session #	Unit #	Trial # (Att.)	Trial # (Unatt.)
1	50	702	361	37	21	364	744
2	80	361	702	38	38	744	364
3	27	408	349	39	22	306	338
4	49	349	408	40	44	338	306
5	26	795	1038	41	46	516	480
6	35	1038	795	42	62	480	516
7	56	280	624	43	43	620	509
8	73	624	280	44	55	509	620
9	26	1109	992	45	43	601	543
10	47	992	1109	46	52	543	601
11	30	673	824	47	40	484	658
12	47	824	673	48	62	658	484
13	33	597	622	49	37	452	320
14	79	622	597	50	63	320	452
15	24	242	619	51	37	548	657
16	51	619	242	52	65	657	548
17	14	261	583	53	39	369	282
18	39	583	261	54	57	282	369
19	26	399	666	55	41	413	501
20	46	666	399	56	53	501	413
21	25	696	800	57	35	601	486
22	46	800	696	58	49	486	601
23	30	751	861	59	36	392	414
24	56	861	751	60	61	414	392
25	25	670	724	61	35	483	421
26	50	724	670	62	50	421	483
27	28	470	414	63	32	592	433
28	52	414	470	64	55	433	592
29	25	1728	1843	65	35	692	439
30	50	1843	1728	66	53	439	692
31	32	515	643	67	36	655	437
32	67	643	515	68	49	437	655
33	16	412	566	69	36	390	352
34	42	566	412	70	47	352	390
35	19	752	730	71	32	507	385
36	42	730	752	72	71	385	507

Table S1: Related to Fig. 1C, Fig. 6B,D and Fig. S1. The number of units and trials of each recording session for the Factor analysis of the multi-electrode recordings from V4 (Cohen and Maunsell, 2009).

distance (mm)	0.4000	0.5657	0.8000	0.8944	1.1314	1.2000	1.2649
pair #	6368	5290	4995	8373	3314	3791	6247
distance (mm)	1.4422	1.6000	1.6492	1.6971	1.7889	2.0000	
pair #	4926	2622	4522	1883	3635	4223	

Table S2: Related to Fig. 6B,D. The number of pairs at each distance value for computing the distance dependent functions of total covariance as well as the first covariance components from factor analysis of the multi-electrode recordings from V4 (Cohen and Maunsell, 2009).