Internally generated population activity in cortical networks hinders information transmission

Chengcheng Huang^{1,2,3}, Alexandre Pouget⁴ & Brent Doiron^{2,3*}

¹Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA, USA

²Department of Mathematics, University of Pittsburgh, Pittsburgh, PA, USA

³Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

⁴Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland

*To whom correspondence should be addressed; E-mail: bdoiron@pitt.edu.

Abstract

How neuronal variability impacts neuronal codes is a central question in systems neuroscience, often with complex and model dependent answers. Most population models are parametric, with a tacitly assumed structure of neuronal tuning and population-wide variability. While these models provide key insights, they purposely divorce any mechanistic relationship between trial average and trial variable neuronal activity. By contrast, circuit based models produce activity with response statistics that are reflection of the underlying circuit structure, and thus any relations between trial averaged and trial variable activity are emergent rather than assumed. In this work, we study information transfer in networks of spatially ordered spiking neuron models with strong excitatory and inhibitory interactions, capable of producing rich population-wide neuronal variability. Motivated by work in the visual system we embed a columnar stimulus orientation map in the network and measure the population estimation of an orientated input. We show that the spatial structure of feedforward and recurrent connectivity are critical determinants for population code performance. In particular, when network wiring supports stable firing rate activity then with a sufficiently large number of decoded neurons all

available stimulus information is transmitted. However, if the inhibitory projections place network activity in a pattern forming regime then the population-wide dynamics compromise information flow. In total, network connectivity determines both the stimulus tuning as well as internally generated population-wide fluctuations and thereby dictates population code performance in complicated ways where modeling efforts provide essential understanding.

Introduction

A prominent feature of cortical response to sensory stimuli is that neuronal activity varies significantly across presentation trials^{1,2}, even when efforts are taken to control or account for variable animal behavior^{3–6}. A component of this variability is coordinated among neurons in a brain area, often leading to shared fluctuations in spiking activity^{7–11}. How stimulus processing is affected by this large, population-wide neuronal variability is a longstanding question in both experimental and theoretical neuroscience communities.

Recording from neuronal populations while simultaneously monitoring an animal's behavior during a structured task offers a glimpse into how neuronal activity supports computation. Correlations between spike counts from pairs of neurons in response to repeated stimulus, often referred to as noise correlations, are modulated by a variety of cognitive factors that are known to affect task performance⁴. For example, noise correlations decrease with animal arousal^{12,13} or task engagement¹⁴. In the visual pathway noise correlations are decreased when spatial attention is directed into the receptive field of a recorded population^{15,16}. In addition, perceptual learning^{17,18} and visual experience¹⁹ can also result in an attenuation of noise correlations. The common theme in all of these studies is that a reduction in noise correlations co-occurs with cognitive shifts that improve task performance. This supports the often cited idea that shared variability is deleterious to neuronal coding because it cannot be reduced by ensemble averaging^{20,21}, and thus it is expected that its reduction will enhance neural coding.

While the above narrative is appealing, it oversimplifies a long debate in the computational neuroscience community^{20,22–26}. It is popular (and pragmatic) to restrict analysis to a linear decoding of neuronal response, where performance is often measured using the linear Fisher information between the estimated and actual stimulus^{23,27–30}. Linear Fisher information depends on two components: the set of neuronal tuning curves and the population covariance matrix^{23,27,28}. Whether noise correlations degrade or improve stimulus coding depends on the relationship between these components. For example, while it is true that correlations between similarly tuned neurons limit Fisher information^{21,31}, correlations between dissimilarly tuned neurons can actually increase information compared to an asynchronous population^{22,23}. Several more recent studies have built upon this idea and shown how stimulus dependent correlations can also improve linear information transmission^{32–34}. Finally, Moreno-Bote and colleagues³⁵ showed that information is only limited by one specific type of correlations, termed differential correlations, which align with population activity in the direction defined by the gain of population tuning. In all of these varied modelling studies both tuning curves and covariance structure were assumed and thus had a prescribed relation to one another. In this study we adopt an alternative modelling approach and consider biophysically grounded population models, where cellular spiking dynamics and synaptic wiring are assumed, and population response and its variability are emergent properties of the circuit.

A serious obstacle in using circuit models to study neuronal coding is the lack of complete mechanistic theory underlying population-wide variability³⁶. An often used framework assumes neuronal fluctuations are inherited from external sources, and any circuit wiring filters and propagate this variability throughout the population^{37–40}. Along these lines recent modelling work in a feedforward circuit model of primary visual cortex shows that when such inherited variability originates from a noisy stimulus then population-wide correlations which limit information transfer naturally develop throughout the circuit⁴¹. Neuronal variability can also emerge through recurrent interactions within the network, when synaptic weights are large and excitation is dynamically balanced by recurrent

inhibition^{1,42,43}. However, these networks famously support an asynchronous state⁴⁴, making them at the surface ill-suited to probe how internally generated population-wide variability affects neuronal processing. Recent extensions of balanced networks which include structured wiring has provided an understanding of how shared variability can emerge through internal mechanics^{45–52}. These networks are a useful modelling framework to probe how circuit produced shared variability impacts population coding.

In this work, we study information transmission in spatially ordered spiking neuron networks where a balance between excitation and inhibition produce population-wide variability. Modelling simple visual inputs to cortex with a columnar organization of stimulus orientation preference allows a study of how neuronal stimulus tuning and associated population variability are shaped by circuit wiring and modulatory inputs. We found that narrow feedforward projections increase information in the output layer for finite number of decoding neurons, while the recurrent projection width with spatially balanced excitation and inhibition has little effect. Interestingly, the total information is preserved when the network is in a regime where firing rates are stable. In contrast, when the network exhibits coherent internal dynamics through a pattern forming instability in firing rates, the information transfer is largely reduced. Our work thus begins to connect the emerging theory of how neuronal circuits produce trial-variable activity with ongoing theories of how population-wide variability affect stimulus representation.

Results

Our study measures the propagation of stimulus information in a two-layer network of spatially ordered neurons. As motivation our model aims to capture layer (L)4 and L2/3 neurons in primary visual cortex (V1) (Fig. 1A, see Methods), however our model is sufficiently general to capture an arbitrary layered network. We model L4 neuronal spiking activity as a Poisson process with Gabor receptive fields defining the stimulus to firing rate transfer. The orientation preference of L4 neurons

are determined from a superimposed pinwheel orientation map (Fig. 1A, bottom right; see Ref.⁵³). The Gabor visual images are corrupted by spatially independent noise obeying a temporal Ornstein-Uhlenbeck process; this provides a bound on the stimulus information entering our network. L2/3 is modeled as a recurrently connected network of both excitatory and inhibitory spiking neuron models. The connection probability of all recurrent projections within L2/3 and the feedforward projections from L4 to L2/3 decay with distance with spatial widths α_{rec} and α_{ffwd} , respectively (see Methods and Ref.⁴⁷ for details).

Networks with large, balanced excitatory and inhibitory connections are an attractive model framework for cortical activity because they capture several aspects of reported *in vivo* population response. When large synaptic connections are paired with random network wiring they naturally produce significant heterogeneity in spiking activities across the network^{42,54,55}. Indeed, in our spatially ordered network the L2/3 neurons have very heterogeneous tuning curves with various widths and magnitudes (Fig. 1B). This produces a large spread of orientation selectivity across the population (Figs. 1B, 2B), capturing the broad heterogeneity of orientation tuning reported in primary visual cortex⁵⁶.

Balanced networks produce significant dynamic and trial-to-trial spiking variability through internal mechanisms^{42,57}. While balanced networks with disordered connectivity produce asynchronous activity⁴⁴, networks with structured wiring can produce correlated variability^{45,48}, that under certain conditions can be population-wide⁴⁷. These past models were concerned with the mechanics of neuronal variability and did not model a spatially distributed stimulus to drive network response. In our model when L4 columnar stimulus structure is enforced then a broad L2/3 columnar activation is recruited accompanied by significant population-wide trial-to-trial variability (Fig. 1C). Despite the lack of feature-based coupling (i.e connectivity was determined only by the spatial distance between neurons), there is a clear positive relation between pairwise signal and noise correlations (Fig. 1D), matching diverse *in vivo* datasets^{15,17,40,58,59}. This occurs for two reasons. First, the stimulus has a spatial columnar profile, so any spatial wiring does partially overlap with the signal profile. Second, it

Figure 1: Spatially ordered balanced networks generate heterogeneous tuning curves and structured trial-to-trial variability. **A**, Model schematic of a two-layer network of spatially ordered spiking neurons modeling Layer 4 neurons and Layer 2/3 neurons, respectively, from visual area V1. The visual input to the model is a Gabor image with orientation θ . The L4 network consists of Poisson units with Gabor receptive fields. The orientation preferences of L4 neurons are assigned according to a pinwheel orientation map (bottom right). The L2/3 network consists of both excitatory and inhibitory neurons modeled with integrate-and-fire dynamics, all arranged on a unit square. The spatial widths of the feedforward projections from L4 to L2/3 and the recurrent projections within L2/3 are denoted as α_{ffwd} and α_{rec} , respectively. **B**, The L2/3 neurons have heterogeneous orientation tuning curves. Ten examples of tuning curves are shown with different color representing different neurons (smoothed with a Gaussian kernel of width 9°). **C**, The model internally generates trial-to-trial variability. Three trials of network spike counts (200 ms time window) are shown from a network with $\alpha_{\text{ffwd}} = 0.05$ and $\alpha_{\text{rec}} = 0.1$. Images are smoothed with a Gaussian kernel of width 0.01. **D**, Noise correlation matrix with neurons ordered by their preferred stimulus orientation. Responses were simulated for a Gabor input with orientation at $\theta = 0^\circ$ without (left) and with stimulus noise (right).

is well known that spike count correlations increase with the firing rates of the neuron pair^{60,61}, so that when neuron pairs with similar stimulus preference are co-activated then any significant trial-to-trial covariability in their synaptic inputs is better expressed in their output spiking. In our model, the relation between signal and noise correlations occurs in response to Gabor images which are noiseless (Fig. 1D, left) or those contaminated by sensory noise (Fig. 1D, right). The latter case naturally provides feedforward noise correlations⁴¹ with a columnar spatial scale. These externally imposed correlations are reduced for neuron pairs with a stimulus preference that match the driving stimulus, in agreement with past models where correlations are inherited from outside the circuit⁴⁰. In total, our spiking network captures many of the trial-averaged and trial-variable aspects of real cortical population response.

We focus on the network’s ability to discriminate two Gabor images with similar orientations, a paradigm commonly used in experiments⁶². Gabor images with orientation θ are presented to the L4 neurons repetitively with an ON interval of 200 ms and an OFF interval of 300 ms (see Methods). In what follows we consider a decoder that has access to N model L2/3 pyramidal neurons (randomly chosen from the 4×10^4 L2/3 neurons). We simulate the spiking activities of L2/3 neurons and collect spike counts from the observed population, $\mathbf{n} = [n_1, \dots, n_N]$, during each stimulus presentation (ON interval). The linear Fisher information about θ available from $\mathbf{n}$ is defined as:

$$I_F = \mathbf{f}'(\theta)^T \Sigma^{-1} \mathbf{f}'(\theta), \quad (1)$$

where $\mathbf{f}$ is the expectation of $\mathbf{n}$ conditioned on θ (population tuning curve), $'$ denotes differentiation with respect to θ , and Σ is the covariance matrix of $\mathbf{n}$ (noise covariance matrix). Fisher information is a useful metric when considering the estimation of orientation θ from $\mathbf{n}$. Let $\hat{\theta}$ be the optimal linear estimator of the θ for a given Gabor image, then we have that $I_F = 1/\text{Var}(\hat{\theta})$ ⁶³. Put more simply, I_F measures the accuracy of the optimal linear estimator of orientation. In practice we measure the linear Fisher information from the spike counts of the L2/3 excitatory neurons using a bias-corrected

estimation (see Methods Eq. (7) in Ref.⁶⁴).

The spatial scales of connectivity in our network are key determinants of overall population spiking activity in unstimulated conditions^{45–47}. In the next sections we explore how the spatial scales of the feedforward L4 to L2/3 connectivity (α_{ffwd} ; Fig. 1A) and those of the L2/3 to L2/3 recurrent projections (α_{rec} ; Fig. 1A) influence information transfer about stimulus θ within the network.

Narrow feedforward projections increase information saturation rate

Linear Fisher information (Eq. 1) depends on two components: the trial-averaged response gain to stimulus orientation (f') and the trial-to-trial covariance matrix Σ . As we will show, the joint influence of α_{ffwd} on gain f' and covariance Σ sets a conflict between how α_{ffwd} should ultimately affect I_F . A similar situation can easily exist in models where the form of f' and Σ are assumed^{21,23,31–33}. However, in many of these cases the justification for how f' and Σ are related is heuristic, and will depend on the specific parametrization of the response model. While our circuit model also has parameters, such as α_{ffwd} , there is a mechanistic understanding for how these control population responses, and consequently for how they relate f' and Σ to one another.

On one hand, the population averaged tuning curve broadens as α_{ffwd} increases, resulting in reduced orientation gain f' (Fig. 2A). In agreement, the orientation selectivity index (OSI, Eq. 6) of individual neurons decreases with α_{ffwd} (Fig. 2B). This reduction in orientation tuning is due to broader spatial filtering of tuned inputs from L4, where the tuning preferences of neurons are spatially clustered in a columnar organization. In particular, when the connection probabilities of all projections are uniform in space ($\alpha_{\text{ffwd}} \rightarrow \infty$), the columnar structure is not preserved in L2/3 and neurons are least tuned. Hence, considering these changes in tuning selectivity alone predict that the I_F about θ from L2/3 neurons decoding would decrease with α_{ffwd} .

On the other hand, the noise correlations between nearby neurons decrease with increasing α_{ffwd} (Fig. 2C). Rosenbaum and colleagues⁴⁶ showed that for spatially distributed balanced networks where

$\alpha_{\text{ffwd}} < \alpha_{\text{rec}}$ an asynchronous solution does not exist. Rather, network activity is organized so that nearby cell pairs have positive noise correlations while more distant cell pairs have negative correlations (Fig. 2C, pink curves). Previous work on population coding has suggested that positive correlation between similarly tuned neurons can limit information while negative correlation between oppositely tuned neurons can benefit coding^{20–22}. Since the tuning preferences of neurons are spatially clustered with a pinwheel orientation map, nearby neurons are likely to share similar tuning. Therefore, in contrast to the effect of the tuning selectivity which would suggest that I_F could increase with α_{ffwd} , the overall reduction in the spatial structure of pairwise correlations implies that I_F could increase with α_{ffwd} .

We divide our analysis of how I_F depends on α_{ffwd} into two cases: first, when the decoder is restricted to a small ($N \sim 10^2$) populations of neurons, and second for decoders that have access to large ($N \sim 10^4$) populations.

For decoders restricted to small populations the combined effects of reduced tuning selectivity and correlations with larger α_{ffwd} , results in the linear Fisher information being largely reduced as α_{ffwd} increases (Fig. 3A). Consistently, the neural thresholds of single neurons, measured as the ratio of the standard deviation of the spike counts and the derivative of the tuning curve function (σ_i/f'_i), also increase with α_{ffwd} (Fig. 3B). In total, the population code is less informative with larger α_{ffwd} , in line with the single neuron reduction in orientation selectivity (Fig. 2B), despite the associated decrease in pairwise noise correlations (Fig. 2C).

The noise in the Gabor image creates shared fluctuations in the L2/3 neurons that cannot be distinguished from signal, limiting the available information about θ that is possible to be decoded (I_F^{in} ; Fig. 3A dashed line, see Methods). For decoders that have access to large populations the linear Fisher information of the L2/3 neurons saturates to a level close to I_F^{in} , regardless of the feedforward projection width (Fig. 3A). It suggests that the network is very efficient at transmitting information. In particular, even networks with spatially disordered connections ($\alpha_{\text{ffwd}} \rightarrow \infty$), where neurons are

Figure 3: The linear Fisher information saturates faster with smaller feedforward projection width. **A**, The linear Fisher information increases with the number of neurons sampled from the excitatory population in L2/3 (color scheme in top panel). With large number of neurons, the linear Fisher information saturates close to the input information bound from the L4 network (I_F^{in} , dashed line). The input noise on the Gabor images is chosen such that I_F^{in} corresponds to a discrimination threshold of about 1.8 degrees, which is consistent with psychophysical experiments⁶⁵ (see Methods). **B**, The probability density distributions of neural thresholds (σ_i/f'_i) of the L2/3 excitatory neurons.

very weakly tuned, still contain most of the information from the input layer.

Moreno-Bote and colleagues³⁵ explicitly make the observation that while networks may show significant noise correlations, it is the shared fluctuations that align with the direction of response gain that limit information. In particular, they consider the decomposition of the covariance $\Sigma = \Sigma_0 + \epsilon \mathbf{f}' \mathbf{f}'^T$, with ϵ measuring the shared variance along the coding direction ($\mathbf{f}'$). They show that for decoders that have access to a large number of neurons then $I_F \sim 1/\epsilon$. For our network it is not (to our knowledge) possible to formally decompose Σ into $\Sigma_0 + \epsilon \mathbf{f}' \mathbf{f}'^T$, meaning we do not have a way of estimating how network interactions contribute to ϵ . However, the fact that the information eventually saturates close to I_F^{in} for all values of α_{ffwd} indicates that α_{ffwd} does not affect the value of ϵ . This is consistent with our network being in a regime where the origin of information limiting correlations stems from strictly external fluctuations, as has been previously studied⁴¹.

Figure 4: The recurrent projection width has small effects on the linear Fisher information. **A**, The linear Fisher information is slightly lower when the recurrent projection width (α_{rec}) is narrower than the feedforward width (α_{ffwd}) (light shade). When α_{rec} is broader than α_{ffwd} , the saturation curve of the Fisher information overlaps (darker shades). Examples of two α_{ffwd} values are shown (red: $\alpha_{ffwd} = 0.05$; green: $\alpha_{ffwd} = 0.075$). **B**, The probability density distributions of neural thresholds (σ_i/f'_i) of the L2/3 excitatory neurons. **C**, The probability density distributions of orientation selectivity indexes of the L2/3 excitatory neurons. **D**, Pairwise correlation as a function of the distance between a neuron pair.

Recurrent connections with spatially balanced excitation and inhibition have small effects on information transfer

Next we study how the spatial scales of recurrent connections (α_{rec}) affect information transmission. We first consider networks with the same spatial scale of recurrent excitatory and inhibitory projections. As remarked in the previous section, the relative spatial scale between recurrent and feedforward projections has a large impact on the spatial structure of pairwise correlations (Fig. 2C, Ref.⁴⁶). For fixed α_{ffwd} then as α_{rec} increases orientation selectivity increases (Fig. 4C), while pairwise noise correlations become spatially structured (Fig. 4D).

For decoders that consider only small populations ($N \sim 10^2$ to 10^3 neurons) we see a clear difference between how α_{rec} and α_{ffwd} affect I_F . Specifically, I_F is relatively insensitive to α_{rec} (the brown curves scanning over α_{rec} for $\alpha_{\text{ffwd}} = 0.05$ are clustered in Fig. 4A; similar for the green curves with $\alpha_{\text{ffwd}} = 0.075$). Indeed, both the distributions of neural thresholds and OSI show similar ordering as the Fisher information (Fig. 4B,C). Further, the spatial structures of pairwise correlations do not contribute to a difference in Fisher information (Fig. 4D), again suggestive that the population correlations are not information limiting. In agreement, for decoders with access to large populations I_F saturates to I_F^{in} . Overall, for spatially balanced recurrent excitatory and inhibitory projections the specific scale of α_{rec} is of little importance when considering information transfer.

Broad inhibitory projections reduce information

Our cortical network can exist in one of two regimes. The preceding sections consider a regime where circuit parameters are such that the firing rate dynamics across the network are stable, despite pairwise correlations being structured. The second regime is where the firing rates are no longer dynamically stable, and large rate fluctuations are produced through internal interactions (as we will formalize below). In past work⁴⁷ we showed how model networks in this second regime produced the low-dimensional, population-wide shared variability characteristic of a variety of cortical areas^{9–11,47,66–69}. In this section we consider how networks in this second regime transfer information across layers.

Dynamical networks with spatially ordered coupling can produce rich spatio-temporal patterns of activity⁷⁰. To explore how the network circuitry in our cortical model determines its pattern forming dynamics we first implement two simplifications. First, we restrict our analysis to an associated firing rate model^{71,72}, sharing the same spatial connectivity as our network of model spiking neurons (see Methods). This greatly simplifies any linear stability analysis, and it provides qualitatively similar macroscopic network firing rates when spiking neuron models are in a fluctuation driven regime^{73,74}. Second, we analyze the dynamics of the network when driven by spatially uniform inputs. This

provides a spatial translational symmetry in the network needed for Fourier analysis.

A stable, spatially uniform solution of the firing rate model corresponds to an asynchronous state in the network of spiking neuron models. We linearize the firing rate network dynamics around this uniform solution and obtain a set of eigenvalues and associated eigenmodes. This eigenstructure governs the linearized firing rate dynamics near the uniform solution. Each eigenmode has a wavenumber indicating its spatial scale across the network; the zero wavenumber eigenmode describes a spatially uniform solution, while higher wavenumber eigenmodes contribute to spatially structured solutions (at the spatial period of the wavenumber). The stability of the solution at each eigenmode is given by the sign of the real component of the associated eigenvalue: negative (positive) eigenvalues imply dynamics about that eigenmode are stable (unstable). If all eigenmodes are stable then we say that the solution is stable.

Networks with spatially balanced recurrent excitatory and inhibitory projections ($\sigma_e = \sigma_i$) and strong static inputs to the inhibitory neurons (μ_i) have a stable spatially uniform solution (Figs. 5A, grey region; 5B, orange curves). In this regime, the network dynamics is approximately linear with weak perturbations, meaning that the network fluctuations are linearly related with input fluctuations. Since linear Fisher information is conserved under an invertible linear transformation (see Discussion), a network in a stable regime preserves almost all of its input information. Indeed, as shown in previous sections, I_F converges to I_F^{in} for large N decoders, and this convergence is independent of the connectivity scales of both the feedforward and the recurrent pathways (Fig. 3A, Fig. 4A).

When the excitatory and inhibitory projections differ in spatial scale, recurrent connectivity can have a large impact on network dynamics. As the inhibitory projection width (σ_i) increases the uniform solution in the spatial firing rate model loses stability at a band of non-zero wavenumbers (Fig. 5A,B). This is via a Turing-Hopf bifurcation⁷⁰, a pattern-forming transition common to spatially distributed neuronal network models^{71,72,75}. The unstable eigenmodes that emerge from the bifurcation contribute to population dynamics confined to the spatial scale of their wavenumbers.

Figure 5: Networks with broad inhibitory projections exhibit pattern forming dynamics. **A**, Bifurcation diagram of a firing rate model as a function of the inhibitory projection width σ_i and the depolarization current to the inhibitory neurons (μ_i). The excitatory projection width and the drive to the excitatory neurons are fixed at $\sigma_e = 0.1$ and $\mu_e = 0.5$, respectively. Color represents the wavenumber whose eigenvalue has the largest positive real part and the gray region is marked stable since all eigenvalues have a negative real part. **B**, The maximum real part of eigenvalues as a function of wavenumber for increasing σ_i ($\sigma_e = 0.1$), when $\mu_i = \mu_e$. Nonzero wavenumbers lose stability as σ_i increases, indicating that the network will exhibit firing rate dynamics with spatial scales of the unstable eigenmodes of the uniform solution. **C**, The power spectrum at temporal frequency $\omega = 0$ for different spatial Fourier modes (n_x, n_y) of the spontaneous spiking activity from the spiking neuron networks with $\sigma_i = 0.1$ (left, the nonzero spatial frequency power peaks around wavenumber 3.6) and $\sigma_i = 0.3$ (middle). Right: The average power at $\omega = 0$ across wavenumbers ($k = \sqrt{n_x^2 + n_y^2}$). **D**, Activities of spiking neuron networks with different σ_i when driven by a Gabor image (right: three trials of spike counts within 200 ms window; left: mean spike counts). Images are smoothed with a Gaussian kernel of width 0.01. Other parameters for the spiking neuron networks (panels C, D) are $\sigma_e = 0.1$, $\alpha_{\text{ffwd}} = 0.05$, and feedforward strengths $J_{\text{eF}} = 240$ mV and $J_{\text{iF}} = 400$ mV.

Such pattern forming transitions are also observed in the spontaneous activity of spiking neuron networks^{45,47,51,76,77}. To quantify the spatial scales of population activity we compute the spatial power of population spiking (at zero temporal frequency) in the spontaneous state. When σ_i equals σ_e the power spectrum is small (Fig. 5C, right, orange curve) showing a band at a wavenumber of approximately 3.6 (Fig. 5C, left), reflecting the spatial scale of the feedforward projections (see Supp Fig 3 in Ref.⁴⁶). As σ_i increases, there is a large increase in power at some non-zero wavenumbers (Fig. 5C), reflecting large, internally generated fluctuations at non-zero spatial frequencies.

When the spiking network is driven by a Gabor image, the internal dynamics of the network interacts with the spatial patterns of the feedforward inputs. In the stable regime, where excitation is spatially balanced with inhibition ($\sigma_i = \sigma_e$), the spatial patterns of the network activity are similar across trials and are inherited from the input spatial scale (Fig. 5D, first row). As σ_i increases, the internal pattern forming dynamics of the network become dominant over the feedforward input patterns (Fig. 5D, top to bottom rows). The broad inhibition suppresses significant activity, resulting in only a few stimulus evoked areas compared to the spatially balanced case. The locations of the evoked areas vary from trial to trial, hence introducing excessive trial-to-trial variability in the population activity patterns. The mean population activity pattern across trials also degrades as σ_i increases (Fig. 5D, left). In total, this suggests that there will be information loss in networks with larger σ_i .

Indeed, the Fisher information I_F about stimulus θ is markedly reduced when the network loses stability through broader inhibitory projections (Fig. 6A). The orientation selectivity of neurons only changes slightly with σ_i (Fig. 6B), which means that a reduction in tuning selectivity cannot be the primary reason for the decrease in I_F . However, the magnitude of pairwise correlations increases drastically as σ_i increases. Due to the internal pattern forming dynamics in networks with large σ_i , nearby neurons are strongly correlated while neuron pairs that are separated by half the distance between active zones are strongly anti-correlated. Consider the decomposition of network covariance $\Sigma = \Sigma_0 + \epsilon \mathbf{f} \mathbf{f}'^T$ (see Ref.³⁵), recalling that in the limit of large decoded populations (N) we have that

Figure 6: **Broad inhibitory projections reduce information flow.** **A**, The linear Fisher information from the L2/3 excitatory neurons is largely reduced as the inhibitory projection width (σ_i) increases. **B**, The probability density distributions of orientation selectivity indexes of the excitatory neurons from L2/3. **C**, Pairwise correlations as a function of distance between a neuron pair in L2/3.

the Fisher information $I_F \sim 1/\epsilon$, so that ϵ measures the strength of information limiting correlations. In the stable regime I_F converges to the input information I_F^{in} (Fig. 6A, orange curve), implying that all information limiting correlations are inherited through the feedforward pathway (as reported earlier in Figs. 3 and 4). However, for broader inhibition it appears that for large N the information I_F saturates to a level that is below I_F^{in} . This argues that a component of the internally generated network covariability actually limits information transfer (i.e. contributes to ϵ). While our simulations give strong support to this conclusion we admit that since it is computationally prohibitive to consider N beyond 10^4 neurons we cannot verify how I_F truly converges. In total, the pattern-forming dynamics in networks with broad inhibitory projections largely reduce the transmitted information.

Depolarizing inhibitory neurons improves information flow

Even when excitatory and inhibitory projections are spatially matched a network can nevertheless have an unstable uniform solution. This occurs when the static input to the inhibitory neurons (μ_i) is low (Fig. 7A, blue region) and the firing rate network is unstable through the eigenmode with zero wavenumber (Fig. 7B). In this case the stability of the uniform solution is mediated by a Hopf bifurcation, through which bulk population-wide firing dynamics are produced (Fig. 7A, transition from

Figure 7: Lack of inhibitory drive gives rise to network-wide fluctuations. **A**, Bifurcation diagram of a firing rate model as a function of the inhibitory projection width σ_i and the depolarization current to the inhibitory neurons. Same as Fig. 5A. **B**, The real part of eigenvalues as a function of wave number for increasing μ_i ($\mu_e = 0.5$) when $\sigma_i = \sigma_e$. The zero wave number loses stability when μ_i is small, indicating that the network will exhibit network-wide nonlinear dynamics. **C**, The power spectrum at temporal frequency $\omega = 0$ for different spatial Fourier modes (n_x, n_y) of the spontaneous spiking activity from the spiking neuron networks with $\mu_i = 0$ (left) and $\mu_i = 1.2$ (middle). Right: The average power at $\omega = 0$ across wave number ($k = \sqrt{n_x^2 + n_y^2}$). **D**, Activities of spiking neuron networks with different μ_i when driven by a Gabor image (right: three trials of spike counts within 200 ms window; left: mean spike counts). Images are smoothed with a Gaussian kernel of width 0.01. Other parameters for the spiking neuron networks (panels C, D) are $\sigma_e = \sigma_i = 0.1$, $\alpha_{\text{ffwd}} = 0.05$, $J_{\text{eF}} = 140$ mV, $J_{\text{iF}} = 0$ mV, and the tonic current to excitatory neurons is $\mu_e = 0$.

Figure 8: Depolarizing inhibitory neurons improves information flow. **A**, The linear Fisher information increases with depolarizing the inhibitory neurons (μ_i). The saturation curves of the linear Fisher information overlap for $\mu_i \geq 0.8$. **B**, The probability density distributions of orientation selectivity indexes of the L2/3 excitatory neurons. **C**, Pairwise correlations as a function of the distance between neuron pairs from the L2/3 excitatory population. The feedforward strengths are $J_{eF} = 140$ mV and $J_{IF} = 0$ mV, and the tonic current to excitatory neurons is $\mu_e = 0$.

blue to grey regions), as opposed to the spatially confined dynamics produced via a Turing instability (Fig. 5). The spiking network shows a similar instability; when the static current to the inhibitory neurons is small the spontaneous activities of spiking neuron networks exhibit large power at zero wave number, indicating large magnitude global fluctuations (Fig. 7C). Further, the global fluctuations in the spontaneous activities of the spiking neuron network are largely suppressed with more input to the inhibitory neurons (Fig. 7C). In response to a stimulus θ the population activity patterns show large fluctuations in the overall spiking activity level when μ_i is small (Fig. 7D). However, unlike the case for broad inhibition, since the instability is at zero wavenumber then no spatially patterned internal dynamics occur that would compete with the stimulus evoked spatial patterns. Consequently, the spatial patterns of the evoked activities are similar across trials (Fig. 7D).

The Fisher information about θ increases with μ_i as the network becomes more stable (Fig. 8A). Once the network is in the stable regime, a further increase in μ_i has little effect on the I_F (Fig. 8A, $\mu_i = 0.8$ compared $\mu_i = 1.2$). To understand this we again consider how response gain and population covariability are affected by μ_i . First, as μ_i increases the tuning curves of the excitatory

neurons are sharpened (Fig. 8B). Second, since the unstable dynamics correlate the whole network, they give rise to positive correlations between neurons across long distances. As μ_i increases these population-wide fluctuations are suppressed and overall neurons are less correlated (Fig. 8C). Indeed, this combination could in principle conspire to produce the increase in information. However, we note that the impact of global fluctuations upon information transfer is much less than that of the spatially confined fluctuations induced by broad inhibition (compare Figs. 6A and 8A).

The small improvement in information transfer between two layers can be compounded during multi-layer processing. We consider a four-layer network model with the same spatial wiring (Fig. 9A). The first and the second layer are the same as the L4 and L2/3 networks, respectively, in previous model. We compare the two networks with different inhibitory biases (μ_i) applied to the neurons in Layers 2, 3, and 4. The feedforward strength was adjusted such that the firing rates were similar across layers 2 to 4 (Supp Fig. S1C), as well as between the two conditions of μ_i . As expected the information about θ decreases as it propagates to higher-order layers (Fig. 9B,C), this is simply an expression of the data-processing inequality⁷⁸. This suggests that information becomes more diffused in higher-order layers (Fig. 9B,C); indeed, the pairwise correlations are higher and the thresholds of single neurons are larger in higher-order layers (Supp Fig. S1A,B). However, the information deteriorates much faster across layers in networks with smaller μ_i (Fig. 9B,C), not only due to the compounding effect, but also because of the increase in synchrony in spike trains among neurons for small μ_i (Supp Fig. S1D). Since the inhibition is insufficient to cancel correlations at each layer, the synchrony builds up as signal propagates to higher-order layers^{79,80}. With stronger input to the inhibitory neurons, excitation is more balanced by inhibition, leading to stable asynchronous dynamics (Supp Fig. S1D). In total, the improvement in the Fisher information by increasing μ_i dramatically increases with the number of processing layers (Fig. 9D).

Figure 9: Information deteriorates in a multi-layer model with internal dynamics. **A**, Schematic of a multi-layer model. Layer 2-4 are recurrent networks of excitatory and inhibitory neurons, with the same projection widths of both recurrent and feedforward connections ($\alpha_{\text{rec}} = 0.1$ and $\alpha_{\text{ffwd}} = 0.05$). Recurrent connections of layer 2 and layer 3 are not shown for better illustration. Layer 1 network is the same as the V1 L4 network in Fig. 1A. **B**, The linear Fisher information is much more reduced in the higher-order layers when there is no input to the inhibitory neurons ($\mu_i = 0$). Dashed line is the total input Fisher information from layer 1 ($F_{I_{\text{in}}}$). **C**, Same as **B** with larger input to the inhibitory neurons ($\mu_i = 1.2$). **D**, The ratio between the Fisher information in networks with $\mu_i = 1.2$ (F_{I_2}) and that with $\mu_i = 0$ (F_{I_1}) across layers, encoded in a large neuron ensemble size ($N = 12800$). The feedforward strengths are adjusted such that the firing rates are similar between the two conditions of μ_i , and are maintained across layers (Supp Fig. S1C).

Discussion

The performance of a population code depends on the relationship between the individual tuning curves and the shared variability among neurons^{22,26,27}. Previous studies have tacitly assumed a prescribed structure between tuning and variability, allowing a dissection of their respective impacts on population coding^{23,31–33,81}. While this approach has given some critical insights, an understanding of how the mechanics of a neural circuit impact population coding has remained elusive. It is well known that the trial averaged^{82,83} and trial-to-trial variability^{29,41,46,47} of a population response are shaped by both feedforward and recurrent network connectivity. In this study we estimated the information available (to a linear decoder) about a simple, one dimensional stimulus that is contained in the activity of a recurrently coupled network of spatially ordered spiking neuron models. We show that simply understanding how pairwise noise correlations are determined through circuitry is insufficient to predict how information will be transferred across layers. Rather, an analysis of how circuitry determines the stability of network firing activity provides a better understanding of information transfer.

A central result of our paper is that in the limit of a large number of decoded neurons the linear Fisher information I_F can either saturate to the input information I_F^{in} for networks with stable firing rate dynamics (Figs. 3,4), or fall short of that bound for networks in a pattern forming regime (Figs. 6,8,9). To appreciate why this is the case it is useful to recall a simple fact about how linear systems encode inputs. Consider a one dimensional stimulus θ that drives a vector of noisy inputs $\mathbf{x}(\theta) = \mathbf{s}(\theta) + \eta$, where the noise process η is zero mean with covariance matrix Σ_x . Let the output $\mathbf{y}(\theta)$ be a linear mapping of the input; specifically we take $\mathbf{y}(\theta) = L\mathbf{x}(\theta)$ for some invertible matrix L . Then we have that the linear Fisher information available to a decoder from $\mathbf{y}$, termed I_F^y , is simply:

$$I_F^y(\theta) = \mathbf{y}'^T \Sigma_y^{-1} \mathbf{y}' = \mathbf{x}'^T L^T [L \Sigma_x L^T]^{-1} L \mathbf{x}' = \mathbf{x}'^T \Sigma_x^{-1} \mathbf{x}' = I_F^x(\theta). \quad (2)$$

In other words, deterministic linear mappings do not distort or degrade information transfer (when the decoder is also linear). We remark that $I_F^y(\theta)$ does not depend on the linear mapping L .

When the cortical network is in the stable regime and any input noise is weakly correlated then we can linearize network dynamics about an operating point^{37,73,84,85}. In this regime the noisy spike counts during a trial, $\mathbf{n}(\theta)$, in response to stimulus θ obey:

$$\mathbf{n}(\theta) = G (J\mathbf{n}(\theta) + \mathbf{x}(\theta)) + \xi \implies \mathbf{n}(\theta) = (I - GJ)^{-1} (G\mathbf{x}(\theta) + \xi).$$

Here $G_{ij} = g_i\delta_{ij}$ is a diagonal matrix of the neuron response gains (depends on the operating point) and J_{ij} is the synaptic coupling from neuron j to i . In addition to stimulus noise in $\mathbf{x}(\theta)$ there is an internal noise process ξ that can reflect any trial fluctuations accrued from the digitization from continuous inputs to spike counts³⁵ or a ‘background’ noise that is present even when coupling and stimulus are absent³⁷; in all cases ξ is independent across neurons. If sufficiently large numbers of neurons are decoded or sufficiently long observation times are used then the contribution of ξ to population statistics is negligible and we have that $\mathbf{n} \approx (I - GJ)^{-1} G\mathbf{x}$, or in other words the population output is a linear function of the input. Given this linearization and Eq. (2) we would then expect that $I_F \approx I_F^{\text{in}}$, so long as the coupling J and input $\mathbf{x}$ provides a stable operating point around which to linearize. In contrast to this stable case, when the network is in a pattern forming regime then there is not a stable operating point about which to linearize. This means that Eq. (2) is no longer valid and we would expect that $I_F < I_F^{\text{in}}$. In total, this argument sketches how an understanding of the network firing rate stability translates quite naturally to an understanding of what circuit conditions will allow the network to faithfully transfer all the information available (to a linear decoder) in an input signal.

Broad inhibitory projections, also referred to as lateral inhibition, is a common circuit structure used to support various neural computations, such as working memory^{71,86}, sharpening of tuning curves^{29,82}, formation of clustered networks⁸⁷, and the periodic spatial receptive fields of grid cells⁸⁸. In this work, we show that the pattern-forming dynamics generated by lateral inhibition degrade information transmission, which is consistent with previous results²⁹. Anatomical measurements

of local cortical circuitry in visual cortex show that excitatory and inhibitory neurons project with similar spatial scales^{89,90}. Our results suggest that such spatially balanced excitatory and inhibitory projections are important for maintaining faithful representations of sensory stimuli. Nevertheless, lateral inhibition may be important for transforming sensory information into task-related information in higher order cortices, such as prefrontal cortex.

Information transmission in feedforward networks has been studied in several models. Bejjanki and colleagues⁹¹ show that by improving feedforward template matching of Gabor filters, perceptual learning can increase the information gleaned from a presented image. In our model, the L4 network is similar to such models^{41,91}, where information loss depends on the Gabor filters of the L4 neurons. In contrast, the feedforward projections from L4 to L2/3 neurons in our model are random and expansive (meaning there are much more neurons in L2/3 than L4), which is known to minimize information loss^{30,92}. Therefore, there is little information loss in the feedforward projections from L4 to L2/3 in our model, regardless of the projection width. In addition, Renart and van Rossum⁹³ and Zylberberg and colleagues⁹² study information transmission with external noise added to the output layer, and compute the optimal connectivity and input covariance, respectively, that maximize the information in the output layer. In our model, there is no external noise imposed on the L2/3 neurons. All the neuronal variability in L2/3 network is either internally generated or inherited from the feedforward inputs^{42,46,47}.

Our results can potentially explain the improvement in discriminability by selective attention. Spatial attention has been shown to significantly reduce the global fluctuations in cortical recordings and meanwhile improve animals' performance in orientation change detection task^{15,16,94,95}. However, the magnitude of global fluctuations is found to have little effect on information, and instead information is only limited by input noise^{35,41,81}. Recently, we developed a circuit model where large-scale wave dynamics give rise to low-dimensional shared variability⁴⁷, thus capturing properties of population recordings in visual cortex^{9,58,66,68,96}. By depolarizing the inhibitory neurons, the attentional modula-

tion in our model stabilizes the network and decreases noise correlations. Therefore, the present study suggests that attention can improve information flow by quenching the internal turbulent dynamics in the recurrent circuit (Fig. 8A).

Spontaneous dynamics of cortex have been shown to mimic the response patterns evoked by external stimuli^{97–99}. It has been hypothesized that the spontaneous dynamics reflect the prior distribution of the stimulus statistics, which is critical for optimal Bayesian inference^{99,100}. Interestingly, our results show that the internal dynamics of recurrent networks do not improve information flow and can reduce it drastically when they are excessive. This is consistent with the correlation between the reduction in global fluctuations in population activity by attention and learning, and the improvement in animal’s performance^{17,18}. However, the spontaneous dynamics in our model is not related to the stimulus since we only consider spatial wiring in the present study. Feature dependent wiring can generate spontaneous dynamics that resemble evoked responses¹⁰¹, and is likely to affect information transmission differently. The exact role of spontaneous dynamics in information processing as well as neural computation remains to be elucidated in future studies.

Methods

Network model description The network consists of two stages, modeling for the layer 4 (L4) and layer 2/3 (L2/3) neurons in V1 respectively (Fig. 1A). Neurons on the two layers are arranged on a uniform grid covering a unit square $\Gamma = [-0.5, 0.5] \times [-0.5, 0.5]$. L4 population is modeled as in Ref.⁴¹. There are $N_x = 2,500$ neurons in L4. The firing rate of each neuron is

$$r_i(\theta, t) = [F_i \times \tilde{I}(\theta, t)]_+$$

where $F_i = F_{(x,y)}(\theta_{\text{pref}}^i)$ is a Gabor filter and $\tilde{I}(\theta, t)$ is a Gabor image corrupted by independent noise following Ornstein-Uhlenbeck process.

$$\tilde{I}(\theta, t) = I(\theta) + \xi(t) \quad \text{and} \quad \tau_n d\xi_i = -\xi_i dt + \sigma_n dW,$$

with $\tau_n = 40$ ms and $\sigma_n = 3.5$. The orientation θ is normalized between 0 and 1. Spike trains of L4 neurons are generated as inhomogeneous Poisson process with rate $r_i(t)$.

The Gabor image is defined on Γ with 25×25 pixels ($\Delta x = 0.04$),

$$I_{(x,y)}(\theta) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda}(x \cos(\theta\pi) + y \sin(\theta\pi)) + \phi\right).$$

The size of the spatial Gaussian envelope was $\sigma = 0.2$, and the spatial wavelength was $\lambda = 0.6$ and phase was $\phi = 0$. The Gabor filters were

$$F_{(x,y)}(\theta_{\text{pref}}) \propto \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda}(x \cos(\theta_{\text{pref}}\pi) + y \sin(\theta_{\text{pref}}\pi)) + \phi\right),$$

with the same σ , λ and ϕ as the image. The orientation map was generated using the formula from Ref.⁵³ (Supp Materials Eq. 20). The preferred orientation at (x, y) is $\theta_{\text{pref}}(x, y) = \text{angle}(z(x, y))/(2\pi)$ and

$$z(x, y) = \sum_{j=0}^{n-1} \exp\left(i \frac{2\pi}{\Lambda} (l_j (\cos(j\pi/n)x + \sin(j\pi/n)y) + \phi_j)\right), n = 30$$

where $\Lambda = 0.2$ is the average column spacing, $l_j = \pm 1$ is a random binary vector and the phase ϕ_j is uniformly distributed in $[0, 2\pi]$.

The L2/3 network consists of recurrently coupled excitatory ($N_e = 40,000$) and inhibitory ($N_i = 10,000$) neurons. Each neuron is modeled as an exponential integrate-and-fire (EIF) neuron whose membrane potential is described by:

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

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, $V_{\text{th}} = -10$ mV, $\Delta_T = 2$ mV, $V_{\text{re}} = -65$ mV and $\tau_{\text{ref}} = 1.5$ ms. Inhibitory neurons are the same except $\tau_m = 10$ ms, $\Delta_T = 0.5$ mV and $\tau_{\text{ref}} = 0.5$ ms. The total current to each neuron is:

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

where $N = N_e + N_i$ is the total number of neurons in the network. Postsynaptic current is

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

where $\tau_{er} = 1$ ms, $\tau_{ed} = 5$ ms for excitatory synapses and $\tau_{ir} = 1$ ms, $\tau_{id} = 8$ ms for inhibitory synapses. The feedforward synapses from L4 to L2/3 have the same kinetics as the recurrent excitatory synapse, i.e. $\eta_F(t) = \eta_e(t)$.

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 measured on Γ with periodic boundary condition:

$$p_{\alpha\beta}(\mathbf{x}, \mathbf{y}) = \bar{p}_{\alpha\beta} g(x_1 - y_1; \sigma_\beta) g(x_2 - y_2; \sigma_\beta).$$

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

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

is a wrapped Gaussian distribution. A presynaptic neuron is allowed to make more than one synaptic connection to a single postsynaptic neuron.

The mean recurrent connection probabilities were $\bar{p}_{ee} = 0.01$, $\bar{p}_{ei} = 0.04$, $\bar{p}_{ie} = 0.03$, $\bar{p}_{ii} = 0.04$, and the recurrent synaptic weights were $J_{ee} = 80$ mV, $J_{ei} = -240$ mV, $J_{ie} = 40$ mV and $J_{ii} = -300$ mV. The feedforward connection probabilities were $\bar{p}_{eF} = 0.1$ and $\bar{p}_{iF} = 0.05$. The feedforward connection strengths were $J_{eF} = 240$ mV and $J_{iF} = 400$ mV for Figure 1- 6, and $J_{eF} = 140$ mV and $J_{iF} = 0$ mV for Figure 7 and 8. The static inputs to inhibitory neurons were $\mu_i = 0$ mV/ms for Figure 1- 6, and $\mu_i = 0, 0.4, 0.8, 1.2$ mV/ms for Figure 7 and 8. The static inputs to excitatory neurons were $\mu_e = 0$ mV/ms for all simulations.

The feedforward connection widths, α_{ffwd} , were 0.05, 0.625, 0.75, 0.1 and ∞ and the excitatory and inhibitory connection widths were fixed to be $\sigma_e = \sigma_i = 2\alpha_{\text{ffwd}}$ for Figure 2 and 3. The inhibitory connection widths σ_i were 0.1, 0.2 and 0.3 for Figure 5 and 6, and the excitatory and the feedforward connection widths were $\sigma_e = 0.1$ and $\alpha_{\text{ffwd}} = 0.05$ respectively. The feedforward connection width

was $\alpha_{\text{ffwd}} = 0.05$ and the excitatory and inhibitory connection widths were $\sigma_e = \sigma_i = 0.1$ for Figure 1,7-9.

In the multi-layer network (Fig 9), the recurrent and the feedforward connections of all layers are $\alpha_{\text{rec}} = 0.1$ and $\alpha_{\text{ffwd}} = 0.05$, respectively. The first and the second layer are the same as the L4 and L2/3 networks, respectively, described above. The other layers are modeled the same as the L2/3 network. In the condition when $\mu_i = 0$ mV/ms, the feedforward connection strength from layer 1 to layer 2 excitatory neurons is $J_{\text{eF}} = 140$ mV, and the feedforward connection strengths between other layers are $J_{\text{eF}} = 35$ mV. In the condition when $\mu_i = 1.2$ mV/ms, the feedforward connection strength from layer 1 to layer 2 excitatory neurons is $J_{\text{eF}} = 180$ mV, and the feedforward connection strengths between other layers are $J_{\text{eF}} = 54$ mV. All the feedforward connections to inhibitory neurons and the static inputs to excitatory neurons were set to be zero, i.e. $J_{\text{iF}} = 0$ mV and $\mu_e = 0$ mV/ms. The feedforward strengths were chosen such that the mean firing rates of layer 2-4 are similar across layers as well as between the two conditions of μ_i .

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 0.05 ms.

Neural field model and stability analysis We use a two dimensional neural field model to describe the dynamics of population rate (Fig. 4 and 6). The neural field equations are

$$\tau_\alpha \frac{\partial r_\alpha(x, t)}{\partial t} = -r_\alpha + \phi(w_{\alpha e} * r_e + w_{\alpha i} * r_i + \mu_\alpha) \quad (5)$$

where $r_\alpha(x, t)$ is the firing rate of neurons in population $\alpha = e, i$ near spatial coordinates $x \in [0, 1] \times [0, 1]$. The symbol $*$ denotes convolution in space, μ_α is a constant external input and $w_{\alpha\beta}(x) = \bar{w}_{\alpha\beta} g(x; \sigma_\beta)$ where $g(x; \sigma_\beta)$ is a two-dimensional wrapped Gaussian with width parameter σ_β , $\beta = e, i$. The transfer function is a threshold-quadratic function, $\phi(x) = [x]_+^2$. The timescale of synaptic and firing rate responses are implicitly combined into τ_α .

For constant inputs, μ_e and μ_i , there exists a spatially uniform fixed point, r_α^* . Linearizing around this fixed point in Fourier domain gives a Jacobian matrix at each spatial Fourier mode⁴⁵

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

where $\vec{n} = (n_x, n_y)$ is the two-dimensional Fourier mode, $\tilde{w}_{\alpha\beta}(\vec{n}) = \bar{w}_{\alpha\beta} \exp(-2\|\vec{n}\|^2 \pi^2 \sigma_\beta^2)$ is the Fourier coefficient of $w_{\alpha\beta}(x)$ with $\|\vec{n}\|^2 = n_x^2 + n_y^2$ and $g_a = \phi'(w_{ae} * r_e^* + w_{ai} * r_i^* + \mu_a)$ is the gain. 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 Figures 5 and 7, μ_i varied from 0.1 to 0.7 with step size 0.002, σ_i varied from 0.05 to 0.2 with step size 0.0005, and $\mu_e = 0.5$ and $\sigma_e = 0.1$. The rest of the parameters were $\bar{w}_{ee} = 80$, $\bar{w}_{ei} = -72$, $\bar{w}_{ie} = 120$, $\bar{w}_{ii} = -90$, $\tau_e = 5$ ms and $\tau_i = 8$ ms.

Statistical methods Each simulation was 20 sec long consisting of alternating OFF (300 ms) and ON (200 ms) intervals. Image was presented during ON intervals, where the average firing rate of L4 neurons was $r_X = 10$ Hz. During OFF intervals, spike trains of L4 neurons were independent Poisson process with rate $r_X = 5$ Hz. Spike counts from the L2/3 excitatory neurons during the ON intervals were used to compute the linear Fisher information and noise correlations. The first spike count in each simulation was excluded. For each parameter condition, the connectivity matrices were fixed for all simulations. The initial states of each neuron's membrane potential were randomized in each simulation.

To compute tuning curve functions, the orientation for each ON interval was randomly sampled from 50 orientations uniformly spaced between 0 and 1. There were 9,750 spike counts in total for all orientations. Tuning curves were smoothed with a Gaussian kernel of width 0.05. Orientation

selectivity index for neuron i is computed as

$$\text{OSI}_i = \frac{\sqrt{\left(\sum_{k=1}^{N_{th}} r^i(\theta_k) \sin(2\pi\theta_k)\right)^2 + \left(\sum_{k=1}^{N_{th}} r^i(\theta_k) \cos(2\pi\theta_k)\right)^2}}{\sum_{k=1}^{N_{th}} r^i(\theta_k)} \quad (6)$$

where $r^i(\theta)$ is the tuning curve function of neuron i , and $\theta_k = k/N_{th}$.

To compute the linear Fisher information and noise correlation, the orientations of the Gabor images during ON intervals, were randomly chosen from $\theta_1 = \theta + \delta\theta/2$ and $\theta_2 = \theta - \delta\theta/2$, where $\theta = 0.5$ and $\delta\theta = 0.01$. The linear Fisher information of L2/3 neurons is computed using the bias-corrected estimate⁶⁴,

$$\hat{I}_{bc} = \frac{(\mathbf{f}_2 - \mathbf{f}_1)^T}{\delta\theta} \left(\frac{Q_1 + Q_2}{2} \right)^{-1} \frac{(\mathbf{f}_2 - \mathbf{f}_1)}{\delta\theta} \frac{2N_{tr} - N - 3}{2N_{tr} - 2} - \frac{2N}{N_{tr}\delta\theta^2} \quad (7)$$

where $\mathbf{f}_i$ and Q_i are the empirical mean and covariance, respectively, for θ_i . N_{tr} was the number of trials for each θ_i . The number of neurons were $N = 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800$, randomly sampled without replacement from the excitatory population of L2/3.

The linear Fisher information of L4 neurons can be estimated analytically, where

$$\begin{aligned} f'(\theta) &= T \cdot F \cdot I'(\theta), \\ \Sigma(\theta) &= T \cdot (FF^T) * \text{Var}(\xi^T) + T \cdot \text{diag}(F \cdot I(\theta)), \end{aligned}$$

with $\xi^T(t) = \int_t^{t+T} \xi(s)ds$ and T be the time window for spike counts. With $T = 200$ ms and $\tau_n = 40$ ms, $\text{Var}(\xi^T) \approx 0.81\sigma_n^2$.

The noise correlation was computed with $N = 1600$ neurons randomly sampled without replacement from the excitatory population of L2/3. The Pearson correlation coefficients were computed from the average covariance matrix $(Q_1 + Q_2)/2$.

For the computation of both noise correlations and linear Fisher information, there were 10 sampling of neurons for each N . Neurons of firing rates less than 1 Hz were excluded. There were 58,500 spike counts in total for θ_1 and θ_2 .

Code availability Computer code for all simulations and analysis of the resulting data will be available at <https://github.com/hcc11/>.

References

1. Shadlen, M. N. and Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience* **18**(10), 3870–3896 (1998).
2. Goris, R. L., Movshon, J. A., and Simoncelli, E. P. Partitioning neuronal variability. *Nature neuroscience* **17**(6), 858–865 (2014).
3. Renart, A. and Machens, C. K. Variability in neural activity and behavior. *Current opinion in neurobiology* **25**, 211–220 (2014).
4. Ruff, D. A., Ni, A. M., and Cohen, M. R. Cognition as a window into neuronal population space. *Annual review of neuroscience* **41**, 77–97 (2018).
5. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., and Harris, K. D. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**(6437), eaav7893 (2019).
6. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., and Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nature neuroscience* **22**(10), 1677–1686 (2019).
7. Cohen, M. and Kohn, A. Measuring and interpreting neuronal correlations. *Nature neuroscience* **14**(7), 811–819 (2011).

8. Luczak, A., Bartho, P., and Harris, K. D. Gating of sensory input by spontaneous cortical activity. *Journal of Neuroscience* **33**(4), 1684–1695 (2013).
9. Rabinowitz, N. C., Goris, R. L., Cohen, M., and Simoncelli, E. Attention stabilizes the shared gain of v4 populations. *eLife*, e08998 (2015).
10. Whiteway, M. R. and Butts, D. A. Revealing unobserved factors underlying cortical activity with a rectified latent variable model applied to neural population recordings. *Journal of neurophysiology* **117**(3), 919–936 (2016).
11. Engel, T. A. and Steinmetz, N. A. New perspectives on dimensionality and variability from large-scale cortical dynamics. *Current opinion in neurobiology* **58**, 181–190 (2019).
12. Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., and Tolias, A. S. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* **84**(2), 355–362 (2014).
13. Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J. A. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron* **86**(3), 740–754 (2015).
14. Downer, J. D., Niwa, M., and Sutter, M. L. Task engagement selectively modulates neural correlations in primary auditory cortex. *Journal of Neuroscience* **35**(19), 7565–7574 (2015).
15. Cohen, M. and Maunsell, J. Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience* **12**(12), 1594–1600 (2009).
16. Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area v4. *Neuron* **63**(6), 879–888 (2009).

17. Gu, Y., Liu, S., Fetsch, C. R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G. C., and Angelaki, D. E. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**(4), 750–761 (2011).
18. Ni, A., Ruff, D., Alberts, J., Symmonds, J., and Cohen, M. Learning and attention reveal a general relationship between population activity and behavior. *Science* **359**(6374), 463–465 (2018).
19. Smith, G. B., Sederberg, A., Elyada, Y. M., Van Hooser, S. D., Kaschube, M., and Fitzpatrick, D. The development of cortical circuits for motion discrimination. *Nature neuroscience* **18**(2), 252 (2015).
20. Zohary, E., Shadlen, M. N., and Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**(6485), 140 (1994).
21. Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. Population coding in neuronal systems with correlated noise. *Physical Review E* **64**(5), 051904 (2001).
22. Averbeck, B. B., Latham, P. E., and Pouget, A. Neural correlations, population coding and computation. *Nature reviews neuroscience* **7**(5), 358 (2006).
23. Abbott, L. F. and Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural computation* **11**(1), 91–101 (1999).
24. Wu, S., Amari, S.-i., and Nakahara, H. Population coding and decoding in a neural field: a computational study. *Neural Computation* **14**(5), 999–1026 (2002).
25. Wu, S., Amarib, S.-i., and Nakahara, H. Information processing in a neuron ensemble with the multiplicative correlation structure. *Neural Networks* **17**, 205–214 (2004).

26. Kohn, A., Coen-Cagli, R., Kanitscheider, I., and Pouget, A. Correlations and neuronal population information. *Annual review of neuroscience* **39**, 237–256 (2016).
27. Seung, H. S. and Sompolinsky, H. Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences* **90**(22), 10749–10753 (1993).
28. Brunel, N. and Nadal, J.-P. Mutual information, fisher information, and population coding. *Neural computation* **10**(7), 1731–1757 (1998).
29. Seriès, P., Latham, P. E., and Pouget, A. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nature neuroscience* **7**(10), 1129 (2004).
30. Beck, J., Bejjanki, V. R., and Pouget, A. Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. *Neural computation* **23**(6), 1484–1502 (2011).
31. Shamir, M. and Sompolinsky, H. Implications of neuronal diversity on population coding. *Neural computation* **18**(8), 1951–1986 (2006).
32. Josić, K., Shea-Brown, E., Doiron, B., and de la Rocha, J. Stimulus-dependent correlations and population codes. *Neural computation* **21**(10), 2774–2804 (2009).
33. Zylberberg, J., Cafaro, J., Turner, M. H., Shea-Brown, E., and Rieke, F. Direction-selective circuits shape noise to ensure a precise population code. *Neuron* **89**(2), 369–383 (2016).
34. Franke, F., Fiscella, M., Sevelev, M., Roska, B., Hierlemann, A., and da Silveira, R. A. Structures of neural correlation and how they favor coding. *Neuron* **89**(2), 409–422 (2016).
35. Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. Information-limiting correlations. *Nature neuroscience* **17**(10), 1410–1417 (2014).

36. Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G., and Josic, K. The mechanics of state dependent neural correlations. *Nature neuroscience* **19**(3), 383–393 (2016).
37. Ocker, G. K., Hu, Y., Buice, M. A., Doiron, B., Josić, K., Rosenbaum, R., and Shea-Brown, E. From the statistics of connectivity to the statistics of spike times in neuronal networks. *Current opinion in neurobiology* **46**, 109–119 (2017).
38. Kanashiro, T., Ocker, G. K., Cohen, M. R., and Doiron, B. Attentional modulation of neuronal variability in circuit models of cortex. *eLife* **6** (2017).
39. Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K. D., and de la Rocha, J. Stochastic transitions into silence cause noise correlations in cortical circuits. *Proceedings of the National Academy of Sciences* **112**(11), 3529–3534 (2015).
40. Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M., and Miller, K. D. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron* **98**(4), 846–860 (2018).
41. Kanitscheider, I., Coen-Cagli, R., and Pouget, A. Origin of information-limiting noise correlations. *Proc Natl Acad Sci* **112**(50), E6973–E6982 (2015).
42. van Vreeswijk, C. and Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
43. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience* **8**(3), 183–208 (2000).
44. Renart, A., de La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. The asynchronous state in cortical circuits. *Science* **327**(5965), 587–590 (2010).

45. Rosenbaum, R. and Doiron, B. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X* **4**(2), 021039 (2014).
46. Rosenbaum, R., Smith, M., Kohn, A., Rubin, J., and Doiron, B. The spatial structure of correlated neuronal variability. *Nature Neuroscience* **20**, 107–114 (2017).
47. Huang, C., Ruff, D. A., Pyle, R., Rosenbaum, R., Cohen, M. R., and Doiron, B. Circuit models of low-dimensional shared variability in cortical networks. *Neuron* **101**(2), 337–348 (2019).
48. Darshan, R., van Vreeswijk, C., and Hansel, D. Strength of correlations in strongly recurrent neuronal networks. *Physical Review X* **8**(3), 031072 (2018).
49. Landaau, I. D. and Sompolinsky, H. Coherent chaos in a recurrent neural network with structured connectivity. *PLoS computational biology* **14**(12), e1006309 (2018).
50. Mastrogiuseppe, F. and Ostojic, S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**(3), 609–623 (2018).
51. Pyle, R. and Rosenbaum, R. Spatiotemporal dynamics and reliable computations in recurrent spiking neural networks. *Physical Review Letters* **118**(1), 018103 (2017).
52. Baker, C., Ebsch, C., Lampl, I., and Rosenbaum, R. Correlated states in balanced neuronal networks. *Physical Review E* **99**(5), 052414 (2019).
53. Kaschube, M., Schnabel, M., Löwel, S., Coppola, D. M., White, L. E., and Wolf, F. Universality in the evolution of orientation columns in the visual cortex. *science* **330**(6007), 1113–1116 (2010).
54. Kass, R. E., Amari, S.-I., Arai, K., Brown, E. N., Diekman, C. O., Diesmann, M., Doiron, B., Eden, U. T., Fairhall, A. L., Fiddymment, G. M., et al. Computational neuroscience: Mathemat-

ical and statistical perspectives. *Annual Review of Statistics and Its Application* **5**(1), 183–214 (2018).

55. Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. On the distribution of firing rates in networks of cortical neurons. *The Journal of neuroscience* **31**(45), 16217–16226 (2011).

56. Ringach, D. L., Shapley, R. M., and Hawken, M. J. Orientation selectivity in macaque v1: diversity and laminar dependence. *Journal of Neuroscience* **22**(13), 5639–5651 (2002).

57. Monteforte, M. and Wolf, F. Dynamic flux tubes form reservoirs of stability in neuronal circuits. *Physical Review X* **2**(4), 041007 (2012).

58. Lin, I.-C., Okun, M., Carandini, M., and Harris, K. D. The nature of shared cortical variability. *Neuron* **87**(3), 644–656 (2015).

59. Smith, M. A. and Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience* **28**(48), 12591–12603 (2008).

60. de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. Correlation between neural spike trains increases with firing rate. *Nature* **448**(7155), 802 (2007).

61. Shea-Brown, E., Josić, K., de La Rocha, J., and Doiron, B. Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Physical review letters* **100**(10), 108102 (2008).

62. Doshier, B. A. and Lu, Z.-L. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences* **95**(23), 13988–13993 (1998).

63. Kay, S. M. *Fundamentals of statistical signal processing*. Prentice Hall PTR, (1993).

64. Kanitscheider, I., Coen-Cagli, R., Kohn, A., and Pouget, A. Measuring fisher information accurately in correlated neural populations. *PLoS computational biology* **11**(6), e1004218 (2015).
65. Doshier, B. A. and Lu, Z.-L. Mechanisms of perceptual learning. *Vision research* **39**(19), 3197–3221 (1999).
66. 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. State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**(1), 235–248 (2014).
67. Ruff, D. A., Xue, C., Kramer, L. E., Baqai, F., and Cohen, M. R. Low rank mechanisms underlying flexible visual representations. *bioRxiv* , 730978 (2019).
68. Williamson, R. C., Cowley, B. R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M. A., and Byron, M. Y. Scaling properties of dimensionality reduction for neural populations and network models. *PLOS Computational Biology* **12**(12), e1005141 (2016).
69. Muller, L., Chavane, F., Reynolds, J., and Sejnowski, T. J. Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience* **19**(5), 255 (2018).
70. Cross, M. C. and Hohenberg, P. C. Pattern formation outside of equilibrium. *Reviews of modern physics* **65**(3), 851 (1993).
71. Ermentrout, B. Neural networks as spatio-temporal pattern-forming systems. *Reports on progress in physics* **61**(4), 353 (1998).
72. Bressloff, P. C. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical* **45**(3), 033001 (2011).
73. Ledoux, E. and Brunel, N. Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. *Frontiers in computational neuroscience* **5**, 25 (2011).

74. Schaffer, E. S., Ostojic, S., and Abbott, L. F. A complex-valued firing-rate model that approximates the dynamics of spiking networks. *PLoS computational biology* **9**(10), e1003301 (2013).
75. Coombes, S. Waves, bumps, and patterns in neural field theories. *Biological cybernetics* **93**(2), 91–108 (2005).
76. Keane, A. and Gong, P. Propagating waves can explain irregular neural dynamics. *Journal of Neuroscience* **35**(4), 1591–1605 (2015).
77. Kriener, B., Helias, M., Rotter, S., Diesmann, M., and Einevoll, G. T. How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime. *Frontiers in computational neuroscience* **7**, 187 (2014).
78. Cover, T. M. and Thomas, J. A. *Elements of information theory*. John Wiley & Sons, (2012).
79. Reyes, A. D. Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro. *Nature neuroscience* **6**(6), 593 (2003).
80. Kumar, A., Rotter, S., and Aertsen, A. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nature reviews neuroscience* **11**(9), 615 (2010).
81. Ecker, A. S., Denfield, G. H., Bethge, M., and Tolias, A. S. On the structure of neuronal population activity under fluctuations in attentional state. *Journal of Neuroscience* **36**(5), 1775–1789 (2016).
82. Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences* **92**(9), 3844–3848 (1995).
83. Ferster, D. and Miller, K. D. Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience* **23**(1), 441–471 (2000).

84. Lindner, B., Doiron, B., and Longtin, A. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Physical Review E* **72**(6), 061919 (2005).
85. Ginzburg, I. and Sompolinsky, H. Theory of correlations in stochastic neural networks. *Physical review E* **50**(4), 3171 (1994).
86. Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex* **10**(9), 910–923 (2000).
87. Litwin-Kumar, A. and Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience* **15**(11), 1498–1505 (2012).
88. Burak, Y. and Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS computational biology* **5**(2), e1000291 (2009).
89. Mariño, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., and Sur, M. Invariant computations in local cortical networks with balanced excitation and inhibition. *Nature neuroscience* **8**(2), 194 (2005).
90. Rossi, L. F., Harris, K., and Carandini, M. Excitatory and inhibitory intracortical circuits for orientation and direction selectivity. *bioRxiv* , 556795 (2019).
91. Bejjanki, V. R., Beck, J. M., Lu, Z.-L., and Pouget, A. Perceptual learning as improved probabilistic inference in early sensory areas. *Nature neuroscience* **14**(5), 642 (2011).
92. Zylberberg, J., Pouget, A., Latham, P. E., and Shea-Brown, E. Robust information propagation through noisy neural circuits. *PLoS computational biology* **13**(4), e1005497 (2017).
93. Renart, A. and van Rossum, M. C. Transmission of population-coded information. *Neural computation* **24**(2), 391–407 (2012).

94. Snyder, A. C., Byron, M. Y., and Smith, M. A. Distinct population codes for attention in the absence and presence of visual stimulation. *Nature communications* **9** (2018).
95. Verhoef, B.-E. and Maunsell, J. H. Attention-related changes in correlated neuronal activity arise from normalization mechanisms. *Nature neuroscience* **20**(7), 969 (2017).
96. Scholvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., and Carandini, M. Cortical state determines global variability and correlations in visual cortex. *Journal of Neuroscience* **35**(1), 170–178 (2015).
97. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**(6961), 954 (2003).
98. Luczak, A., Barthó, P., and Harris, K. D. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**(3), 413–425 (2009).
99. Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331**(6013), 83–87 (2011).
100. Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* **92**(2), 530–543 (2016).
101. Goldberg, J. A., Rokni, U., and Sompolinsky, H. Patterns of ongoing activity and the functional architecture of the primary visual cortex. *Neuron* **42**(3), 489–500 (2004).

Acknowledgments

The work was funded by the Swartz Foundation Fellowship #2017-7 (C.H.), grants from the Simons Foundation Collaboration on the Global Brain (C.H., A.P and B.D.), the Swiss National Science Foundation (www.snf.ch), #31003A_143707 and #31003A_165831 (A.P.), National Institutes of

Health (Grants #1U19NS107613-01 and #R01 EB026953), and the Vannevar Bush Faculty (Fellowship #N00014-18-1-2002). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC).

Author Contributions

C.H., A.P., and B.D. conceived the project; C.H. performed the simulations and data analysis; B.D. supervised the project; all authors contributed to writing the manuscript.

Author Information

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.D. (bdoiron@pitt.edu).