Hierarchical Neural Circuit Theory of Normalization and Inter-areal Communication

Asit Pal[1,2,*] Shivang Rawat[2,3], David J. Heeger[4,5,†] and Stefano Martiniani[1,2,3,5,‡]

[1]Simons Center for Computational Physical Chemistry, Dept. of Chemistry, New York University, NY, USA
[2]Center for Soft Matter Research, Dept. of Physics, New York University, NY, USA
[3]Courant Institute of Mathematical Sciences, New York University, NY, USA
[4]Dept. Of Psychology, New York University, NY, USA
[5]Center for Neural Science, New York University, NY, USA

Highlights

- Hierarchical neural circuit with reciprocal connections implementing divisive normalization.
- Unified theory of inter-areal communication.
- Emergent properties of the theory explain key experimental findings.
- Feedback from higher cortical areas dictates functional connectivity.

In brief

A hierarchical recurrent neural circuit theory with reciprocal connections, that dynamically implements divisive normalization, predicts key experimental findings. The theory offers a unified view of inter-areal communication in which both communication through subspace and communication through coherence are emergent properties, predicting that high-coherence frequencies exhibit stronger inter-areal communication and reduced subspace dimensionality.

Summary

The primate brain exhibits a hierarchical, modular architecture with conserved microcircuits executing canonical computations across reciprocally connected cortical areas. Though feedback connections are ubiquitous, their functions remain largely unknown. To investigate the role of feedback, we present a hierarchical neural circuit theory with feedback connections that dynamically implements divisive normalization across its hierarchy. In a two-stage instantiation (V1 $\leftrightarrow$ V2), increasing feedback from V2 to V1 amplifies responses in both areas, more so in the higher cortical area, consistent with experiments. We analytically derive power spectra (V1) and coherence spectra (V1-V2), and validate them against experimental observations: peaks in both spectra shift to higher frequencies with increased stimulus contrast, and power decays as $1/f^4$ at high frequencies (f). The theory further predicts distinctive spectral signatures of feedback and input gain modulation. Crucially, the theory offers a unified view of inter-areal communication, with emergent features commensurate with empirical observations of both communication subspaces and inter-areal coherence. It admits a low-dimensional communication subspace, where inter-areal communication is lower-dimensional than within-area communication, and frequency bands characterized by high inter-areal coherence. It further predicts that: i) increasing feedback strength enhances inter-areal communication and diminishes within-area communication, without altering the subspace dimensionality; ii) high-coherence frequencies are characterized by stronger communication (ability to estimate neural activity in one brain area from neural activity in another brain area) and reduced subspace dimensionality. Finally, a three-area (V1 $\leftrightarrow$ V4 and V1 $\leftrightarrow$ V5) instantiation of the theory demonstrates that differential feedback from higher to lower cortical areas dictates their dynamic functional connectivity. Altogether, our theory provides a robust and analytically tractable framework for generating experimentally-testable predictions about normalization, inter-areal communication, and functional connectivity.

*asit.pal@nyu.edu
†david.heeger@nyu.edu
‡stefano.martiniani@nyu.edu

Introduction

The primate brain is a complex hierarchical network of interconnected cortical areas that perform computations essential for sensorimotor processing and cognitive functions. Among these, certain computations are repeated across cortical areas and are thus said to be "canonical computations". Divisive normalization [1, 2] is one such fundamental canonical principle.

Feedback connections are ubiquitous in the primate brain [3–5], yet their precise roles in neural computation and communication remain largely unknown [6–8]. They have been hypothesized to serve various functions, including attentional modulation [9–11], inter-areal communication [12], learning [13, 14], and generative modeling [15]. Although divisive normalization has been extensively studied in feedforward recurrent networks [16–19], its implementation and effects in hierarchical networks with feedback connections remain unexplored.

Effective communication between brain areas is essential for complex cognitive tasks, yet the precise mechanisms of communication remain an active area of investigation. One prominent hypothesis, "communication through coherence" (CTC) [20, 21], suggests that synchronized oscillations (coherence) across brain regions during diverse cognitive, perceptual, and behavioral tasks are crucial for effective inter-areal communication [22–26]. However, the causal nature of the relationship remains a key unresolved question. It is currently debated whether coherence is a fundamental mechanism enabling inter-areal communication or rather an epiphenomenon arising from it [27–31]. Another prominent alternate hypothesis proposes that brain areas communicate through a low-dimensional "communication subspace (CS)" embedded within the high-dimensional neural activity [32–35]. This concept, also identified in artificial neural networks [36], posits that only activity within these shared subspaces between the source and target areas facilitates communication. Currently, a comprehensive theoretical framework capable of simultaneously exploring these different communication mechanisms and explaining their dynamic modulation by factors such as feedback and stimulus contrast is lacking.

To address these unresolved questions, we introduce a novel hierarchical recurrent neural circuit theory (Fig. 1) built on two *a priori* constraints: i) a hierarchy of brain areas with recurrent, feedforward, and feedback connections, and ii) recurrently implemented normalization within each cortical area. Based on the theory, we derive analytical expressions to compute inter-areal coherence and communication subspaces, offering experimentally testable predictions about the effects of both input gain and feedback strength on inter-areal communication. As a consequence of these two *a priori* constraints, our theoretical framework gives rise to emergent properties that explain a wide spectrum of neurophysiological phenomena, including modulation of contrast response functions, oscillatory dynamics evident in both power and coherence spectra, communication subspaces, and the dynamic control of functional connectivity. A central contribution of this work is to present a unified view of inter-areal communication. The theory simultaneously predicts empirical observations of both communication subspaces and inter-areal coherence, and reveals that coherence shapes the communication subspace.

Note that throughout our analysis, we neither adjust model parameters (such as time constants, weight matrices, and noise amplitudes) to reproduce specific experimental trends nor fit data when making direct comparisons. All results presented were obtained using a *single* set of model parameters.

Results

Theory

Initially conceived to explain the responses in the primary visual cortex (V1) [1], the normalization model has emerged as a canonical computation across sensory and cognitive domains [2], including higher level vision [37, 38], olfaction [39], audition [40, 41], somatosensation [42], attention [43–55], multisensory integration [56–60], working memory [61–64], and decision-making [65–68]. At its core, the normalization model proposes that the response of an individual neuron is scaled by the collective activity of its neighboring neurons (viz., the normalization pool), analogous to normalizing the length of a vector. Functionally, this computation serves as an automatic gain control mechanism, enabling neurons to operate effectively over a broad dynamic range of inputs [1, 69, 70]. It also simplifies the interpretation of neural signals by downstream brain regions [70–72] and maintains response consistency across different input conditions [1, 39, 69, 70, 73]. In addition, normalization decorrelates population responses to reduce redundancy [39, 74–77], prevents excessive recurrent excitation to maintain network stability, [18, 19], and fosters efficient learning [78, 79].

Normalization is a computational-level description of neural responses that can be implemented using various mechanisms. Experimental evidence suggests that normalization operates via recurrent amplification [80–83], amplifying weak inputs more than

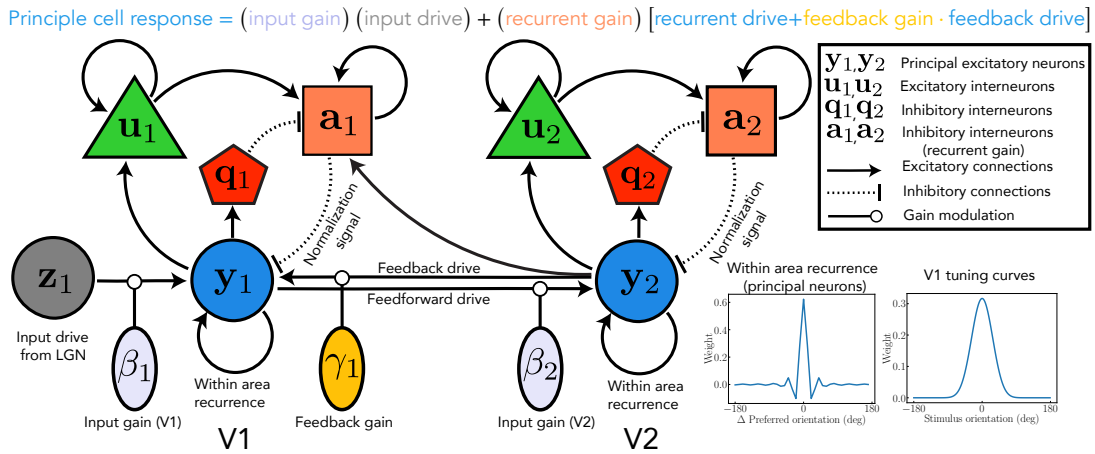

Figure 1: ***Schematic of the two-stage recurrent circuit model.*** *Solid and dashed lines represent excitatory and inhibitory connections, respectively. Principal excitatory neurons are denoted by* $\mathbf{y}$*, modulatory excitatory neurons by* $\mathbf{u}$*, and modulatory inhibitory neurons by* $\mathbf{a}$ *and* $\mathbf{q}$*. Subscripts 1 and 2 designate neurons in visual areas V1 and V2, respectively.* $\mathbf{z}_1$ *denotes the input drive from the LGN to V1, a weighted sum of the responses of LGN neurons. Parameters* β_1 *and* γ_1 *modulate the input and feedback gain to V1, respectively.*

strong inputs, consonant with evidence that cortical circuits are dominated by recurrent excitation [84–89]. ORGaNICS is one such theoretical framework that dynamically implements divisive normalization by modulating recurrent amplification [16, 17]. Previous work on normalization models (including previous work on ORGaNICs) has focused on a single neural circuit within a brain region, such as V1, ignoring the abundant top-down feedback connections. Here, we develop a new hierarchical ORGaNICs theory with feedback in addition to feedforward and recurrent connections.

The hierarchical ORGaNICs theory describes neuronal population responses as dynamic processes evolving in a recurrent circuit, modeled by nonlinear differential equations (see Fig. 1 for illustration and Methods for equations). The circuit comprises distinct cell types: principal (excitatory), modulator (excitatory), and modulator (inhibitory) cells (n.b., "modulator" refers to a multiplicative computation, not necessarily implemented with neuromodulators). Output firing rates of principal cells depend on three terms: 1) feedforward input drive modulated (viz., scaled) by input gain; 2) feedback drive modulated by feedback gain; 3) the sum of recurrent drive and scaled feedback drive, collectively modulated by recurrent gain. Unlike generic E-I models in which all neurons perform the same computation, ORGaNICs assign unique computations to each cell type. Because normalization ensures circuit stability over a broad range of parameters [18, 19], the theory is uniquely well suited to analytical treatment, e.g., to derive closed-form solutions for the power spectral density, coherence, and communication subspaces (see Methods and Appendix).

Each neuron's input drive is a weighted sum of responses from a population of neurons at lower stages of the hierarchy. For example, LGN responses provide input drive to V1 ($\mathbf{z_1}$ in Fig.1), and V1 responses similarly drive V2 (solid arrow from $\mathbf{y_1}$ to $\mathbf{y_2}$ in Fig.1, representing principal neuron responses in each area). This weighting is defined by an embedding weight matrix composed of positive synaptic weights, which specify each neuron's stimulus preferences, such as orientation selectivity (Fig.1, V1 tuning curves), analogous to a standard receptive field model. Input gain is determined independently by input modulator cells (β_1 and β_2 in Fig.1).

The feedback drive is a weighted sum of the responses of a population of neurons in one or more brain areas from a higher stage in the hierarchy, e.g., V2 provides feedback to V1 (Fig. 1, solid arrow from $\mathbf{y_2}$ to $\mathbf{y_1}$). The feedback drive is specified by a feedback weight matrix. For simplicity, the feedback weights from V2 to V1 were chosen to be the transpose of the feedforward weight matrix from V1 to V2 [15]. For the exact structure of the different matrices, see Fig. 7. The feedback gain from V2 to V1 is denoted by γ_1 in Fig. 1.

The recurrent drive is a weighted sum of principal cell responses (Fig. 1, curved arrows from $\mathbf{y_1}$ onto itself and from $\mathbf{y_2}$ onto itself). For the results reported in this paper, the recurrent weights were selected to have a center-surround architecture in which the closest recurrent connections are excitatory and the more distant ones are inhibitory (Fig. 1, recurrent synaptic weights). The recurrent gain depends on the responses of recurrent modulator cells ($\mathbf{a}_1$ and $\mathbf{a}_2$ in Fig. 1), which in turn depend on: 1) the principal cell responses ($\mathbf{y_1}$ and $\mathbf{y_2}$ in Fig. 1) via excitatory

($\mathbf{u}_1$ and $\mathbf{u}_2$ in Fig. 1) and 2) inhibitory ($\mathbf{q}_1$ and $\mathbf{q}_2$ in Fig. 1) interneurons; 3) feedback from principal cells at a higher stage (solid arrow from $\mathbf{y_2}$ to $\mathbf{a}_1$ in Fig. 1). Normalization is achieved dynamically through the action of the modulator cells, which carry the "normalization signal".

Contrast response functions

The two-stage recurrent model with feedback implements divisive normalization exactly, across multiple cortical areas simultaneously (i.e., the model's steady-state responses align precisely with the analytical steady-state solutions given by the normalization equation Eq. 2 when all the gains are set to $\gamma = \beta = 1$ and the recurrent weight matrix is identity, see Methods and Appendix). As shown in Fig. 2a, both V1 (dashed) and V2 (solid) exhibit the characteristic saturating, sigmoidal dependence on contrast. The model predicts a lower semisaturation constant (the contrast at which the firing rate reaches half-saturation) and a steeper slope for cortical area V2 compared to V1. This prediction is commensurate with experimental observations in the literature that used optimal stimulus sizes (Fig. 2b) [90, 91]. Similar hierarchical trends, specifically an increase in slope and a decrease in the semisaturation constant, have also been observed across other brain areas like the LGN, V1, and MT with progression up the visual hierarchy [92].

Increasing the feedback gain, γ_1, enhances neural responses across cortical areas (Fig. 3a). However, this enhancement is not uniform across cortical areas. The gain change is notably higher for neurons in the higher cortical area (V2). Elevating input gain, β_1, also enhances neural responses across both cortical areas (Fig. 3b). However, unlike the effects observed with feedback gain modulation, the response enhancement from changing input gain is nearly identical in both cortical areas.

Power Spectra

Power spectra analysis of electroencephalographic (EEG) and local field potential (LFP) signals has been used extensively to identify and quantify oscillatory activity in various frequency bands (e.g., alpha, beta, gamma, theta), hypothesized to be linked to various cognitive functions, behavioral states, and neurological conditions [93–96]. Given that our model has a fixed point solution and is analytically tractable, we can calculate the power and coherence spectra analytically, given a model of the noise in the circuit (see Methods).

We adopted an analytically tractable noise model that adds low-pass filtered synaptic noise to the membrane potential and multiplicative Gaussian white noise to the firing rate dynamics. Furthermore, we explicitly accounted for long-range correlated noise in LFP measurements. This type of noise, propagating through brain tissue, affects LFPs with negligible impact on spiking activity and firing rates [97–99]. Further details are provided in the Appendix ("Intrinsic neural circuit noise" and "Extrinsic noise in the local field potential").

Our theory successfully replicates key experimental observations of V1 power spectra across varying contrast levels. These theoretical V1 predictions (Fig. 2c) are compared with experimental spectra from macaque V1 (Fig. 2d) under identical contrast conditions [22]. The model accurately captures three critical experimental trends: 1) the power spectrum of the noise is shaped by recurrent normalization to exhibit a resonant peak in the gamma-frequency range ($\approx$ 40-60 Hz); 2) the peak frequency shifts rightward with increasing contrast; and 3) the peak power decreases after reaching a maximum around 40% contrast [100–102]. The oscillatory activity in the gamma-frequency range depends on normalization; removing normalization eliminates the peaks (see Suppl. Fig. 11e). The peak frequencies depend on the values of the intrinsic time constants (see Table 1); larger time constants would shift the peaks to lower frequencies (see Discussion).

The theory also reproduces the steep fall-off in power reported at higher frequencies [103], where power approximately scales as $1/f^4$ with increasing frequency (f) (Fig. 2c, inset). This occurs because the synaptic noise added to the membrane potentials is first low-pass filtered by the synapse's own time constant, and then by the network model itself that acts as a low-pass filter (each step contributing a factor of $1/f^2$).

Modulation of feedback and input gains elicits distinct changes in the power spectra for a constant stimulus contrast. Figures 3c-f illustrate these distinct effects on the normalized power spectra when feedback gain (left panels) or input gain (right panels) is varied. These comparisons are presented for both a low (3%; Figs. 3c,d) and a high (50%; Figs. 3e,f) stimulus contrast. See Supplementary Material for comprehensive results across other contrast levels (Suppl. Fig. 8).

The model predicts oscillatory activity in the alpha-frequency range ($\approx$ 10 Hz) for low stimulus contrasts and small feedback gain values (Fig. 3*c*). As the feedback gain to the excitatory neurons increases, this alpha peak diminishes and ultimately vanishes at high feedback gain levels, while a broad peak concurrently emerges in the gamma frequency range (Figs. 3c,e). No such oscillatory activity in the alpha-frequency range is observed by manipulating the

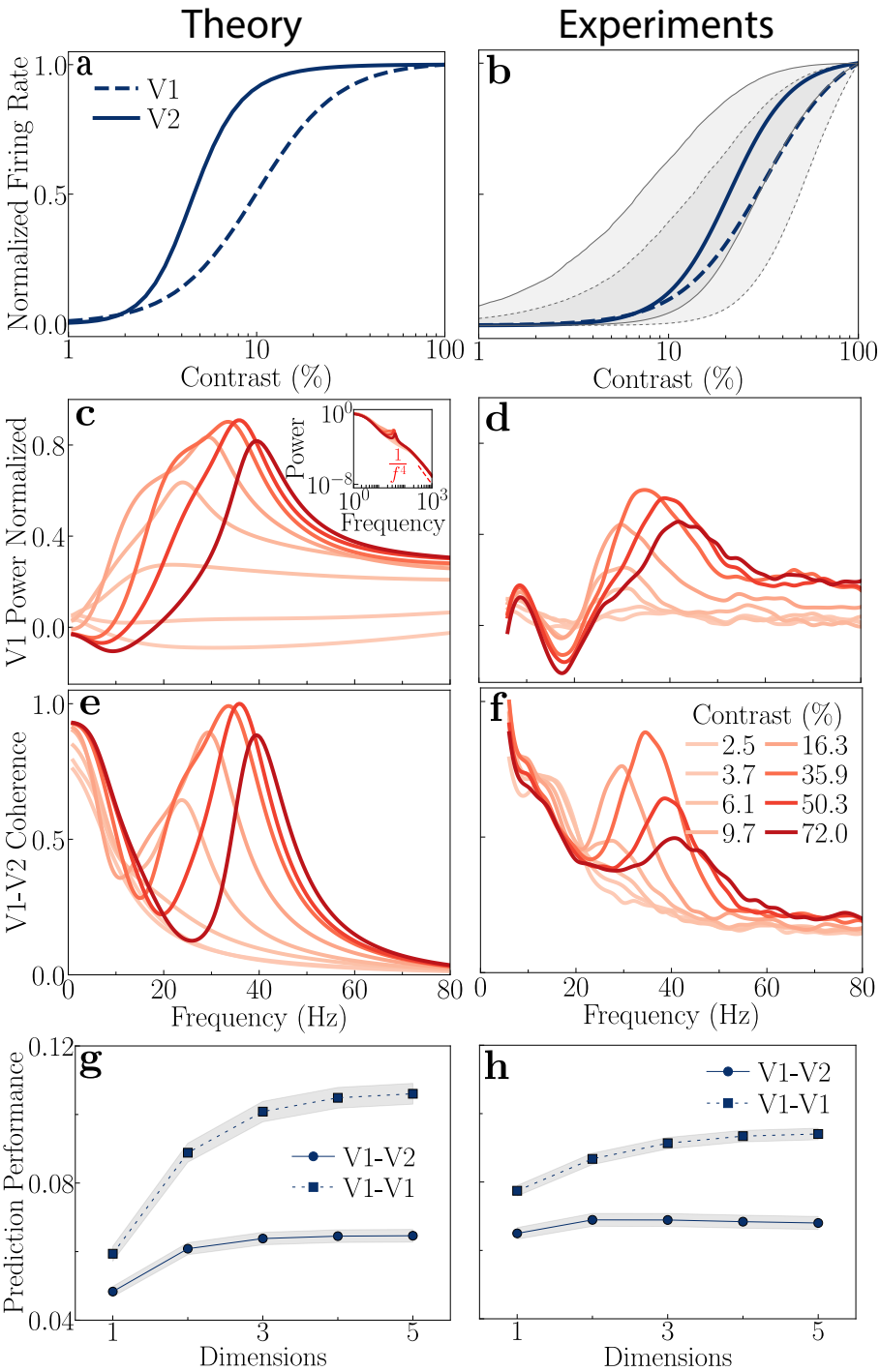

Figure 2: ***Comparison of theoretical predictions (left column) and experimental observations (right column) in V1 and V2.*** *All theoretical predictions are generated using the same baseline parameters (Table 1).* ***a,*** *Theoretical mean firing rates as a function of stimulus contrast (V1: dashed line, V2: solid line).* ***b,*** *Experimental mean firing rates of V1 and V2 (data from [90] and [91], replotted for comparison). The shaded area with a solid border indicates the 25th to 75th percentile range for V1, and the one with the dashed border indicates the same for V2.* ***c,*** *Theoretical V1 power spectra at various stimulus contrast levels. Power spectra were normalized using the equation* $\frac{(power - baseline)}{(power + baseline)}$*, where baseline power is taken to be spontaneous activity at 0% contrast. The peak power shifts towards higher frequency with increasing contrast. The inset shows the* $\frac{1}{f^4}$ *power law decay at high frequency.* ***d,*** *Experimental power spectra from macaque V1 (data from [22], replotted for comparison).* ***e,*** *Theoretical coherence spectra between maximally firing neurons in V1 and V2. The peak coherence shifts towards higher frequency with increasing contrast.* ***f,*** *Experimental coherence spectra from macaque V1-V2 (data from [22], replotted for comparison).* ***g,*** *Theoretical communication subspaces, prediction performance as a function of dimensionality for inter-area (circles) and within-area (squares) communication. Simulation results indicate that the inter-area communication subspace is lower dimensional than the within-area communication.* ***h,*** *Experimental communication subspaces from macaque V1-V2 (data from [32], replotted for comparison). The plotted prediction performance in both panels* ***g*** *and* ***h*** *is an average across different subsets of the source and target neural populations, and the shaded areas represent the standard error of the mean (SEM).*

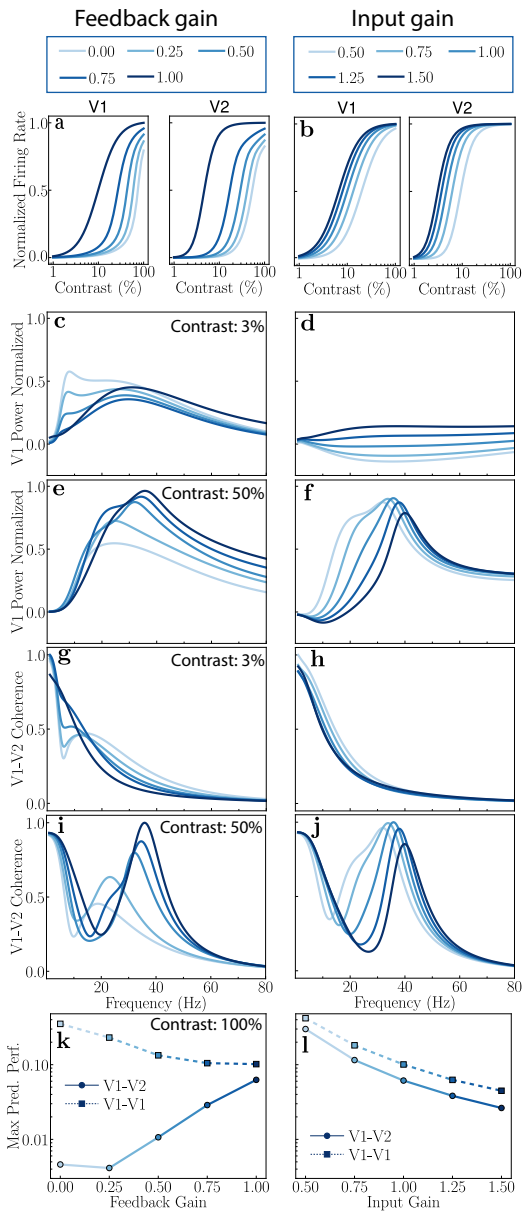

Figure 3: ***Theoretical predictions: Modulating feedback (left) and input gain (right). a,b,*** *Firing rates as a function of contrast. Increasing either feedback gain or input gain enhances neural responses, with feedback gain showing a greater impact in the higher cortical area V2.* ***c,d,*** *V1 power spectra: 3% contrast. An alpha peak is observed for low feedback gain, but the peak diminishes with increasing feedback gain. The alpha peak is absent with input gain changes.* ***e,f,*** *V1 power spectra: 50% contrast. A consistent gamma peak is observed, which shifts toward higher frequencies with increasing feedback gain (**e**) and input gain (**f**).* ***g,h,*** *Coherence spectra: 3% contrast. A broad peak in the beta band is observed at low feedback gain, which vanishes at high feedback gain. No such peak is observed with changes in input gain.* ***i,j,*** *Coherence spectra: 50% contrast. A beta peak is observed for low contrasts, which shifts toward higher (gamma) frequencies with increasing feedback gain. No beta peak is observed for changes in input gain, but the gamma peak shifts toward higher frequencies with increasing input gain.* ***k,l,*** *Communication subspaces. Increasing feedback gain enhances inter-areal (circles) communication while decreasing within-area (squares) communication. Conversely, increasing input gain decreases both inter- and within-area communication.*

input gain, for any contrast (Figs. 3d,f).

The power spectra are dominated by oscillatory activity in the gamma-frequency range at high stimulus contrast (Figs. 3e,f). Increasing the feedback gain causes this gamma peak to shift towards higher frequencies. A similar shift of the gamma peak to higher frequencies is observed when the input gain is increased. However, a key distinction arises in their impact on the peak's bandwidth: the gamma peak bandwidth narrows with increasing input gain but broadens with increasing feedback gain.

Coherence

Coherence measures the statistical correlation between two signals (e.g., LFPs in two different brain areas) as a function of frequency. Neuronal synchrony is hypothesized to contribute to the dynamic selection and routing of information both within and between brain areas [21, 104, 105].

Our model's predictions for V1-V2 coherence spectra accurately capture key experimental trends observed across various stimulus contrasts. We analytically determined the coherence between the maximally firing neurons of V1 and V2 (see Methods). Figure 2e displays the theoretically predicted V1-V2 coherence spectra for different contrasts. These model predictions are compared with experimental V1-V2 coherence data from macaque monkeys, obtained under identical contrast conditions (Fig. 2f; [22]). The model accurately captures several crucial experimental observations: 1) high coherence at low frequencies; 2) a shift of the peak coherence toward higher frequencies with increasing contrast; and 3) a decrease in peak coherence after reaching a maximum around 40% contrast.

Feedback and input gain differentially modulate V1-V2 coherence across contrasts (Fig. 3g-j). At low contrast (3%), a coherence peak in the beta band (12-30 Hz) is observed with low feedback gain (Fig. 3g); this peak disappears as feedback gain is increased. Notably, at low feedback gain, a dip in V1-V2 coherence occurs in the alpha band, corresponding to the frequencies where peak power is observed in V1. This suggests that alpha band activity, despite its prominence in V1 power, may not effectively propagate to V2 or participate in inter-areal communication. No such peaks are observed with changes in the input gain for low contrast (Fig. 3h). At high contrast (50%), increasing feedback gain shifts the peak coherence towards higher frequencies and increases both its magnitude and width (Fig. 3i). A similar frequency shift in peak coherence occurs with increasing input gain, but its magnitude and width decrease (Fig. 3j).

Communication subspaces

Neural communication is hypothesized to be channeled through "communication subspaces (CS)" [32–35]. Although neural activity space is high-dimensional, neural communication may not utilize this vast space. CS are lower-dimensional manifolds within the broader neural activity space that preferentially engage in inter-areal communication. In contrast, within-area communication utilizes a distinct, higher-dimensional "private" subspace [32].

Our theory successfully captures key characteristics of the CS (Figs. 2g,h). We derived an analytical model to examine both inter-areal (V1-V2) and within-area (V1-V1) communication (see Methods). We found that the CS is low-dimensional, much lower than the number of neurons (30 neurons for the analysis) in each brain area. In addition, the inter-areal communication (V1-V2) subspace is lower-dimensional than the within-area (V1-V1) communication subspace. Figure 2g illustrates the model prediction performance (i.e., the ability to predict activity fluctuations of a target subpopulation from the activity fluctuations of a source subpopulation using a linear model) as a function of the dimensionality of the CS. These theoretical predictions align well with experimental data (Fig. 2h; [32]).

Feedback and input gain differentially modulate communication subspaces (Fig. 3k,l). Increasing feedback gain from the higher cortical area (V2) to the lower cortical area (V1) enhanced inter-areal communication while diminishing within-area communication (Fig. 3k), without substantially altering the subspace dimensionality (see Supplementary Material, Fig. 10a,b). Conversely, increasing input gain reduced both the inter-area and within-area communication (Fig. 3l). These results are for 100% contrast; see Supplementary Material (Fig. 10c) for other contrasts. The observed enhancement of inter-areal communication with increased feedback suggests a key role for top-down feedback, to dynamically modulate the efficiency of information transfer between cortical areas without altering the underlying structure of the communication subspace (see below, Dynamic modulation of functional connectivity through feedback), a finding consistent with previous work on subspace dimensionality [33].

Frequency decomposition of communication subspaces

The theory predicts that the effectiveness of communication between brain areas varies systematically with frequency and contrast (Fig. 4). We quantified the efficacy of communication by analytically calculating the prediction performance for each fre-

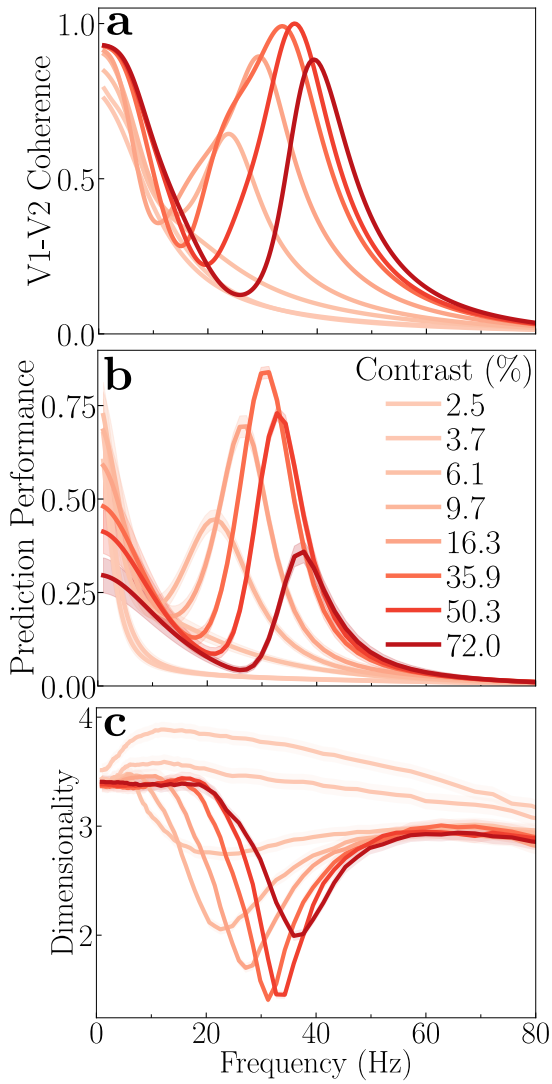

Figure 4: ***Theoretical prediction: Frequency-dependence of communication.*** ***a,*** *V1-V2 coherence spectra for different contrasts; same as in Fig. 2e.* ***b,*** *Prediction performance versus frequency at different contrasts, averaged across different subsets of the source and target populations. At low stimulus contrast (≈ 10 %), a peak in communication efficacy is observed around 20 Hz. As stimulus contrast increases, the preferred frequency of communication shifts towards higher frequencies (40 Hz). The magnitude of communication reaches a maximum around 40% contrast, before declining at higher contrast levels. Under conditions of very low contrast (< 10%), communication is mostly concentrated at very low frequencies. These frequency-specific trends in communication parallel the V1-V2 coherence patterns shown in panel* ***a****.* ***c,*** *Dimensionality of the communication subspace versus frequency, averaged across different subsets of the source and target populations. A notable dip in dimensionality occurs at frequencies corresponding to peaks in communication (panel* ***b****) and coherence (panel* ***a****). The shaded areas in panels* ***b*** *and* ***c*** *represent the standard error of the mean (SEM).*

quency component (see Methods). To achieve this, we transformed the noise correlations of the source and target populations into power spectra using the Wiener-Khinchin theorem. This conversion enabled us to predict the power of the target population from the source population (using a linear readout) at each frequency, thus providing a measure of communication efficacy (prediction performance) as a function of frequency.

V1-V2 communication is dynamically modulated by stimulus contrast in a frequency-dependent manner. At low stimulus contrasts (e.g, approximately 10%), a distinct peak in communication is observed at $\approx$ 20 Hz (Fig. 4b). As stimulus contrast increased, this preferred frequency for communication shifts toward higher frequencies (Fig. 4b). Furthermore, the magnitude of this communication peak exhibits a non-monotonic relationship with contrast, reaching its maximum at around 40% contrast before declining at higher contrasts. This particular effect is due to the normalization; removing normalization eliminates the peaks in both coherence and prediction performance (see Suppl. Fig. 11f-g). Under conditions of very low contrast, no distinct peak in communication is observed; instead, communication is mostly concentrated at very low frequencies (<10 Hz). Notably, these observed frequency-specific trends in communication bear a striking resemblance to the V1-V2 coherence patterns (Fig. 4a). This parallelism indicates a relationship between the degree of synchronization of neural activity between brain areas and the effectiveness of communication between these brain regions, a prediction that can be tested experimentally (see Discussion).

The dimensionality of the inter-areal (V1-V2) CS also varies with frequency and stimulus contrast (Fig. 4c). The dimensionality of the communication subspace decreases at frequencies where inter-areal coherence and communication are maximal, for each contrast (Fig. 4c).

The dimensionality of the CS is given by the rank of the covariance matrix between the source (V1) and target (V2) subpopulations (see Methods: Interpretation of the results). The dimensionality of the covariance matrix is influenced by three main factors. First, the dimensionality remains low in the absence of normalization (Suppl. Fig. 11h). However, including normalization specifically reduces dimensionality at coherent frequencies (Fig. 4c). Second, the dimensionality is increased for a fully interconnected (all-to-all) connectivity between V1 and V2 (Suppl. Fig. 11l). Third, delocalized inputs (e.g., noise images or naturalistic images) lead to a higher dimensionality than localized inputs (e.g., gratings) (Suppl. Fig. 11p) as observed in experimental results [32].

Dynamic modulation of functional connectivity through feedback

Anatomical connections influence neural communication but do not completely determine it. Rather, it has been observed experimentally that functional connectivity between brain areas changes dynamically [106–108].

Our theory predicts that functional connectivity is controlled by modulating feedback gain (Fig. 5). We expanded our model to encompass three cortical areas: V1, V4, and V5 (also called MT). In this expanded model, V1 has reciprocal connections with both V4 and V5, representing the parallel dorsal and ventral visual pathways, respectively [109–111]. For simplicity, we assume there are no direct connections between V4 and V5 in the model. Analysis of this three-area model provides compelling evidence for

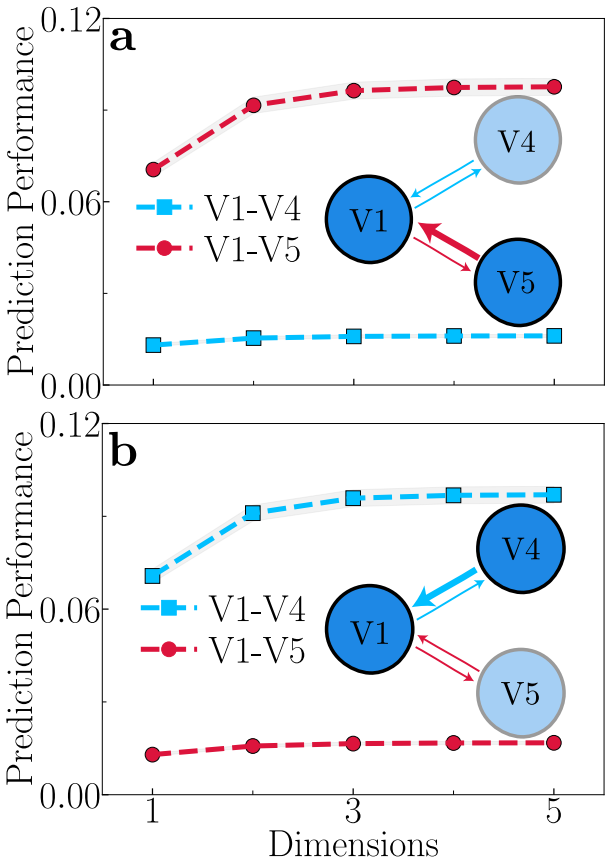

Figure 5: ***Theoretical prediction: Feedback-dependent modulation of functional connectivity. a,*** *Feedback from V5 → V1 is stronger than feedback from V4 → V1, and V1 preferentially enhances communication with V5.* ***b,*** *Feedback from V4 → V1 is stronger than feedback from V5 → V1, and V1 preferentially enhances communication with V4. The plotted prediction performance is an average across different subsets of the source and target populations, and the shaded areas represent the standard error of the mean (SEM). These results demonstrate that top-down feedback can dynamically route information flow between cortical areas, i.e., modulating functional connectivity.*

the role of top-down feedback in controlling functional connectivity. When the feedback strength from V5 $\rightarrow$ V1 is greater than the feedback from V4 $\rightarrow$ V1, we observe that V1 strengthens communication with V5 (Fig. 5a). Conversely, when the feedback from V4 to V1 is stronger, V1 shows enhanced communication with V4 (Fig. 5b). This suggests that the relative strength of feedback from the higher cortical area (V5 or V4) can dictate the information flow from the lower cortical area (V1). One possible way this could be achieved is by selectively enhancing the neural activity in either V5 or V4, which will, in turn, increase the feedback to V1. This is commensurate with experimental evidence that feature-based attention and task demands may selectively boost activity in one or another cortical area [112–118], and specifically that attending to motion increases the transfer of motion information (communication) between V1 and V5 [119–121]. Therefore, top-down signals can dynamically modulate functional connectivity from lower to higher cortical areas, simply by adjusting the gain of neural activity in one higher cortical area (e.g., V4) relative to another (e.g., V5). This mechanism enables flexible cortical information routing, which is essential for cognitive flexibility.

Note that, to isolate the variable of interest, we modeled V4 and V5 with identical parameters for their recurrent, feedforward (V1 $\rightarrow$ V4/V5) and feedback (V1 $\leftarrow$ V4/V5) connectivity. This ensures that the observed increase in communication with either V4 or V5 is a direct consequence of the stronger feedback from that specific area.

Robustness with respect to model parameters

All the results in this paper are obtained from the same set of parameters (Table 1) unless otherwise specified. To ensure the robustness of our findings, we verified them against a wide range of model parameters.

$\tau_{\mathbf{y}_1,\mathbf{u}_1,\mathbf{a}_1,\mathbf{q}_1}$	$\tau_{\mathbf{y}_2,\mathbf{u}_2,\mathbf{a}_2,\mathbf{q}_2}$	γ_1	$\beta_{\{1,2\}}$	$\alpha_{\{1,2\}}$	$\sigma_{\{1,2\}}$
1 ms	1 ms	1.0	1.0	10	0.07

Table 1: *Default values of the parameters*

Parameter	Tested stability range
γ_1	[0.0 - 1.1]
$\beta_{\{1,2\}}$	[0.0 - 2.5]
$\alpha_{\{1,2\}}$	[7.5 - 50]
$\sigma_{\{1,2\}}$	[0.035 - 0.14]

Table 2: *Valid parameter ranges*

We systematically repeated the calculations and simulations described above, varying each parameter individually while holding others at their baseline values. These parameter ranges are specified in Table 2. Overall, our key findings remain robust across a wide range of parameter regimes without requiring precise parameter tuning:

- Contrast-response functions (Fig. 2a). Higher cortical area (V2) consistently exhibits greater contrast gain, reflected by a lower semi-saturation contrast and a steeper exponent than lower cortical area (V1).
- Gamma oscillations (Fig. 2c,e). Both the power spectra and coherence spectra exhibit prominent gamma-band peaks that shift to higher frequencies as contrast increases.
- Alpha oscillations (Fig. 3c). Low contrast inputs and weak feedback gain produce clear alpha-band peaks in the power spectra.
- Communication subspaces (Fig. 2g). The V1-V2 subspace is lower dimensional than the V1-V1 subspace.
- Communication subspaces and feedback gain (Fig. 3k). The V1-V2 communication strength consistently increases with increasing feedback gain.
- Communication subspaces and input gain (Fig. 3l). V1-V1 and V1-V2 communication strength consistently decrease with increasing input gain.
- Frequency-dependence of communication subspaces (Fig. 4). The communication strength peaks at the same frequencies as coherence for each contrast level, and subspace dimensionality is minimized at those frequencies.
- Dynamic functional connectivity (Fig. 5). Enhancing feedback from V4 to V1, selectively boosts communication between V1 and V4, and likewise, enhancing feedback from V5 to V1, selectively enhances communication between V1 and V5.
- Oscillatory activity depends on normalization (Suppl. Fig. 11e,f). Removing normalization abolishes the gamma frequency peaks in both power and coherence spectra.
- Dimensionality of inter-areal communication subspaces (Suppl. Fig. 11l,p). The dimensionality is low when the input drive is localized (e.g., gratings). However, delocalized input drives (e.g., naturalistic images) result in a higher dimensionality. Fully all-to-all connections between V1 and V2 also produce a higher subspace dimensionality.

Discussion

We introduce an analytically tractable theory of neural circuit function that implements divisive normalization dynamically across multiple interconnected brain areas. Emergent properties of the theory account for a diverse array of neurophysiological phenomena, including oscillatory activity (measurable as peaks in LFP power and coherence spectra) and communication subspaces. Based on the theory, we propose a unified view of inter-areal communication with the following experimentally-testable predictions: i) increasing feedback strength should enhance inter-areal communication and diminish within-area communication, without altering the subspace dimensionality; and ii) frequencies exhibiting high coherence should correspond to both stronger communication and reduced subspace dimensionality. Furthermore, the theory proposes a mechanism for dynamic functional connectivity: top-down feedback signals enhance communication between brain areas simply by adjusting the gain of neural activity in one or more cortical areas. This modulation of neural gain offers a broader mechanism for dynamically changing the functional connections between brain areas over time.

Neural circuit function

The original ORGaNICs model [16, 17], in its initial form without feedback connections, serves as a biophysically-plausible counterpart to Long Short Term Memory units (LSTMs) [122], a class of artificial recurrent neural networks (RNNs) prevalent in machine-learning applications [123–126]. Unlike LSTMs or any other previously proposed RNNs, ORGaNICs have built-in recurrent normalization, which offers several advantages, including stability in the network even in circuits with exuberant recurrent connections [18, 19].

The hierarchical version of the ORGaNICs theory introduces feedback connections, drawing a partial analogy to "hierarchical predictive coding" models of sensory and perceptual processing [127–132]. However, the hierarchical model diverges from predictive coding in its fundamental mechanisms. In predictive coding, it is proposed that forward-propagating signals represent prediction errors, while feedback pathways convey predictions to lower levels of hierarchy. The goal is to minimize prediction error by matching sensory input with top-down predictions. In hierarchical ORGaNICs, instead of simply propagating prediction errors, the interplay between feedforward and feedback drives adjusts neuronal responses to a value that balances the feedforward processing of its inputs with the feedback (see Ref. [15] for details). This mechanism is more consistent with neurophysiological and psychophysical phenomena than predictive coding models. For instance, neurons maintain sustained activity in response to a predictable stimulus [133], whereas predictive coding theories would suggest that the activity, representing prediction error, should be diminished or "explained away" as the stimulus becomes predictable.

Normalization

The proposed theory has an analytical fixed point that follows the normalization equation (Eq. 2), for arbitrary (non-negative) normalization weights, when the input drive is constant over time (see Appendix for derivation) [18, 19]. The model achieves normalization dynamically via recurrent amplification, which is inversely regulated by the activity of inhibitory modulator cells. When the input is weak, the modulator cells have a small response, resulting in strong recurrent amplification. Conversely, a strong input generates a large response from modulator cells, which suppresses the amplification. This process modulates both excitatory and inhibitory recurrent signals, which is consistent with the experimental observation that normalization involves a decrease in both recurrent excitatory and inhibitory conductances [134]. Because the theory successfully implements normalization, it is commensurate with a wide range of physiological and psychophysical phenomena previously explained by the normalization model (see Introduction for references). The functional importance of this mechanism is further validated by our model. When divisive normalization is removed from the model, oscillatory peaks in both the power and coherence spectra are eliminated (Suppl. Fig. 11e,f). Also, the relationship between coherence, prediction performance, and subspace dimensionality is lost (Suppl. Fig. 11g,h).

Role of feedback

This paper provides an experimentally testable computational model of cortical function where neural activity is shaped by a combination of feedforward drive (bottom-up) and feedback drive (top-down). Feedback in the model may serve a variety of functions. First, feedback provides a means for controlling inter-areal communication (functional connectivity) simply by changing the gain of neural activity in one or more cortical areas (see below, Attention and functional connectivity). Second, the feedback drive may contribute to selective attention (see below, Attention and functional connectivity). Third, the relative strength of the input gain and feedback gain determines whether neural responses are driven

bottom-up, top-down, or a combination of the two. Fourth, feedback may contribute to learning to minimize prediction error [15, 135]. In addition, feedback to inhibitory neurons is essential for maintaining network stability.

The interplay between the input gain and the feedback gain determines the primary driver of neural responses. If the input gain is strong and the feedback is weak, the system behaves like traditional feedforward models. Conversely, with strong feedback and weak input, the network is similar to generative models [136–138], creating sensory representations from higher-level abstract knowledge, akin to recalling an image from memory (see Ref. [15] for details). When input and feedback are balanced, the framework integrates prior beliefs with sensory evidence, akin to models based on Bayesian inference [139–145]. The model predicts differential effects of changing input gain versus feedback gain on the power, coherence spectra, and inter-areal communication (see Fig. 3). Hence, these metrics may be used as a signature for distinguishing changes in input gain from changes in feedback.

Unified view of inter-areal communication

Our theoretical framework offers a novel perspective on inter-areal communication. Rather than viewing "communication through coherence" (CTC) and "communication subspace" (CS) as alternative hypotheses, our theory unifies them, showing that coherence and communication subspaces are complementary empirical phenomena emerging naturally from the same underlying mechanism. We found that the frequency decomposition of communication subspaces closely mirrors the patterns of inter-areal coherence at different contrasts. Specifically, frequencies exhibiting high coherence are also those for which the communication efficacy within the subspace is maximal, and the subspace dimensionality is minimized. Consequently, coherence might dynamically shape the communication subspaces, for instance, by reducing their dimensionality at preferred frequencies [34]. Moreover, the predicted frequency-dependence of communication subspaces might serve as a marker for testing neural circuitry theories (such as ours).

Attention and functional connectivity

The theory provides a possible explanation for the efficient routing of sensory information between cortical areas, i.e., for changing functional connectivity. Lower cortical areas, such as V1, can flexibly relay information to specific higher cortical areas (e.g., V4 vs. V5) based on the relative strength of the feedback from each of the higher cortical areas. The strength of the feedback can be controlled either by changing the feedback gain or (perhaps more simply) by changing the gain of the neural responses in each of the higher cortical areas. We hypothesize that this feedback-mediated process may account for changes in functional connectivity [108, 146, 147] and feature-based attention [112–116]. Consistent with this hypothesis, experimental evidence demonstrates that feature-based attention is spatially global, i.e., that neural activity is boosted in feature-selective neurons with receptive fields covering the visual field [113, 148].

The theory may also provide insights into the neural processes underlying spatially-selective attention. According to the theory, there are two processes by which attention may modulate the gain of neural activity: i) attention may modulate the input gain of the sensory neurons, following earlier work on the normalization model of attention [53]; ii) attention may modulate the activity of the sensory neurons via the feedback drive. One or the other of these two proposed mechanisms for attentional modulation (e.g., changing the input gain or the feedback drive) may offer a better fit to electrophysiology and behavioral data. Alternatively, one of the mechanisms (e.g., changing the input gain) may offer a better explanation of exogenous attention, whereas the other mechanism (e.g., the feedback drive) may offer a better explanation of endogenous attention. Importantly, the theory predicts different effects of changing input gain versus feedback gain on the power and coherence spectra, as noted above, so these metrics may be used as a signature for distinguishing changes in input gain from changes in feedback.

Thalamo-cortical model

Alpha oscillations are a defining characteristic of thalamo-cortical dynamics, particularly prominent within the visual system during states of "relaxed wakefulness" [149–151]. Specifically, alpha oscillations are hypothesized to be generated and modulated by thalamo-cortical loops [151, 152], suggesting that alpha rhythms signify a state of functional inhibition or "idling" in sensory pathways, potentially gated by corticothalamic feedback [152, 153]. Our theory predicts the emergence of oscillatory activity in the alpha band (8-12 Hz) under conditions of low stimulus contrast and low feedback to excitatory neurons, and high feedback to inhibitory neurons ($\gamma < 1$, Figs. 3c,g). This aligns well with the existing literature demonstrating the role of alpha oscillations in mediating cortical inhibition [153, 154]. This positions the theory as a potential tool for modeling the

thalamocortical interactions underlying alpha waves.

Furthermore, we hypothesize that the input gain modulator responses depend in part on thalamocortical loops, hence positioning the theory as a means for modeling the interplay between bottom-up sensory input and top-down feedback within the thalamocortical circuit, the gating of visual information flow, thalamic control of functional cortical connectivity [155] and thalamocortical loops in other neural systems [156–160].

Limitations

We assume the feedback to be instantaneous, but there is an associated feedback delay [161, 162]. The delay in receiving feedback signals from higher cortical areas is due to the finite speed of signal conduction, processing, and computation. These delays arise from factors like axonal conduction times and synaptic transmission. It will be interesting to investigate how the synchronization between cortical areas changes as a function of feedback delay time.

The feedforward, feedback, and recurrent weight matrices may comprise positive and negative weights. Hence, we posit additional inhibitory interneurons to invert the sign of the responses, corresponding to negative weights in the synaptic weight matrices. These inhibitory neurons need not be one-to-one with their excitatory inputs. Rather, each may compute a partial weighted sum of their inputs corresponding to the negative terms in the matrix multiplications. Regardless, these inhibitory interneurons will introduce an additional delay.

Most of the computations posited by the theory (see Methods, Eqs. 1a, 1c, 1b, and 1d) are biophysically plausible (see below and Ref. [17]), but there are some elements of computation for which the underlying mechanisms are unclear. For example, the weighted sums may be implemented as a push-pull balanced increase in excitatory and decrease in inhibitory synaptic inputs (or vice versa) that injects current without changing conductance [163] and the gain control may be implemented with a push-push balanced increase in excitatory and inhibitory synaptic inputs that changes conductance without injecting current [163, 164]. However, we do not have a hypothesis for the cellular mechanism underlying the element-by-element product of firing rates $\mathbf{y}_1^+ \odot \mathbf{u}_1^{+2}$ in Eq. 1b.

The intrinsic time constants are unrealistically short for individual neurons (see Table 1). If more realistic time constants were used, the oscillatory activity and coherence in the gamma-frequency band would shift toward a lower frequency. Instead, one might interpret the time constants in the theory as effective time constants of a population of neurons rather than interpreting them literally as the intrinsic time constants (capacitance/conductance) of individual neurons.

Biology plausibility

We hypothesize a correspondence between the variables in the equations that describe the theory and cortical microcircuits composed of identified cell types. Unlike generic E-I models (e.g., Wilson and Cowan [165]), we propose that different cell types (different morphology and biochemistry) perform different computations, each with a different output nonlinearity (i.e., F-I curve) transforming membrane potentials to firing rates (see Appendix). Specifically, principal cells **y** may correspond to pyramidal cells, inhibitory modulatory cells **a** may correspond to a combination of parvalbumin-expressing (PV+) cells and somatostatin-expressing (SST+) cells, and inhibitory interneurons **q** may correspond to vasoactive intestinal polypeptide (VIP) cells. Principal cells **y** provide excitatory input to both inhibitory modulatory cells **a** (via **u**) and inhibitory interneurons **q**, inhibitory interneurons **q** provide inhibitory input to inhibitory modulatory cells **a**, and inhibitory modulatory cells **a** provide inhibitory input to principal cells **y** (Fig. 1). This arrangement is commensurate with empirical observations of cortical circuits: pyramidal cells provide excitatory input to PV+, SST+, and VIP cells, VIP cells provide inhibitory input to SST+ cells, and both PV+ and SST+ cells provide inhibitory input to pyramidal cells [166–170]. The inhibitory modulator cells **a** are hypothesized to have large RFs and broad orientation-selectivity (reflecting properties of the normalization pool), consistent with the response properties of SST+ and PV+ neurons, respectively [167, 171, 172]. We also hypothesize excitatory feedback from neurons in higher cortical areas to both principal cells **y** and inhibitory modulatory cells **a** (Fig. 1). Feedback to principal cells may be received by the distal apical dendrites of pyramidal cells in layer 1 (see below). Feedback to inhibitory modulator cells is commensurate with experimental evidence that feedback from higher cortical areas targets SST+ neurons [173–175], and that normalization depends in part on loops through higher visual cortical areas [176–179]. A possible discrepancy between the model and the actual cortical circuit is that the excitatory input from principal cells **y** to inhibitory modulatory cells **a** passes through another excitatory cell type **u** whereas pyramidal cells provide direct excitatory input to PV+ and SST+ cells. One could, however, rewrite the equations for the theory (see Methods, Eqs. 1c and 1b) such that principal cells provide direct excitatory input to inhibitory modulatory cells (specifically by substituting the expression

for **u** from Eq. 1b into the expression for **a** in Eq. 1c).

We hypothesize a correspondence between the computations performed by the principal cells (**y**) and the computational processes occurring within the dendritic compartments of pyramidal cells. The equation governing the responses of the principal cells (see Fig. 6 and Methods, Eq. 1a) expresses the following synaptic current: (input gain) (input drive) + (recurrent gain) [recurrent drive + (feedback gain) (feedback drive)]. The input drive may be computed as a weighted sum of feedforward inputs in the distal basal dendrites, commensurate with evidence that feedforward synaptic inputs contact basal dendrites. The input gain may correspond to inhibitory neurons that contact the most proximal trunk of the basal dendritic tree. The feedback drive may be computed as a weighted sum of feedback inputs received by distal apical dendrites in layer 1, which are known to receive signals from higher cortical areas. The feedback gain might be attributed to the amplification of the feedback drive within the apical trunk of pyramidal cells [180]. The recurrent drive may correspond to a weighted sum of lateral inputs at more proximal apical dendrites, and the recurrent gain may correspond to inhibitory inputs that contact the most proximal trunk of the apical dendritic tree.

Experimental evidence suggests that acetylcholine (ACh) selectively adjusts the balance between bottom-up sensory inputs and top-down signals related to expectancy and uncertainty [181]. It has also been hypothesized that ACh signals when bottom-up sensory inputs are known to be reliable [182, 183]. Consequently, it is reasonable to hypothesize that the input gain and/or feedback gain might be controlled (at least in part) by ACh. Although ACh is released broadly throughout the cortex, its effect can be regionally specific [184], possibly offering a mechanism for how the values of the state parameters can differ across the hierarchy of brain areas.

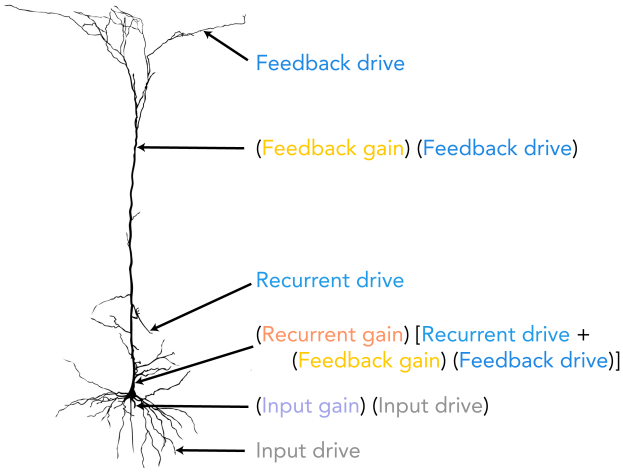

Figure 6: ***Hypothesized mapping of computational components in the model onto the dendritic compartments of a pyramidal cell.*** *The input drive is hypothesized to be the weighted sum of feedforward inputs arriving at the distal basal dendrites. This is modulated by an input gain, corresponding to inhibition at the proximal basal dendritic trunk. In the apical tuft, the feedback drive represents a weighted sum of feedback inputs from higher cortical areas. This signal is amplified by a feedback gain within the distal apical trunk. The recurrent drive corresponds to the sum of lateral inputs on the proximal apical dendrites. The combined recurrent and feedback signals are then modulated by a recurrent gain at the proximal apical trunk. The overall synaptic current is a combination of these gain-modulated drives. The figure is taken from [185].*

Methods

The model

We propose a multistage hierarchical recurrent circuit that incorporates feedforward connections from the primary visual cortex (V1) to higher visual cortical areas (e.g., V2, V4, MT) and feedback connections from these higher brain areas to V1 (Fig. 1). We refer to V2 throughout this paper for concreteness, but it is merely an example of a higher brain area that is reciprocally connected with V1. This model builds on the ORGaNICs framework developed by Heeger et al. [16, 17]. Divisive Normalization (DN) is achieved in each brain area by local inhibitory neuronal populations.

The response of principal excitatory neurons in the model is a sum of three key components: 1) input drive, a weighted sum of neural responses from a brain area that provides feedforward input (e.g., LGN $\rightarrow$ V1 and V1 $\rightarrow$ V2), modulated by input gain; 2) feedback drive, a weighted sum of neural responses from the higher cortical area (e.g., V1 $\leftarrow$ V2 or V1 $\leftarrow$ V4), modulated (viz., scaled) by feedback gain; and 3) the sum of recurrent drive—a weighted sum of neuronal responses from other neurons in the same brain area (e.g., V1 $\rightarrow$ V1 and V2 $\rightarrow$ V2)—and scaled feedback drive, collectively modulated by recurrent gain.

The proposed neural circuit comprises four distinct cell types, the principal excitatory neurons (**y**), the modulatory excitatory neurons (**u**), the modulatory inhibitory neurons (**a** and **q**), with the following dynamics:

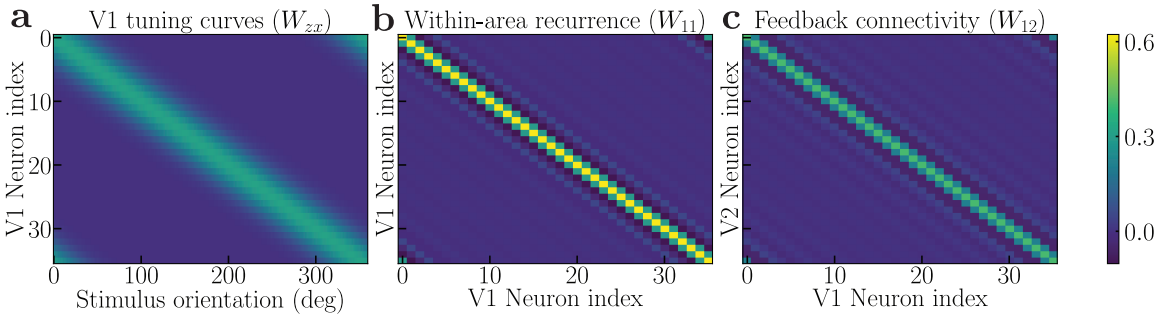

Figure 7: ***Connectivity matrices in the hierarchical recurrent circuit model.*** *a, The V1 encoding matrix ($\mathbf{W}_{zx}$), each row represents the tuning curve of a V1 neuron.* ***b****, The recurrent connectivity matrix for V1 ($\mathbf{W}_{11}$). The V2 recurrent connectivity matrix $\mathbf{W}_{22}$ has a similar structure.* ***c****, The Feedback connectivity matrix from V2 to V1 ($\mathbf{W}_{12}$). The corresponding feedforward matrix from V1 to V2, ($\mathbf{W}_{21}$), is its transpose. Since $\mathbf{W}_{21}$ is symmetric, the feedforward and feedback matrices are identical ($\mathbf{W}_{12} = \mathbf{W}_{21}$).*

$$\tau_{y_1}\frac{\mathrm{d}\mathbf{y}_1}{\mathrm{d}t} = -\mathbf{y}_1 + \frac{\beta_1}{2}\odot\mathbf{z}_1 + \left(\frac{1}{1+\mathbf{a}_1^+}\right)\left(\mathbf{W}_{11}\sqrt{\mathbf{y}_1^+} + \frac{\gamma_1}{2}\odot\mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}\right) \tag{1a}$$

$$\tau_{u_1}\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} = -\mathbf{u}_1 + \left(\frac{\sigma_1}{2}\right)^2 + \mathbf{N}_1\left(\mathbf{y}_1^+\odot\mathbf{u}_1^{+2}\right) \tag{1b}$$

$$\tau_{a_1}\frac{\mathrm{d}\mathbf{a}_1}{\mathrm{d}t} = -\mathbf{a}_1 + \frac{\mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}}{2\cdot\mathbf{q}_1^+} + \mathbf{u}_1^+ + \mathbf{a}_1\odot\mathbf{u}_1^+ + \alpha_1\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} \tag{1c}$$

$$\tau_{q_1}\frac{\mathrm{d}\mathbf{q}_1}{\mathrm{d}t} = -\mathbf{q}_1 + \sqrt{\mathbf{y}_1^+}. \tag{1d}$$

The above equations are for V1; V2 follows a similar set of equations (see Appendix). The membrane potentials of neurons in cortical areas V1 and V2 are denoted by $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively, while their corresponding firing rates are represented by $\mathbf{y}_1^+$ and $\mathbf{y}_2^+$. The $\odot$ symbol represents the Hadamard product, element-wise multiplication of vectors. To estimate the firing rates $\mathbf{y}_1^+$ and $\mathbf{y}_2^+$ from their respective membrane potentials, we employ a modified Gaussian Rectification (GR) model [186, 187]. The variable $\mathbf{z}_1$ denotes the input drive to V1 from the preceding brain area (the lateral geniculate nucleus of the thalamus, LGN), computed as a weighted ($\mathbf{W}_{zx}$) sum of the LGN firing rates. $\tau_\mathbf{k}$ is the intrinsic time constant for each subtype of neuron, where $\mathbf{k} = \mathbf{y}, \mathbf{u}, \mathbf{a}, \mathbf{q}$. $\mathbf{W}_{11}\sqrt{\mathbf{y}_1^+}$ and $\mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}$ are the recurrent and feedback drive, respectively, where $\mathbf{W}_{11}$ is the recurrent weight matrix for V1 . $\mathbf{W}_{11}$ has a center-surround structure where the closer connections are excitatory and the more distant connections are inhibitory (Fig. 7b), same as in [17] where the feedforward ORGaNICs was introduced. The recurrent weight matrix for V2, $\mathbf{W}_{22}$, has a similar structure. The maximum eigenvalue for these recurrent weight matrices is set to unity to ensure stability. However, we have shown that this need not be the case because normalization ensures stability [18, 19]. $\mathbf{W}_{12}$ represents the inter-areal feedback connectivity matrix from V2 to V1, its entries are all non-negative, and the matrix is diagonally dominant (Fig. 7c). The feedforward matrix from V1 to V2, $\mathbf{W}_{21}$, is taken to be the transpose of $\mathbf{W}_{12}$. Given that $\mathbf{W}_{12}$ is symmetric, it follows that $\mathbf{W}_{12} = \mathbf{W}_{21}$. The normalization matrix in V1 ($\mathbf{N}_1$) and V2 ($\mathbf{N}_2$) is taken to be all ones, meaning each neuron contributes equally to the normalization. The input gain and feedback gain for V1 are modulated by β_1 and γ_1, respectively. The recurrent and feedback amplifications in V1 are modulated dynamically by the inhibitory neuron population $\mathbf{a}_1$.

Both the V1 and V2 neurons follow the normalization equation *exactly* when β and γ are set to unity (i.e. when there is identical feedback to excitatory ($\mathbf{y}_1$) and inhibitory ($\mathbf{a}_1$) neurons) and $\mathbf{W}_{11}$ is the identity matrix (i.e., each neuron recurrently amplifies itself), but we have shown that the circuit closely approximates normalization for a wide range of recurrent weight matrices ($\mathbf{W}_{11}$) [19]. For V1 neurons, normalization is expressed as:

$$\mathbf{y}_1^+ = \frac{\lfloor\mathbf{z}_1\rfloor^2}{\sigma^2 + \mathbf{N}_1\lfloor\mathbf{z}_1\rfloor^2}, \tag{2}$$

and likewise for V2

$$\mathbf{y}_2^+ = \frac{\lfloor\mathbf{z}_2\rfloor^2}{\sigma^2 + \mathbf{N}_2\lfloor\mathbf{z}_2\rfloor^2}, \tag{3}$$

where $\mathbf{z}_2 = \mathbf{W}_{21}\mathbf{y}_1^+$ is the drive from V1 to V2, σ is the semi-saturation constant. Normalization is achieved dynamically via recurrent amplification (amplifying weak inputs but not strong inputs) with the three modulatory neurons, $\mathbf{u}_1$, $\mathbf{a}_1$ and $\mathbf{q}_1$ whose dynamics are governed by Eq. 1b, Eq. 1c, and Eq. 1d respectively. The equations for modulatory neurons

are derived by substituting the normalization equation (Eq. 2) in the equation for the principal neurons (Eq. 1a) (see Appendix for derivation). An increase in the activity of principal neurons ($\mathbf{y}_1$) leads to enhanced activation in $\mathbf{u}_1$ and $\mathbf{q}_1$. This, in turn, drives a net increase of activity in modulatory neurons ($\mathbf{a}_1$). The elevated $\mathbf{a}_1$ activity subsequently exerts an inhibitory effect on $\mathbf{y}_1$, thereby implementing normalization by the recurrent interactions between the neural populations (see Appendix for derivation).

The parameter γ_1 modulates the feedback gain to the principal excitatory neuron population ($\mathbf{y}_1$). This determines the relative feedback to the excitatory neurons ($\mathbf{y}_1$) compared to the inhibitory neurons ($\mathbf{a}_1$). As a rule of thumb, $\gamma_1 < 1$ indicates greater feedback inhibition, $\gamma_1 = 1$ denotes a balance between feedback excitation and inhibition, and $\gamma_1 > 1$ signifies more feedback excitation. For simplicity, we have set the feedback to the inhibitory neurons ($\mathbf{a}_1$) at a fixed value of 0.50 (hence, the factor of $\frac{1}{2}$ in equation Eq. 1c and Eq. 1a, see Appendix for details). Similarly, the parameter β_1 modulates the relative input gain to the principal excitatory neuron population ($\mathbf{y}_1$) compared to the modulatory excitatory neuron population ($\mathbf{u}_1$). For simplicity, we set the input gain to the modulatory neurons ($\mathbf{u}_1$) to 0.50 (hence, the factor of $\frac{1}{2}$ in equation Eq. 1a and Eq. 1b, see Appendix). Note that when β_1 and γ_1 are both set to unity and the recurrent matrix is identity, the network achieves excitatory-inhibitory balance exactly, and the normalization (Eq. 2) is followed precisely.

We modeled each brain area using a ring model, where principal neurons are arranged in a circle to represent continuous features like stimulus orientation. The recurrent connectivity between these neurons is determined by the distance between them (Fig. 7b). Each modeled area consists of 72 neurons for each of the four specified types ($\mathbf{y}_1$,$\mathbf{u}_1$,$\mathbf{a}_1$,$\mathbf{q}_1$). To simulate sensory input, a visual stimulus with a defined contrast and orientation is presented. The stimulus is then convolved with the tuning curves of the V1 neurons to provide the input drive to the V1 principal neurons. The tuning curves are modeled as raised cosine functions, each peaking at the neuron's preferred orientation (Fig. 7a).

Power and Coherence Spectra

The time evolution of neural population activity in V1 and V2 can be described by a general stochastic differential equation (SDE) with an additive noise term $\mathbf{L}\mathrm{d}\mathbf{W}$:

$$\mathrm{d}\mathbf{x} = \mathbf{f(x, u)}\mathrm{d}t + \mathbf{L}\mathrm{d}\mathbf{W}, \tag{4}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a vector representing the activity of neurons; $\mathbf{f(x, u)} \in \mathbb{R}^n$ represents the deterministic part of the dynamics explained above (Eq. 1); $\mathrm{d}\mathbf{W}$ is a vector of independent Gaussian increments with correlation matrix $\mathbf{D}$ ($\mathbb{E}[\mathrm{d}\mathbf{W} \cdot \mathrm{d}\mathbf{W}^\mathsf{T}]$), and $\mathbf{L} \in \mathbb{R}^{n\times n}$ is the dispersion matrix defining how noise enters the system. Given the analytical solution for the fixed point of the dynamical system (Eq. 2), we can linearize the system around the fixed point ($\mathbf{x}^*$):

$$\mathrm{d}\mathbf{x} = \mathbf{J}_{\mathbf{x}^*}(\mathbf{x} - \mathbf{x}^*)\mathrm{d}t + \mathbf{L}\mathrm{d}\mathbf{W}, \tag{5}$$

where $\mathbf{J}_{\mathbf{x}^*}$ is the Jacobian computed at the fixed point $\mathbf{x}^*$. This linearization leads to a wide-sense stationary Gaussian process, $\mathbf{x}(t)$, whose power spectral density matrix, $\boldsymbol{\mathcal{S}}(\omega) \in \mathbb{R}^{n\times n}$, is given by [188]:

$$\boldsymbol{\mathcal{S}}(\omega) = \mathbb{E}\left[|\hat{\mathbf{x}}(\omega)|^2\right] = (\imath\omega\mathbf{I}+\mathbf{J})^{-1}\mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}(-\imath\omega\mathbf{I}+\mathbf{J})^{-\mathsf{T}}. \tag{6}$$

From $\boldsymbol{\mathcal{S}}(\omega)$, we can directly compute the coherence, κ_{ij}, between any two i and j neurons in the two areas:

$$\kappa_{ij} = \frac{|\mathcal{S}_{ij}|^2}{\mathcal{S}_{ii}\mathcal{S}_{jj}}. \tag{7}$$

Communication subspaces

Here we present the method used to characterize the communication subspaces between two subpopulations (source and target) of neurons [32, 162, 189, 190]. These subpopulations may comprise distinct neurons in the same brain area or across different brain areas. The basic idea is to determine how well the fluctuations in responses of a target subpopulation can be predicted using a linear transform (a weighted sum) of the fluctuations in responses of the source subpopulation, and to determine the dimensionality of this linear transform.

The fluctuations in the responses and their covariability are fully determined by the correlation matrix at steady-state when the strength of the noise is small. Therefore the correlation matrix, denoted as $\mathbf{C}(0)$, for the responses $\mathbf{x}(t)$ can be computed by solving the Lyapunov equation [188]:

$$\mathbf{J}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{J}^\mathsf{T} = -\mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}. \tag{8}$$

Once $\mathbf{C}(0)$ is obtained, we follow the method proposed by Semedo et al. [32] to construct an analytical model of the communication subspace. To achieve this, we divide the mean-subtracted responses of the entire neuronal population, denoted as $\mathbf{x}$, into two subsets: $\mathbf{s}$ (responses of the source neurons, e.g., in V1) and $\mathbf{t}$ (responses of the target neurons, e.g., in V2 or V1). We consider a linear model for predicting target responses using the source responses, $\hat{\mathbf{t}} = \mathbf{B}^\mathsf{T}\mathbf{s}$, where the mean squared error (MSE) is given by:

$$\varepsilon = \mathbb{E}[|\mathbf{t} - \mathbf{B}^\mathsf{T}\mathbf{s}|^2] = \mathrm{Tr}(\mathbf{C}_2 + \mathbf{B}^\mathsf{T}\mathbf{C}_1\mathbf{B} - 2\mathbf{B}^\mathsf{T}\mathbf{C}_3), \tag{9}$$

where $\mathbf{C}_1 = \mathbb{E}[\mathbf{s}\mathbf{s}^\mathsf{T}]$, $\mathbf{C}_2 = \mathbb{E}[\mathbf{t}\mathbf{t}^\mathsf{T}]$, and $\mathbf{C}_3 = \mathbb{E}[\mathbf{s}\mathbf{t}^\mathsf{T}]$. These matrices can be directly obtained from $\mathbf{C}(0)$. To find the optimal readout matrix $\mathbf{B}_{\text{opt}}$, we minimize ε with respect to $\mathbf{B}$, leading to the Ordinary Least Squares solution:

$$\mathbf{B}_{\text{opt}} = \mathbf{C}_1^{-1}\mathbf{C}_3, \tag{10}$$

First, we define the covariance of the target activity *predicted* by the optimal model (note the "hats"), which we denote as $\hat{\mathbf{C}}_2$:

$$\hat{\mathbf{C}}_2 = \mathbb{E}[\hat{\mathbf{t}}\hat{\mathbf{t}}^\mathsf{T}] = \mathbf{B}_{\text{opt}}^\mathsf{T}\mathbf{C}_1\mathbf{B}_{\text{opt}} \tag{11}$$

By substituting the expression for $\mathbf{B}_{\text{opt}}$, we arrive at a key relationship:

$$\hat{\mathbf{C}}_2 = \left(\mathbf{C}_1^{-1}\mathbf{C}_3\right)^\mathsf{T}\mathbf{C}_1\left(\mathbf{C}_1^{-1}\mathbf{C}_3\right) = \mathbf{C}_3^\mathsf{T}\mathbf{C}_1^{-1}\mathbf{C}_3 \tag{12}$$

The principal dimensions of this predicted covariance matrix, $\hat{\mathbf{C}}_2$, define the communication subspace. These are found by computing the eigendecomposition:

$$\hat{\mathbf{C}}_2 = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\mathsf{T} \tag{13}$$

Here, the columns of the orthogonal matrix $\mathbf{V}$ are the eigenvectors of $\hat{\mathbf{C}}_2$ and $\mathbf{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues, sorted in descending order. The prediction performance for a reduced-rank approximation of dimension i is defined as: A rank-i approximation, $\mathbf{B}_{\text{RRR}}$, is constructed by projecting the optimal solution $\mathbf{B}_{\text{opt}}$ onto the subspace spanned by the top i eigenvectors of $\hat{\mathbf{C}}_2$.

$$\mathbf{B}_{\text{RRR}}(i) = \mathbf{B}_{\text{opt}}\mathbf{V}(i)\mathbf{V}^\mathsf{T}(i) = \mathbf{B}_{\text{opt}}\mathbf{P}_i \tag{14}$$

where $\mathbf{V}(i)$ is a matrix containing first i columns of $\mathbf{V}$. Here $\mathbf{P}_i = \mathbf{V}(i)\mathbf{V}^\mathsf{T}(i)$ is the projection matrix. Note that $\mathbf{P}_i$ is symmetric ($\mathbf{P}_i^\mathsf{T} = \mathbf{P}_i$) and idempotent ($\mathbf{P}_i^2 = \mathbf{P}_i$). We define the prediction performance as:

$$\text{Performance}(i) = 1 - \frac{\varepsilon_i}{\varepsilon_0}, \tag{15}$$

where ε_i is the MSE for the rank-i approximation

$$\varepsilon_i = \text{Tr}(\mathbf{C}_2 + \mathbf{B}_{\text{RRR}}^\mathsf{T}(i)\mathbf{C}_1\mathbf{B}_{\text{RRR}}(i) - 2\mathbf{B}_{\text{RRR}}^\mathsf{T}(i)\mathbf{C}_3). \tag{16}$$

ε_0 is the MSE for the case where the readout matrix is a null matrix, corresponding to the total variance of the target population ($\varepsilon_0 = \text{Tr}(\mathbf{C}_2)$).

To simplify the expressions for ε_i. We substitute $\mathbf{B}_{\text{RRR}}(i) = \mathbf{B}_{\text{opt}}\mathbf{P}_i$ into the expression:

$$\varepsilon_i = \text{Tr}(\mathbf{C}_2) - \text{Tr}\left(2\mathbf{P}_i\mathbf{B}_{\text{opt}}^\mathsf{T}\mathbf{C}_3 - \mathbf{P}_i\mathbf{B}_{\text{opt}}^\mathsf{T}\mathbf{C}_1\mathbf{B}_{\text{opt}}\mathbf{P}_i\right) \tag{17}$$

Now, we substitute $\mathbf{B}_{\text{opt}}^\mathsf{T} = \mathbf{C}_3^\mathsf{T}\mathbf{C}_1^{-1}$ and recognize that the term $\mathbf{B}_{\text{opt}}^\mathsf{T}\mathbf{C}_1\mathbf{B}_{\text{opt}}$ is $\hat{\mathbf{C}}_2$. Using the cyclic property of the trace ($\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$) and the idempotent property of the projection matrix ($\mathbf{P}_i^2 = \mathbf{P}_i$), we can simplify further:

$$\varepsilon_i = \text{Tr}(\mathbf{C}_2) - \text{Tr}(\hat{\mathbf{C}}_2\mathbf{P}_i) \tag{18}$$

By applying the cyclic property and expanding $\mathbf{P}_i$, we can see that the second term is the sum of the top i eigenvalues (λ_j) of $\hat{\mathbf{C}}_2$.

$$\text{Tr}(\hat{\mathbf{C}}_2\mathbf{V}(i)\mathbf{V}(i)^\mathsf{T}) = \text{Tr}(\mathbf{V}(i)^\mathsf{T}\hat{\mathbf{C}}_2\mathbf{V}(i)) = \sum_{j=1}^{i}\lambda_j \tag{19}$$

Now the expression for prediction performance simplifies to:

$$\text{Performance}(i) = \frac{\sum_{j=1}^{i}\lambda_j}{\text{Tr}(\mathbf{C}_2)}, \tag{20}$$

Which can also be interpreted as:

$$\text{Performance}(i) = \frac{\text{Explained variance } (i \text{ dimensions})}{\text{Total target variance}}. \tag{21}$$

The maximum prediction performance (full-rank approximation) is given by:

$$\text{Performance} = \frac{\text{Tr}(\hat{\mathbf{C}}_2)}{\text{Tr}(\mathbf{C}_2)} = \frac{\text{Tr}(\mathbf{C}_3^\mathsf{T}\mathbf{C}_1^{-1}\mathbf{C}_3)}{\text{Tr}(\mathbf{C}_2)} \tag{22}$$

Interpretation of the results: The prediction performance (Eq. 22) is governed by the interplay between the source ($\mathbf{C}_1$), target ($\mathbf{C}_2$), and cross-population ($\mathbf{C}_3$) covariances. The effectiveness of communication (indicated by a large predictive performance) can be understood through three key factors: 1) *Source-target correlation*: Prediction performance is directly enhanced by a stronger correlation between the source and target populations (i.e., larger $\mathbf{C}_3$ norm). 2) *Signal reliability*: Prediction performance is boosted by the reliability of the signals. Less noisy activity in source and target populations, indicated by smaller diagonal entries in $\mathbf{C}_1$ and $\mathbf{C}_2$, leads to better prediction performance. 3) *Alignment*: Prediction performance is maximized when the principal directions of source covariance (eigenvectors of $\mathbf{C}_1$) align with the principal directions of source-target covariance (left singular vectors of $\mathbf{C}_3$).

This framework also explains the results in the frequency domain. The key metric here is the cross-power spectrum, which is the Fourier transform of the cross-covariances, and coherence, which is the squared cross-power spectrum normalized by the product of the auto-spectra, Eq. 7. Frequencies that show high coherence, indicating high cross-covariances between source and target, the prediction

performance is highest at these specific frequencies. Note that the dimensionality of the communication subspace is given by the rank of the matrix $\hat{\mathbf{C}}_2 = \mathbf{C}_3^{\mathsf{T}}\mathbf{C}_1^{-1}\mathbf{C}_3$. Since the source covariance matrix $\mathbf{C}_1$ is a full-rank matrix, the dimensionality of the subspace is determined exactly by the rank of the cross-covariance matrix, $\mathbf{C}_3$. In the frequency domain, this means the dimensionality at a specific frequency is given by the rank of the cross-power spectral density matrix. Our results show that the extent of normalization from surrounding neurons plays a critical role in shaping the dimensionality of the communication subspace. With surround normalization, dimensionality decreases specifically at frequencies where coherence is strongest (Fig. 4c). In contrast, without this normalization, no such correlation between coherence and dimensionality is observed (Fig. 11h).

Frequency decomposition of the communication subspace

To analyze how communication between different brain areas varies across frequencies, we consider the frequency decomposition of the neural responses. The Wiener-Khinchin theorem states that the autocorrelation function $\mathbf{C}(\tau)$ is the inverse Fourier transform of the power spectral density (PSD) matrix ($\boldsymbol{\mathcal{S}}(\omega)$):

$$\mathbf{C}(\tau) = \int_{-\infty}^{\infty} \boldsymbol{\mathcal{S}}(\omega) e^{\imath\omega\tau} d\omega \tag{23}$$

To isolate $\mathbf{C}(\tau)$ at a particular frequency (ω^*), we apply an idealized narrow band-pass filter centered at ω^*. Specifically, the filter's transfer function $H(\omega)$ is given by,

$$H(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where $A = [\omega^* - \epsilon, \omega^* + \epsilon] \cup [-\omega^* - \epsilon, -\omega^* + \epsilon]$ and ϵ is a small positive value defining the narrow bandwidth 2ϵ around ω^*. The spectral density matrix of this filtered signal is given by,

$$\boldsymbol{\mathcal{S}}_{filtered}(\omega) = |H(\omega)|^2 \boldsymbol{\mathcal{S}}(\omega) \tag{25}$$

The resulting autocorrelation of the filtered signal is expressed as:

$$\begin{aligned} \mathbf{C}_{filtered}(\tau) &= \int_{-\infty}^{\infty} \boldsymbol{\mathcal{S}}_{filtered}(\omega) e^{\imath\omega\tau} d\omega \\ &= \int_{-\infty}^{\infty} |H(\omega)|^2 \boldsymbol{\mathcal{S}}(\omega) e^{\imath\omega\tau} d\omega. \end{aligned} \tag{26}$$

Using the definition of $H(\omega)$ and assuming ϵ to be sufficiently small such that $\boldsymbol{\mathcal{S}}(\omega)$ and $e^{\imath\omega\tau}$ are approximately constant over the integration intervals of width 2ϵ, the frequency-specific autocorrelation at ω^* is:

$$\mathbf{C}(\tau)|_{\omega^*} \approx 2\epsilon \left(\boldsymbol{\mathcal{S}}(-\omega^*) e^{-\imath\omega^*\tau} + \boldsymbol{\mathcal{S}}(\omega^*) e^{\imath\omega^*\tau} \right). \tag{27}$$

At zero time lag ($\tau = 0$),

$$\mathbf{C}(0)|_{\omega^*} \approx 2\epsilon \left(\boldsymbol{\mathcal{S}}(-\omega^*) + \boldsymbol{\mathcal{S}}(\omega^*) \right). \tag{28}$$

Since the PSD is Hermitian, i.e., $\boldsymbol{\mathcal{S}}(\omega) = \boldsymbol{\mathcal{S}}^*(-\omega))$, we have

$$\mathbf{C}(0)|_{\omega^*} = 4\epsilon \, \mathrm{Re}\{\boldsymbol{\mathcal{S}}(\omega^*)\} \tag{29}$$

This frequency-specific autocorrelation matrix ($\mathbf{C}(0)|_{\omega^*}$) can then be used in place of $\mathbf{C}(0)$ in the communication subspace analysis described above to evaluate the prediction performance at any frequency. As our analysis does not depend on the absolute scaling of this matrix, the prefactor 4ϵ can be disregarded, and we effectively use $\mathrm{Re}\{\boldsymbol{\mathcal{S}}(\omega^*)\}$ as the frequency-specific correlation matrix.

Acknowledgements

We thank Pascal Fries and Adam Kohn for generously providing the experimental data in Fig 2. We acknowledge Michael Halassa, Hillel Adesnik, Adam Kohn and Flaviano Morone for their valuable feedback on the manuscript, and Ruben Coen-Cagli, John Maunsell, and Jason MacLean for insightful discussions. This work was initiated in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. This work was supported by the National Institute of Health under award numbers R01MH137669 and R01EY035242. S.M. also acknowledges support from the Simons Center for Computational Physical Chemistry (Simons Foundation grant 839534, MT). This work was supported in part through the NYU IT High-Performance Computing resources, services, and staff expertise.

References

[1] David J Heeger. "Normalization of cell responses in cat striate cortex". In: *Visual neuroscience* 9.2 (1992), pp. 181–197.

[2] Matteo Carandini and David J Heeger. "Normalization as a canonical neural computation". In: *Nature reviews neuroscience* 13.1 (2012), pp. 51–62.

[3] Paul-Antoine Salin and Jean Bullier. "Corticocortical connections in the visual system: structure and function". In: *Physiological reviews* 75.1 (1995), pp. 107–154.

[4] Stewart Shipp and Semir Zeki. "The organization of connections between areas V5 and V2 in macaque monkey visual cortex". In: *European Journal of Neuroscience* 1.4 (1989), pp. 333–354.

[5] H Kennedy and J Bullier. "A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey". In: *Journal of Neuroscience* 5.10 (1985), pp. 2815–2830.

[6] Wim Vanduffel et al. "Functional impact of cerebral connections". In: *Proceedings of the National Academy of Sciences* 94.14 (1997), pp. 7617–7620.

[7] JM Alonso et al. "Influence of layer V of area 18 of the cat visual cortex on responses of cells in layer V of area 17 to stimuli of high velocity". In: *Experimental brain research* 93 (1993), pp. 363–366.

[8] Julie H Sandell and Peter H Schiller. "Effect of cooling area 18 on striate cortex cells in the squirrel monkey." In: *Journal of Neurophysiology* 48.1 (1982), pp. 38–48.

[9] Joseph B Hopfinger, Michael H Buonocore, and George R Mangun. "The neural mechanisms of top-down attentional control". In: *Nature neuroscience* 3.3 (2000), pp. 284–291.

[10] Edward Awh, Artem V Belopolsky, and Jan Theeuwes. "Top-down versus bottom-up attentional control: A failed theoretical dichotomy". In: *Trends in cognitive sciences* 16.8 (2012), pp. 437–443.

[11] Farhan Baluch and Laurent Itti. "Mechanisms of top-down attention". In: *Trends in neurosciences* 34.4 (2011), pp. 210–224.

[12] Silvan Quax, Ole Jensen, and Paul Tiesinga. "Top-down control of cortical gamma-band communication via pulvinar induced phase shifts in the alpha rhythm". In: *PLoS computational biology* 13.5 (2017), e1005519.

[13] Hiroshi Makino and Takaki Komiyama. "Learning enhances the relative impact of top-down processing in the visual cortex". In: *Nature neuroscience* 18.8 (2015), pp. 1116–1122.

[14] Wu Li, Valentin Piëch, and Charles D Gilbert. "Perceptual learning and top-down influences in primary visual cortex". In: *Nature neuroscience* 7.6 (2004), pp. 651–657.

[15] David J Heeger. "Theory of cortical function". In: *Proceedings of the National Academy of Sciences* 114.8 (2017), pp. 1773–1782.

[16] David J Heeger and Klavdia O Zemlianova. "A recurrent circuit implements normalization, simulating the dynamics of V1 activity". In: *Proceedings of the National Academy of Sciences* 117.36 (2020), pp. 22494–22505.

[17] David J Heeger and Wayne E Mackey. "Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics". In: *Proceedings of the National Academy of Sciences* 116.45 (2019), pp. 22783–22794.

[18] Shivang Rawat, David Heeger, and Stefano Martiniani. "Unconditional stability of a recurrent neural circuit implementing divisive normalization". In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 14712–14750.

[19] Flaviano Morone et al. "Stabilization of recurrent neural networks through divisive normalization". In: *bioRxiv* (2025), pp. 2025–05.

[20] Andre M Bastos, Julien Vezoli, and Pascal Fries. "Communication through coherence with inter-areal delays". In: *Current opinion in neurobiology* 31 (2015), pp. 173–180.

[21] Pascal Fries. "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence". In: *Trends in cognitive sciences* 9.10 (2005), pp. 474–480.

[22] Mark J Roberts et al. "Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching". In: *Neuron* 78.3 (2013), pp. 523–536.

[23] Emilio Salinas and Terrence J Sejnowski. "Correlated neuronal activity and the flow of neural information". In: *Nature reviews neuroscience* 2.8 (2001), pp. 539–550.

[24] Francisco Varela et al. "The brainweb: phase synchronization and large-scale integration". In: *Nature reviews neuroscience* 2.4 (2001), pp. 229–239.

[25] Pascal Fries et al. "Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry". In: *Proceedings of the National Academy of Sciences* 94.23 (1997), pp. 12699–12704.

[26] Pascal Fries et al. "Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection". In: *Journal of Neuroscience* 22.9 (2002), pp. 3739–3754.

[27] Priscilla E Greenwood and Lawrence M Ward. "Attentional selection and communication through coherence: Scope and limitations". In: *PLOS Computational Biology* 20.8 (2024), e1011431.

[28] Edmund T Rolls and Alessandro Treves. "The neuronal encoding of information in the brain". In: *Progress in neurobiology* 95.3 (2011), pp. 448–490.

[29] Edmund T Rolls, Tristan J Webb, and Gustavo Deco. "Communication before coherence". In: *European Journal of Neuroscience* 36.5 (2012), pp. 2689–2709.

[30] Marius Schneider et al. "A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power". In: *Neuron* 109.24 (2021), pp. 4050–4067.

[31] Supratim Ray and John HR Maunsell. "Do gamma oscillations play a role in cerebral cortex?" In: *Trends in cognitive sciences* 19.2 (2015), pp. 78–85.

[32] João D Semedo et al. "Cortical areas interact through a communication subspace". In: *Neuron* 102.1 (2019), pp. 249–259.

[33] Ramanujan Srinath, Douglas A Ruff, and Marlene R Cohen. "Attention improves information flow between neuronal populations without changing the communication subspace". In: *Current Biology* 31.23 (2021), pp. 5299–5313.

[34] Ryan A Young et al. "Hippocampal-prefrontal communication subspaces align with behavioral and network patterns in a spatial memory task". In: *bioRxiv* (2024), pp. 2024–07.

[35] CJ MacDowell et al. *Multiplexed Subspaces Route Neural Activity Across Brain-wide Networks. bioRxiv, 2023.02. 08.527772*. 2023.

[36] Peter G Poggi, Stefan Mihalas, and Dana Mastrovito. "Communication subspaces align with training in ANNs". In: *bioRxiv* (2024), pp. 2024–11.

[37] Kendrick N Kay et al. "A two-stage cascade model of BOLD responses in human visual cortex". In: *PLoS computational biology* 9.5 (2013), e1003079.

[38] Narges Doostani, Gholam-Ali Hossein-Zadeh, and Maryam Vaziri-Pashkam. "The normalization model predicts responses in the human visual cortex during object-based attention". In: *Elife* 12 (2023), e75726.

[39] Shawn R Olsen, Vikas Bhandawat, and Rachel I Wilson. "Divisive normalization in olfactory population codes". In: *Neuron* 66.2 (2010), pp. 287–299.

[40] Misha B Ahrens, Jennifer F Linden, and Maneesh Sahani. "Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods". In: *Journal of Neuroscience* 28.8 (2008), pp. 1929–1942.

[41] Neil C Rabinowitz et al. "Contrast gain control in auditory cortex". In: *Neuron* 70.6 (2011), pp. 1178–1191.

[42] Gijs Joost Brouwer et al. "Normalization in human somatosensory cortex". In: *Journal of Neurophysiology* 114.5 (2015), pp. 2588–2599.

[43] Rachel N Denison, Marisa Carrasco, and David J Heeger. "A dynamic normalization model of temporal attention". In: *Nature Human Behaviour* 5.12 (2021), pp. 1674–1685.

[44] Amy M Ni and John HR Maunsell. "Neuronal effects of spatial and feature attention differ due to normalization". In: *Journal of Neuroscience* 39.28 (2019), pp. 5493–5505.

[45] Xilin Zhang et al. "A normalization framework for emotional attention". In: *PLoS biology* 14.11 (2016), e1002578.

[46] Philipp Schwedhelm, B Suresh Krishna, and Stefan Treue. "An extended normalization model of attention accounts for feature-based attentional enhancement of both response and coherence gain". In: *PLoS computational biology* 12.12 (2016), e1005225.

[47] Philip L Smith, David K Sewell, and Simon D Lilburn. "From shunting inhibition to dynamic normalization: Attentional selection and decision-making in brief visual displays". In: *Vision Research* 116 (2015), pp. 219–240.

[48] John HR Maunsell. "Neuronal mechanisms of visual attention". In: *Annual review of vision science* 1.1 (2015), pp. 373–391.

[49] Frederik Beuth and Fred H Hamker. "A mechanistic cortical microcircuit of attention for amplification, normalization and suppression". In: *Vision research* 116 (2015), pp. 241–257.

[50] Katrin Herrmann, David J Heeger, and Marisa Carrasco. "Feature-based attention enhances performance by increasing response gain". In: *Vision research* 74 (2012), pp. 10–20.

[51] Amy M Ni and John HR Maunsell. "Spatially tuned normalization explains attention modulation variance within neurons". In: *Journal of neurophysiology* 118.3 (2017), pp. 1903–1913.

[52] Katrin Herrmann et al. "When size matters: attention affects performance by contrast or response gain". In: *Nature neuroscience* 13.12 (2010), pp. 1554–1559.

[53] John H Reynolds and David J Heeger. "The normalization model of attention". In: *Neuron* 61.2 (2009), pp. 168–185.

[54] Joonyeol Lee and John HR Maunsell. "A normalization model of attentional modulation of single unit responses". In: *PloS one* 4.2 (2009), e4651.

[55] Geoffrey M Boynton. "A framework for describing the effects of attention on visual responses". In: *Vision research* 49.10 (2009), pp. 1129–1143.

[56] Tadashi Sugihara et al. "Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex". In: *Journal of Neuroscience* 26.43 (2006), pp. 11138–11147.

[57] Marie Avillac, Suliann Ben Hamed, and Jean-René Duhamel. "Multisensory integration in the ventral intraparietal area of the macaque monkey". In: *Journal of Neuroscience* 27.8 (2007), pp. 1922–1932.

[58] Tomokazu Ohshiro, Dora E Angelaki, and Gregory C DeAngelis. "A normalization model of multisensory integration". In: *Nature neuroscience* 14.6 (2011), pp. 775–782.

[59] Tomokazu Ohshiro, Dora E Angelaki, and Gregory C DeAngelis. "A neural signature of divisive normalization at the level of multisensory integration in primate cortex". In: *Neuron* 95.2 (2017), pp. 399–411.

[60] Maria M Diehl and Lizabeth M Romanski. "Responses of prefrontal multisensory neurons to mismatching faces and vocalizations". In: *Journal of Neuroscience* 34.34 (2014), pp. 11233–11243.

[61] Robert Taylor and Paul M Bays. "Efficient coding in visual working memory accounts for stimulus-specific variations in recall". In: *Journal of Neuroscience* 38.32 (2018), pp. 7132–7142.

[62] Sebastian Schneegans and Paul M Bays. "Drift in neural population activity causes working memory to deteriorate over time". In: *Journal of Neuroscience* 38.21 (2018), pp. 4859–4869.

[63] Wei Ji Ma, Masud Husain, and Paul M Bays. "Changing concepts of working memory". In: *Nature neuroscience* 17.3 (2014), pp. 347–356.

[64] Paul M Bays. "Noise in neural populations accounts for errors in working memory". In: *Journal of Neuroscience* 34.10 (2014), pp. 3632–3645.

[65] Kenway Louie, Lauren E Grattan, and Paul W Glimcher. "Reward value-based gain control: divisive normalization in parietal cortex". In: *Journal of Neuroscience* 31.29 (2011), pp. 10627–10639.

[66] Kenway Louie, Mel W Khaw, and Paul W Glimcher. "Normalization is a general neural mechanism for context-dependent decision making". In: *Proceedings of the National Academy of Sciences* 110.15 (2013), pp. 6139–6144.

[67] Kenway Louie et al. "Dynamic divisive normalization predicts time-varying value coding in decision-related circuits". In: *Journal of Neuroscience* 34.48 (2014), pp. 16046–16057.

[68] Mel W Khaw, Paul W Glimcher, and Kenway Louie. "Normalized value coding explains dynamic adaptation in the human valuation process". In: *Proceedings of the National Academy of Sciences* 114.48 (2017), pp. 12696–12701.

[69] David J Heeger, Eero P Simoncelli, and J Anthony Movshon. "Computational models of cortical visual processing." In: *Proceedings of the National Academy of Sciences* 93.2 (1996), pp. 623–627.

[70] Eero P Simoncelli and David J Heeger. "A model of neuronal responses in visual area MT". In: *Vision research* 38.5 (1998), pp. 743–761.

[71] Eero P Simoncelli. "Local analysis of visual motion". In: *The visual neurosciences* 109 (2003), pp. 1616–1623.

[72] Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. "Marginalization in neural circuits with divisive normalization". In: *Journal of Neuroscience* 31.43 (2011), pp. 15310–15319.

[73] Dario L Ringach. "Population coding under normalization". In: *Vision research* 50.22 (2010), pp. 2223–2232.

[74] Odelia Schwartz and Eero P Simoncelli. "Natural signal statistics and sensory gain control". In: *Nature neuroscience* 4.8 (2001), pp. 819–825.

[75] Zachary M Westrick, David J Heeger, and Michael S Landy. "Pattern adaptation and normalization reweighting". In: *Journal of Neuroscience* 36.38 (2016), pp. 9805–9816.

[76] Siwei Lyu and Eero P Simoncelli. "Nonlinear extraction of independent components of natural images using radial gaussianization". In: *Neural computation* 21.6 (2009), pp. 1485–1519.

[77] Shawn X Luo, Richard Axel, and Larry F Abbott. "Generating sparse and selective third-order responses in the olfactory system of the fly". In: *Proceedings of the National Academy of Sciences* 107.23 (2010), pp. 10713–10718.

[78] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. "Layer normalization". In: *arXiv preprint arXiv:1607.06450* (2016).

[79] Mengye Ren et al. "Normalizing the normalizers: Comparing and extending network normalization schemes". In: *arXiv preprint arXiv:1611.04520* (2016).

[80] Kevin A Bolding and Kevin M Franks. "Recurrent cortical circuits implement concentration-invariant odor coding". In: *Science* 361.6407 (2018), eaat6904.

[81] Hillel Adesnik and Massimo Scanziani. "Lateral competition for cortical space by layer-specific horizontal circuits". In: *Nature* 464.7292 (2010), pp. 1155–1160.

[82] Tatsuo K Sato et al. "An excitatory basis for divisive normalization in visual cortex". In: *Nature neuroscience* 19.4 (2016), pp. 568–570.

[83] Hillel Adesnik. "Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice". In: *Neuron* 95.5 (2017), pp. 1147–1159.

[84] Rodney J Douglas et al. "Recurrent excitation in neocortical circuits". In: *Science* 269.5226 (1995), pp. 981–985.

[85] Christoph Kapfer et al. "Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex". In: *Nature neuroscience* 10.6 (2007), pp. 743–753.

[86] Edward M Callaway. "Local circuits in primary visual cortex of the macaque monkey". In: *Annual review of neuroscience* 21.1 (1998), pp. 47–74.

[87] Yumiko Yoshimura, Jami L Dantzker, and Edward M Callaway. "Excitatory cortical neurons form fine-scale functional networks". In: *Nature* 433.7028 (2005), pp. 868–873.

[88] Rodney J Douglas and Kevan AC Martin. "Recurrent neuronal circuits in the neocortex". In: *Current biology* 17.13 (2007), R496–R500.

[89] Simon Peron et al. "Recurrent interactions in local cortical circuits". In: *Nature* 579.7798 (2020), pp. 256–259.

[90] Jonathan B Levitt, Daniel C Kiper, and J Anthony Movshon. "Receptive fields and functional architecture of macaque V2". In: *Journal of neurophysiology* 71.6 (1994), pp. 2517–2542.

[91] Wilson S Geisler and Duane G Albrecht. "Visual cortex neurons in monkeys and cats: detection, discrimination, and identification". In: *Visual neuroscience* 14.5 (1997), pp. 897–919.

[92] Gary Sclar, John HR Maunsell, and Peter Lennie. "Coding of image contrast in central visual pathways of the macaque monkey". In: *Vision research* 30.1 (1990), pp. 1–10.

[93] Michael J Kahana. "The cognitive correlates of human brain oscillations". In: *Journal of Neuroscience* 26.6 (2006), pp. 1669–1672.

[94] Gyorgy Buzsaki and Andreas Draguhn. "Neuronal oscillations in cortical networks". In: *science* 304.5679 (2004), pp. 1926–1929.

[95] Markus Siegel, Tobias H Donner, and Andreas K Engel. "Spectral fingerprints of large-scale neuronal interactions". In: *Nature Reviews Neuroscience* 13.2 (2012), pp. 121–134.

[96] Peter Uhlhaas et al. "Neural synchrony in cortical networks: history, concept and current status". In: *Frontiers in integrative neuroscience* 3 (2009), p. 543.

[97] György Buzsáki, Costas A Anastassiou, and Christof Koch. "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes". In: *Nature reviews neuroscience* 13.6 (2012), pp. 407–420.

[98] Claude Bédard, Helmut Kröger, and Alain Destexhe. "Model of low-pass filtering of local field potentials in brain tissue". In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 73.5 (2006), p. 051911.

[99] Gaute T Einevoll et al. "Modelling and analysis of local field potentials for studying the function of cortical circuits". In: *Nature Reviews Neuroscience* 14.11 (2013), pp. 770–785.

[100] Xiaoxuan Jia, Dajun Xing, and Adam Kohn. "No consistent relationship between gamma power and peak frequency in macaque primary visual cortex". In: *Journal of Neuroscience* 33.1 (2013), pp. 17–25.

[101] Supratim Ray and John HR Maunsell. "Differences in gamma frequencies across visual cortex restrict their possible use in computation". In: *Neuron* 67.5 (2010), pp. 885–896.

[102] J Andrew Henrie and Robert Shapley. "LFP power spectra in V1 cortex: the graded effect of stimulus contrast". In: *Journal of neurophysiology* 94.1 (2005), pp. 479–490.

[103] Kai J Miller et al. "Power-law scaling in the brain surface electric potential". In: *PLoS computational biology* 5.12 (2009), e1000609.

[104] Andreas K Engel, Pascal Fries, and Wolf Singer. "Dynamic predictions: oscillations and synchrony in top–down processing". In: *Nature Reviews Neuroscience* 2.10 (2001), pp. 704–716.

[105] Pascal Fries. "Rhythms for cognition: communication through coherence". In: *Neuron* 88.1 (2015), pp. 220–235.

[106] R Matthew Hutchison et al. "Dynamic functional connectivity: promise, issues, and interpretations". In: *Neuroimage* 80 (2013), pp. 360–378.

[107] Elena A Allen et al. "Tracking whole-brain connectivity dynamics in the resting state". In: *Cerebral cortex* 24.3 (2014), pp. 663–676.

[108] Vince D Calhoun et al. "The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery". In: *Neuron* 84.2 (2014), pp. 262–274.

[109] Sang-Han Choi et al. "Proposal for human visual pathway in the extrastriate cortex by fiber tracking method using diffusion-weighted MRI". In: *Neuroimage* 220 (2020), p. 117145.

[110] Mortimer Mishkin, Leslie G Ungerleider, and Kathleen A Macko. "Object vision and spatial vision: two cortical pathways". In: *Trends in neurosciences* 6 (1983), pp. 414–417.

[111] Melvyn A Goodale and A David Milner. "Separate visual pathways for perception and action". In: *Trends in neurosciences* 15.1 (1992), pp. 20–25.

[112] Maurizio Corbetta et al. "Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography". In: *Journal of neuroscience* 11.8 (1991), pp. 2383–2402.

[113] John T Serences and Geoffrey M Boynton. "Feature-based attentional modulations in the absence of direct visual stimulation". In: *Neuron* 55.2 (2007), pp. 301–312.

[114] John HR Maunsell and Stefan Treue. "Feature-based attention in visual cortex". In: *Trends in neurosciences* 29.6 (2006), pp. 317–322.

[115] Stefan Treue and Julio C Martinez Trujillo. "Feature-based attention influences motion processing gain in macaque visual cortex". In: *Nature* 399.6736 (1999), pp. 575–579.

[116] Marlene R Cohen and John HR Maunsell. "Using neuronal populations to study the mechanisms underlying spatial and feature attention". In: *Neuron* 70.6 (2011), pp. 1192–1204.

[117] Taosheng Liu et al. "Cortical mechanisms of feature-based attentional control". In: *Cerebral cortex* 13.12 (2003), pp. 1334–1343.

[118] Julio C Martinez-Trujillo and Stefan Treue. "Feature-based attention increases the selectivity of population responses in primate visual cortex". In: *Current biology* 14.9 (2004), pp. 744–751.

[119] Sameer Saproo and John T Serences. "Attention improves transfer of motion information between V1 and MT". In: *Journal of Neuroscience* 34.10 (2014), pp. 3586–3596.

[120] Naseem Al-Aidroos, Christopher P Said, and Nicholas B Turk-Browne. "Top-down attention switches coupling between low-level and high-level areas of human visual cortex". In: *Proceedings of the National Academy of Sciences* 109.36 (2012), pp. 14675–14680.

[121] Yuri B Saalmann, Ivan N Pigarev, and Trichur R Vidyasagar. "Neural mechanisms of visual attention: how top-down feedback highlights relevant locations". In: *Science* 316.5831 (2007), pp. 1612–1615.

[122] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[123] Alex Graves. "Generating sequences with recurrent neural networks". In: *arXiv preprint arXiv:1308.0850* (2013).

[124] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks". In: *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee. 2013, pp. 6645–6649.

[125] Kyunghyun Cho et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation". In: *arXiv preprint arXiv:1406.1078* (2014).

[126] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to sequence learning with neural networks". In: *Advances in neural information processing systems* 27 (2014).

[127] Rajesh PN Rao and Dana H Ballard. "Dynamic model of visual recognition predicts neural response properties in the visual cortex". In: *Neural computation* 9.4 (1997), pp. 721–763.

[128] Catherine Wacongne, Jean-Pierre Changeux, and Stanislas Dehaene. "A neuronal model of predictive coding accounting for the mismatch negativity". In: *Journal of Neuroscience* 32.11 (2012), pp. 3665–3678.

[129] Rajesh PN Rao and Dana H Ballard. "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects". In: *Nature neuroscience* 2.1 (1999), pp. 79–87.

[130] Tai Sing Lee and David Mumford. "Hierarchical Bayesian inference in the visual cortex". In: *Journal of the Optical Society of America A* 20.7 (2003), pp. 1434–1448.

[131] Karl Friston. "A theory of cortical responses". In: *Philosophical transactions of the Royal Society B: Biological sciences* 360.1456 (2005), pp. 815–836.

[132] Andre M Bastos et al. "Canonical microcircuits for predictive coding". In: *Neuron* 76.4 (2012), pp. 695–711.

[133] Samuel P Burns, Dajun Xing, and Robert M Shapley. "Comparisons of the dynamics of local field potential and multiunit activity signals in macaque visual cortex". In: *Journal of Neuroscience* 30.41 (2010), pp. 13739–13749.

[134] Hirofumi Ozeki et al. "Inhibitory stabilization of the cortical network underlies visual surround suppression". In: *Neuron* 62.4 (2009), pp. 578–592.

[135] Walter Senn et al. "A neuronal least-action principle for real-time learning in cortical circuits". In: *ELife* 12 (2024), RP89674.

[136] David J Heeger and James R Bergen. "Pyramid-based texture analysis/synthesis". In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995, pp. 229–238.

[137] Javier Portilla and Eero P Simoncelli. "A parametric texture model based on joint statistics of complex wavelet coefficients". In: *International journal of computer vision* 40 (2000), pp. 49–70.

[138] Geoffrey E Hinton et al. "The" wake-sleep" algorithm for unsupervised neural networks". In: *Science* 268.5214 (1995), pp. 1158–1161.

[139] David J Heeger. "Optical flow using spatiotemporal filters". In: *International journal of computer vision* 1 (1988), pp. 279–302.

[140] David J Heeger and Eero P Simoncelli. "Model of visual motion sensing". In: *Spatial vision in humans and robots* 19 (1993), pp. 367–392.

[141] Wei Ji Ma et al. "Bayesian inference with probabilistic population codes". In: *Nature neuroscience* 9.11 (2006), pp. 1432–1438.

[142] David C Knill and Alexandre Pouget. "The Bayesian brain: the role of uncertainty in neural coding and computation". In: *TRENDS in Neurosciences* 27.12 (2004), pp. 712–719.

[143] Michael S Landy et al. "Measurement and modeling of depth cue combination: in defense of weak fusion". In: *Vision research* 35.3 (1995), pp. 389–412.

[144] Marc O Ernst and Martin S Banks. "Humans integrate visual and haptic information in a statistically optimal fashion". In: *Nature* 415.6870 (2002), pp. 429–433.

[145] Christopher R Fetsch et al. "Neural correlates of reliability-based cue weighting during multisensory integration". In: *Nature neuroscience* 15.1 (2012), pp. 146–154.

[146] Christopher J Honey et al. "Predicting human resting-state functional connectivity from structural connectivity". In: *Proceedings of the National Academy of Sciences* 106.6 (2009), pp. 2035–2040.

[147] André M Bastos and Jan-Mathijs Schoffelen. "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls". In: *Frontiers in systems neuroscience* 9 (2016), p. 175.

[148] Melissa Saenz, Giedrius T Buracas, and Geoffrey M Boynton. "Global effects of feature-based attention in human visual cortex". In: *Nature neuroscience* 5.7 (2002), pp. 631–632.

[149] Edward G Jones. "Thalamic circuitry and thalamocortical synchrony". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357.1428 (2002), pp. 1659–1673.

[150] Stuart W Hughes and Vincenzo Crunelli. "Thalamic mechanisms of EEG alpha rhythms and their pathological implications". In: *The Neuroscientist* 11.4 (2005), pp. 357–372.

[151] Anil Bollimunta et al. "Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations". In: *Journal of Neuroscience* 31.13 (2011), pp. 4935–4943.

[152] Mila Halgren et al. "The generation and propagation of the human alpha rhythm". In: *Proceedings of the National Academy of Sciences* 116.47 (2019), pp. 23772–23782.

[153] Ole Jensen and Ali Mazaheri. "Shaping functional architecture by oscillatory alpha activity: gating by inhibition". In: *Frontiers in human neuroscience* 4 (2010), p. 186.

[154] Wolfgang Klimesch, Paul Sauseng, and Simon Hanslmayr. "EEG alpha oscillations: the inhibition–timing hypothesis". In: *Brain research reviews* 53.1 (2007), pp. 63–88.

[155] L Ian Schmitt et al. "Thalamic amplification of cortical connectivity sustains attentional control". In: *Nature* 545.7653 (2017), pp. 219–223.

[156] Alain Destexhe. "Modelling corticothalamic feedback and the gating of the thalamus by the cerebral cortex". In: *Journal of Physiology-Paris* 94.5-6 (2000), pp. 391–410.

[157] S Murray Sherman and W Martin Usrey. "Exploring Thalamocortical Interactions: Circuitry for Sensation, Action, and Cognition". In: (2021).

[158] Rajeev V Rikhye, Aditya Gilra, and Michael M Halassa. "Thalamic regulation of switching between cortical representations enables cognitive flexibility". In: *Nature neuroscience* 21.12 (2018), pp. 1753–1763.

[159] Rajeev V Rikhye, Ralf D Wimmer, and Michael M Halassa. "Toward an integrative theory of thalamic function". In: *Annual review of neuroscience* 41.1 (2018), pp. 163–183.

[160] Zengcai V Guo et al. "Maintenance of persistent activity in a frontal thalamocortical loop". In: *Nature* 545.7653 (2017), pp. 181–186.

[161] Matthew T Schmolesky et al. "Signal timing across the macaque visual system". In: *Journal of neurophysiology* 79.6 (1998), pp. 3272–3278.

[162] João D Semedo et al. "Feedforward and feedback interactions between visual cortical areas use different population activity patterns". In: *Nature communications* 13.1 (2022), p. 1099.

[163] Matteo Carandini and David J Heeger. "Summation and division by neurons in primate visual cortex". In: *Science* 264.5163 (1994), pp. 1333–1336.

[164] Frances S Chance, Larry F Abbott, and Alex D Reyes. "Gain modulation from background synaptic input". In: *Neuron* 35.4 (2002), pp. 773–782.

[165] Hugh R Wilson and Jack D Cowan. "Excitatory and inhibitory interactions in localized populations of model neurons". In: *Biophysical journal* 12.1 (1972), pp. 1–24.

[166] Therese Riedemann. "Diversity and function of somatostatin-expressing interneurons in the cerebral cortex". In: *International journal of molecular sciences* 20.12 (2019), p. 2952.

[167] Hillel Adesnik et al. "A neural circuit for spatial summation in visual cortex". In: *Nature* 490.7419 (2012), pp. 226–231.

[168] Carsten K Pfeffer et al. "Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons". In: *Nature neuroscience* 16.8 (2013), pp. 1068–1076.

[169] Luke Campagnola et al. "Local connectivity and synaptic dynamics in mouse and human neocortex". In: *Science* 375.6585 (2022), eabj5861.

[170] Jung Hoon Lee and Stefan Mihalas. "Visual processing mode switching regulated by VIP cells". In: *Scientific Reports* 7.1 (2017), p. 1843.

[171] Bassam V Atallah et al. "Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli". In: *Neuron* 73.1 (2012), pp. 159–170.

[172] Alessandro Sanzeni et al. "Inhibition stabilization is a widespread property of cortical networks". In: *Elife* 9 (2020), e54875.

[173] Florian Studer, Tommaso Zeppillo, and Tania Rinaldi Barkat. "Top-down inputs are controlled by somatostatin-expressing interneurons during associative learning". In: *bioRxiv* (2024), pp. 2024–12.

[174] Alexander Naka et al. "Complementary networks of cortical somatostatin interneurons enforce layer specific control". In: *Elife* 8 (2019), e43696.

[175] Shan Shen et al. "Distinct organization of two cortico-cortical feedback pathways". In: *Nature communications* 13.1 (2022), p. 6389.

[176] Alessandra Angelucci, Jonathan B Levitt, and Jennifer S Lund. "Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1". In: *Progress in brain research* 136 (2002), pp. 373–388.

[177] Alessandra Angelucci et al. "Circuits for local and global signal integration in primary visual cortex". In: *Journal of Neuroscience* 22.19 (2002), pp. 8633–8646.

[178] Jonathan J Nassi, Stephen G Lomber, and Richard T Born. "Corticocortical feedback contributes to surround suppression in V1 of the alert primate". In: *Journal of Neuroscience* 33.19 (2013), pp. 8504–8517.

[179] Jonathan J Nassi et al. "Corticocortical feedback increases the spatial extent of normalization". In: *Frontiers in systems neuroscience* 8 (2014), p. 105.

[180] J Julius Zhu. "Maturation of layer 5 neocortical pyramidal neurons: amplifying salient layer 1 and layer 4 inputs by Ca2+ action potentials in adult rat tuft dendrites". In: *The Journal of physiology* 526.3 (2000), pp. 571–587.

[181] Martin Sarter and John P Bruno. "Cognitive functions of cortical acetylcholine: toward a unifying hypothesis". In: *Brain Research Reviews* 23.1-2 (1997), pp. 28–46.

[182] J Yu Angela and Peter Dayan. "Uncertainty, neuromodulation, and attention". In: *Neuron* 46.4 (2005), pp. 681–692.

[183] Louise Marshall et al. "Pharmacological fingerprints of contextual uncertainty". In: *PLoS Biology* 14.11 (2016), e1002575.

[184] Diego E Pafundo et al. "Top-down-mediated facilitation in the visual cortex is gated by subcortical neuromodulation". In: *Journal of Neuroscience* 36.10 (2016), pp. 2904–2914.

[185] John M Bekkers. "Pyramidal neurons". In: *Current biology* 21.24 (2011), R975.

[186] Shivang Rawat, David Heeger, and Stefano Martiniani. "A Comprehensive Large-Scale Model of the Primary Visual Cortex (V1)". In: *Bulletin of the American Physical Society* (2024).

[187] Matteo Carandini. "Amplification of trial-to-trial response variability by neurons in visual cortex". In: *PLoS biology* 2.9 (2004), e264.

[188] Shivang Rawat and Stefano Martiniani. "Element-wise and recursive solutions for the power spectral density of biological stochastic dynamical systems at fixed points". In: *Physical Review Research* 6.4 (2024), p. 043179.

[189] Adam Kohn et al. "Principles of corticocortical communication: proposed schemes and design considerations". In: *Trends in Neurosciences* 43.9 (2020), pp. 725–737.

[190] João D Semedo et al. "Statistical methods for dissecting interactions between brain areas". In: *Current opinion in neurobiology* 65 (2020), pp. 59–69.

[191] David J Heeger. "Half-squaring in responses of cat striate cells". In: *Visual neuroscience* 9.5 (1992), pp. 427–443.

[192] Christina Allen and Charles F Stevens. "An evaluation of causes for unreliability of synaptic transmission." In: *Proceedings of the National Academy of Sciences* 91.22 (1994), pp. 10380–10383.

[193] Zachary F Mainen and Terrence J Sejnowski. "Reliability of spike timing in neocortical neurons". In: *Science* 268.5216 (1995), pp. 1503–1506.

[194] Matthew J Nelson and Pierre Pouget. *Do electrode properties create a problem in interpreting local field potential recordings?* 2010.

[195] Kip A Ludwig et al. "Using a common average reference to improve cortical neuron recordings from microelectrode arrays". In: *Journal of neurophysiology* 101.3 (2009), pp. 1679–1689.

[196] Nathan W Whitmore and Shih-Chieh Lin. "Unmasking local activity within local field potentials (LFPs) by removing distal electrical signals using independent component analysis". In: *Neuroimage* 132 (2016), pp. 79–92.

Appendix

Derivation of the General Hierarchical Recurrent Normalization Circuit

This appendix provides a detailed derivation of the general dynamical equations for the hierarchical recurrent normalization model with reciprocal connections. We begin with the primary visual cortex (V1). Assuming $\mathbf{N}_1$ is the normalization weight matrix for V1 and $\mathbf{z}_1$ is the input drive to V1, we can write the normalization equations for the V1 principal neurons with complementary receptive fields as:

$$\mathbf{y}_1^{s+} = \frac{\lfloor \mathbf{z}_1 \rfloor^n}{\sigma^n + \mathbf{N}_1(\lfloor \mathbf{z}_1 \rfloor^n + \lfloor -\mathbf{z}_1 \rfloor^n)} \qquad \mathbf{y}_1^{s-} = \frac{\lfloor -\mathbf{z}_1 \rfloor^n}{\sigma^n + \mathbf{N}_1(\lfloor \mathbf{z}_1 \rfloor^n + \lfloor -\mathbf{z}_1 \rfloor^n)} \tag{30}$$

Where $\mathbf{y}_1^{s+}$ and $\mathbf{y}_1^{s-}$ represent the steady-state firing rates of principal neurons corresponding to positive ($\lfloor \mathbf{z}_1 \rfloor^n$) and negative $\lfloor -\mathbf{z}_1 \rfloor^n$ input drives, respectively. Here, σ represents the semi-saturation constant. The exponent n is typically observed to be $n \approx 2$ for the primary visual cortex, it can be different for different brain areas. Mathematically:

$$\lfloor \mathbf{z}_1 \rfloor^n + \lfloor -\mathbf{z}_1 \rfloor^n = |\mathbf{z}_1|^n \tag{31}$$

Our goal is to derive the dynamical equations for a general n, that yield the fixed points defined by Eq. 30. First, we distinguish between the membrane potentials and the firing rates of neurons. The membrane potential of a given type of neuron is denoted by vectors such as $\mathbf{y}_1$, $\mathbf{u}_1$, $\mathbf{a}_1$, and $\mathbf{q}_1$. Their corresponding firing rates are $\mathbf{y}_1^+$ (and $\mathbf{y}_1^-$ for complementary responses), $\mathbf{u}_1^+$, $\mathbf{a}_1^+$, and $\mathbf{q}_1^+$. The instantaneous firing rates are related to membrane potentials by applying rectification (denoted by $\lfloor \cdot \rfloor$) followed by power-law activation [191],

$$\mathbf{x}^+ = \lfloor \mathbf{x} \rfloor^\beta. \tag{32}$$

The values of β are taken to be $\{2, \frac{1}{2}, 1, 1\}$ for $\{\mathbf{y}_1, \mathbf{u}_1, \mathbf{a}_1, \mathbf{q}_1\}$, respectively. Combining the firing rates of principal neurons with the complementary receptive fields, $\mathbf{y}_1^{s+}$ and $\mathbf{y}_1^{s-}$, Eq. 30 can be alternatively written as,

$$|\mathbf{y}_1|^n = \mathbf{y}_1^{s+} + \mathbf{y}_1^{s-} = \frac{|\mathbf{z}_1|^n}{\sigma^n + \mathbf{N}_1 |\mathbf{z}_1|^n} \tag{33}$$

To proceed with our derivation, we need to isolate the $|\mathbf{z}_1|^n$ term. From Eq. 33, we have:

$$|\mathbf{z}_1|^n = \sigma^n \odot |\mathbf{y}_1|^n + |\mathbf{y}_1|^n \odot (\mathbf{N}_1 |\mathbf{z}_1|^n) \tag{34}$$

where $\odot$ denotes element-wise multiplication. Using the identity $\mathbf{x}_1 \odot \mathbf{x}_2 = \mathrm{D}(\mathbf{x}_1)\mathbf{x}_2$, where $\mathrm{D}(\mathbf{x}_1)$ is a diagonal matrix with the elements of vector $\mathbf{x}_1$ on its diagonal, we can rewrite and solve for $|\mathbf{z}_1|^n$:

$$|\mathbf{z}_1|^n = \sigma^n \odot |\mathbf{y}_1|^n + \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1 |\mathbf{z}_1|^n \tag{35}$$

Factoring out $|\mathbf{z}_1|^n$:

$$\left[\mathbf{I} - \mathrm{D}\left(|\mathbf{y}_1|^n\right) \mathbf{N}_1\right] |\mathbf{z}_1|^n = \sigma^n \odot |\mathbf{y}_1|^n \tag{36}$$

Multiplying both sides with $[\mathbf{I} - \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1]^{-1}$ from the left yields:

$$|\mathbf{z}_1|^n = \left[\mathbf{I} - \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1\right]^{-1} \left(\sigma^n \odot |\mathbf{y}_1|^n\right) \tag{37}$$

This expression for $|\mathbf{z}_1|^n$ will be used in subsequent steps. Now, we introduce the dynamical equation for the membrane potential $\mathbf{y}_1$ of the principal V1 neurons, consistent with biological computation in pyramidal cells (see Section 'Biology plausibility'):

$$\tau_{y_1} \frac{\mathrm{d}\mathbf{y}_1}{\mathrm{d}t} = -\mathbf{y}_1 + \mathbf{b}_1 \odot \mathbf{z}_1 + \frac{1}{1 + \mathbf{a}_1^+} \odot \left(\mathbf{W}_{11} \left((\mathbf{y}_1^+)^{\frac{1}{n}} - (\mathbf{y}_1^-)^{\frac{1}{n}}\right) + \gamma_1 \odot \mathbf{W}_{12}(\mathbf{y}_2^+)^{\frac{1}{n}}\right). \tag{38}$$

Here, τ_{y_1} is the time constant, $\mathbf{b}_1$ is the input gain, $\mathbf{a}_1^+$ represents firing rates of the inhibitory neuron population. $\mathbf{W}_{11}$ is the recurrent V1 connectivity matrix, $\mathbf{W}_{12}$ is the connectivity matrix containing the synaptic weights from V2 to V1. $\mathbf{W}_{12}(\mathbf{y}_2^+)^{\frac{1}{n}}$ represents the feedback signal from V2 to V1, calculated as the weighted sum of the activities of V2 neurons. γ_1 is the feedback gain modulation parameter. $\mathbf{y}_2^+$ is the firing rate of V2 principal neurons, defined as $\lfloor \mathbf{y}_2 \rfloor^n$.

For simplicity, we assume that the recurrent weight matrix $\mathbf{W}_{11} = \mathrm{I}$. Given that $(\mathbf{y}_1^+)^{1/p} - (\mathbf{y}_1^-)^{1/p} = \lfloor \mathbf{y}_1 \rfloor - \lfloor -\mathbf{y}_1 \rfloor = \mathbf{y}_1$ and $(\mathbf{y}_2^+)^{1/n} = \lfloor \mathbf{y}_2 \rfloor$, the equation for $\mathbf{y}_1$ simplifies to:

$$\tau_{y_1} \frac{\mathrm{d}\mathbf{y}_1}{\mathrm{d}t} = -\mathbf{y}_1 + \mathbf{b}_1 \odot \mathbf{z}_1 + \frac{1}{1+\mathbf{a}_1^+} \odot (\mathbf{y}_1 + \gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor) \tag{39}$$

Now at steady state, $\frac{\mathrm{d}\mathbf{y}_1}{\mathrm{d}t} = 0$,

$$0 = -\mathbf{y}_1 + \mathbf{b}_1 \odot \mathbf{z}_1 + \frac{1}{1+\mathbf{a}_1^+} \odot (\mathbf{y}_1 + \gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor) \tag{40}$$

Expanding the terms:

$$0 = -\mathbf{y}_1 + \mathbf{b}_1 \odot \mathbf{z}_1 + \frac{1}{1+\mathbf{a}_1^+} \odot \mathbf{y}_1 + \frac{1}{1+\mathbf{a}_1^+} \odot (\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor) \tag{41}$$

Rearranging terms:

$$\frac{\mathbf{a}_1^+}{1+\mathbf{a}_1^+} \odot \mathbf{y}_1 = \mathbf{b}_1 \odot \mathbf{z}_1 + \frac{1}{1+\mathbf{a}_1^+} \odot (\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor) \tag{42}$$

Moving the feedback term to the left-hand side:

$$\frac{\mathbf{a}_1^+}{1+\mathbf{a}_1^+} \odot \mathbf{y}_1 - \frac{1}{1+\mathbf{a}_1^+} \odot (\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor) = \mathbf{b}_1 \odot \mathbf{z}_1 \tag{43}$$

Factoring out $\mathbf{y}_1$:

$$\mathbf{y}_1 \odot \left(\frac{\mathbf{a}_1^+}{1+\mathbf{a}_1^+} - \frac{1}{1+\mathbf{a}_1^+} \odot \frac{\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor}{\mathbf{y}_1} \right) = \mathbf{b}_1 \odot \mathbf{z}_1 \tag{44}$$

We define a modulatory term $\mathbf{u}_1^+$, such that:

$$\mathbf{y}_1 \odot \mathbf{u}_1^+ = \mathbf{b}_1 \odot \mathbf{z}_1 \tag{45}$$

where:

$$\mathbf{u}_1^+ = \frac{\mathbf{a}_1^+}{1+\mathbf{a}_1^+} - \frac{1}{1+\mathbf{a}_1^+} \odot \frac{\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor}{\mathbf{y}_1} \tag{46}$$

Next, we derive the dynamical equation for $\mathbf{a}_1$, from the definition of $\mathbf{u}_1^+$:

$$\mathbf{u}_1^+ \odot (1+\mathbf{a}_1^+) = \mathbf{a}_1^+ - \frac{\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor}{\mathbf{y}_1} \tag{47}$$

Upon rearranging,

$$0 = -\mathbf{a}_1^+ + \frac{\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor}{\mathbf{y}_1} + (1+\mathbf{a}_1^+) \odot \mathbf{u}_1^+ \tag{48}$$

This equation describes the steady state of $\mathbf{a}_1$. It exhibits a form analogous to that of $\mathbf{y}_1$, comprising an input drive and a recurrent drive term. Therefore, we propose the following dynamical equation for the membrane potential $\mathbf{a}_1$:

$$\tau_{a_1} \frac{\mathrm{d}\mathbf{a}_1}{\mathrm{d}t} = -\mathbf{a}_1 + \frac{\gamma_1 \odot \mathbf{W}_{12} \lfloor \mathbf{y}_2 \rfloor}{\mathbf{y}_1} + (1+\mathbf{a}_1^+) \odot \mathbf{u}_1^+ \tag{49}$$

According to Dale's Law, excitatory neurons, such as $\mathbf{y}_1$, cannot directly inhibit other neurons. Therefore, the division by $\mathbf{y}_1$ implies the existence of an intermediate inhibitory interneuron. We introduce an inhibitory interneuron population $\mathbf{q}_1$, that tracks $\mathbf{y}_1$:

$$\tau_{q_1} \frac{\mathrm{d}\mathbf{q}_1}{\mathrm{d}t} = -\mathbf{q}_1 + (\mathbf{y}_1^+)^{1/n} \tag{50}$$

With this interneuron, the dynamic equation for $\mathbf{a}_1$ becomes:

$$\tau_{a_1} \frac{\mathrm{d}\mathbf{a}_1}{\mathrm{d}t} = -\mathbf{a}_1 + \frac{\gamma_1 \odot \mathbf{W}_{12} (\mathbf{y}_2^+)^{1/n}}{\mathbf{q}_1^+} + (1+\mathbf{a}_1^+) \odot \mathbf{u}_1^+ \tag{51}$$

Now, we return to Eq. 45 to derive the dynamical equation for the modulatory neuron $\mathbf{u}_1$. Starting with $\mathbf{y}_1 \odot \mathbf{u}_1^+ = \mathbf{b}_1 \odot \mathbf{z}_1$, we take the modulus on both sides and raise to the power of n:

$$|\mathbf{y}_1|^n \odot (\mathbf{u}_1^+)^n = \mathbf{b}_1^n \odot |\mathbf{z}_1|^n \tag{52}$$

Substituting the expression for $|\mathbf{z}_1|^n$ from Eq. 37:

$$|\mathbf{y}_1|^n \odot (\mathbf{u}_1^+)^n = \mathbf{b}_1^n \odot [\mathbf{I} - \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1]^{-1}(\boldsymbol{\sigma}^n \odot |\mathbf{y}_1|^n) \tag{53}$$

We assume that $\mathbf{b}_1$ has equal components, i.e., $\mathbf{b}_1 = b_1\mathbf{1}$. Multiplying both sides by $[\mathbf{I} - \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1]$ and rearranging:

$$[\mathbf{I} - \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1]\left(\frac{(\mathbf{u}_1^+)^n \odot |\mathbf{y}_1|^n}{\mathbf{b}_1^n}\right) = \boldsymbol{\sigma}^n \odot |\mathbf{y}_1|^n \tag{54}$$

Multiplying both sides by $\mathrm{D}\left(\frac{1}{|\mathbf{y}_1|^n}\right)$:

$$\mathrm{D}\left(\frac{1}{|\mathbf{y}_1|^n}\right)[\mathbf{I} - \mathrm{D}(|\mathbf{y}_1|^n)\mathbf{N}_1]\left(\frac{(\mathbf{u}_1^+)^n \odot |\mathbf{y}_1|^n}{\mathbf{b}_1^n}\right) = \mathrm{D}\left(\frac{1}{|\mathbf{y}_1|^n}\right)(\boldsymbol{\sigma}^n \odot |\mathbf{y}_1|^n) \tag{55}$$

This simplifies to:

$$\frac{(\mathbf{u}_1^+)^n}{\mathbf{b}_1^n} - \mathbf{N}_1\left(\frac{(\mathbf{u}_1^+)^n \odot |\mathbf{y}_1|^n}{\mathbf{b}_1^n}\right) = \boldsymbol{\sigma}^n \tag{56}$$

Since we assumed that $\mathbf{b}_1$ has equal components, we get:

$$(\mathbf{u}_1^+)^n - \mathbf{N}_1((\mathbf{u}_1^+)^n \odot |\mathbf{y}_1|^n) = \boldsymbol{\sigma}^n \odot \mathbf{b}_1^n \tag{57}$$

Elementwise multiplying both sides by $\mathbf{u}_1/(\mathbf{u}_1^+)^n$ from left and rearrange:

$$0 = -\mathbf{u}_1 + \frac{\mathbf{u}_1}{(\mathbf{u}_1^+)^n} \odot \left(\mathbf{b}_1^n \odot \boldsymbol{\sigma}^n + \mathbf{N}_1((\mathbf{u}_1^+)^n \odot |\mathbf{y}_1|^n)\right) \tag{58}$$

This equation describes the steady state of $\mathbf{u}_1$. We propose the following dynamical equation for $\mathbf{u}_1$,

$$\tau_{u_1}\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} = -\mathbf{u}_1 + \frac{\mathbf{u}_1}{(\mathbf{u}_1^+)^n} \odot \left(\mathbf{b}_1^n \odot \boldsymbol{\sigma}^n + \mathbf{N}_1((\mathbf{u}_1^+)^n \odot |\mathbf{y}_1|^n)\right) \tag{59}$$

These equations (Eq. 39, Eq. 59, Eq. 51, Eq. 50) describe the dynamics of principal excitatory neurons, modulatory inhibitory neurons, and modulatory excitatory neurons in V1. They exhibit the fixed point solution given by the normalization by design. These equations are general for any choice of n and β. We get a particularly simplified form when we set $n = 2$ and $\beta = \{2, \frac{1}{2}, 1, 1\}$ for $\{\mathbf{y}_1, \mathbf{u}_1, \mathbf{a}_1, \mathbf{q}_1\}$, respectively.
Finally we rewrite the set of four dynamical equations for V1 as:

$$\tau_{y_1}\frac{\mathrm{d}\mathbf{y}_1}{\mathrm{d}t} = -\mathbf{y}_1 + \mathbf{b}_1^{(y_1)} \odot \mathbf{z}_1 + \left(\frac{1}{1+\mathbf{a}_1^+}\right)\left(\mathbf{W}_{11}\sqrt{\mathbf{y}_1^+} + \gamma_1^{(y_1)} \odot \mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}\right) \tag{60}$$

$$\tau_{u_1}\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} = -\mathbf{u}_1 + ((\mathbf{b}_1^{(u_1)})^2 \odot \sigma^2) + \mathbf{N}_1\left(\mathbf{y}_1^+ \odot (\mathbf{u}_1^+)^2\right) \tag{61}$$

$$\tau_{a_1}\frac{\mathrm{d}\mathbf{a}_1}{\mathrm{d}t} = -\mathbf{a}_1 + \frac{\gamma_1^{(a_1)} \odot \mathbf{W}_{12}\sqrt{\mathbf{y}_2^+}}{\mathbf{q}_1^+} + \mathbf{u}_1^+ + \mathbf{a}_1^+ \odot \mathbf{u}_1^+ + \alpha_1\frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t} \tag{62}$$

$$\tau_{q_1}\frac{\mathrm{d}\mathbf{q}_1}{\mathrm{d}t} = -\mathbf{q}_1 + \sqrt{\mathbf{y}_1^+} \tag{63}$$

where we have decoupled the input gain $\mathbf{b}_1$ and feedback gain γ_1 across each equation. When $\mathbf{b}_1^{(y_1)} = \mathbf{b}_1^{(u_1)}$ and $\gamma_1^{(y_1)} = \gamma_1^{(a_1)}$, and $\mathbf{W}_{11} = \mathbf{I}$ we exactly recover the normalization equation, Eq. 2, at steady state for $\mathbf{y}_1$. Then to reduce the number of parameters we re-define the input gain as $\boldsymbol{\beta}_1 \equiv \mathbf{b}_1^{(y_1)}/\mathbf{b}_1^{(u_1)}$ and fix $\mathbf{b}_1^{(u_1)} = 1/2$. Similarly, we re-define the feedback gain as $\gamma_1 \equiv \gamma_1^{(y_1)}/\gamma_1^{(a_1)}$ and fix $\gamma_1^{(a_1)} = 1/2$. We thus arrive at Eq. 1 of

the main text. Finally, note that the term $\alpha_1 \frac{\mathrm{d}\mathbf{u}_1}{\mathrm{d}t}$ is added to the dynamics of $\mathbf{a}_1$ to achieve correct oscillatory behavior. This term does not impact the steady state solution but influences the dynamics.

The dynamical equations for V2 can be derived following a similar procedure:

$$\tau_{y_2} \frac{\mathrm{d}\mathbf{y}_2}{\mathrm{d}t} = -\mathbf{y}_2 + \mathbf{b}_2^{(y_2)} \odot \mathbf{z}_2 + \left(\frac{1}{1+\mathbf{a}_2^+}\right)\left(\mathbf{W}_{22}\sqrt{\mathbf{y}_2^+}\right) \tag{64}$$

$$\tau_{u_2} \frac{\mathrm{d}\mathbf{u}_2}{\mathrm{d}t} = -\mathbf{u}_2 + ((\mathbf{b}_2^{(u_2)})^2 \odot \boldsymbol{\sigma}^2) + \mathbf{N}_2\left(\mathbf{y}_2^+ \odot (\mathbf{u}_2^+)^2\right) \tag{65}$$

$$\tau_{a_2} \frac{\mathrm{d}\mathbf{a}_2}{\mathrm{d}t} = -\mathbf{a}_2 + \mathbf{u}_2^+ + \mathbf{a}_2^+ \odot \mathbf{u}_2^+ + \alpha_2 \frac{\mathrm{d}\mathbf{u}_2}{\mathrm{d}t} \tag{66}$$

$$\tau_{q_2} \frac{\mathrm{d}\mathbf{q}_2}{\mathrm{d}t} = -\mathbf{q}_2 + \sqrt{\mathbf{y}_2^+} \tag{67}$$

where, $\mathbf{z}_2 = \mathbf{W}_{21}\mathbf{y}_1^+$; and, like for $\mathbf{y}_1$, we have decoupled the input gain $\mathbf{b}_2$ across each equation. When $\mathbf{b}_2^{(y_2)} = \mathbf{b}_2^{(u_2)}$ and $\mathbf{W}_{22} = \mathbf{I}$ we exactly recover the normalization equation, Eq. 2, at steady state for $\mathbf{y}_2$. To reduce the number of parameters, we redefine the input gain as $\boldsymbol{\beta}_2 \equiv \mathbf{b}_2^{(y_2)}/\mathbf{b}_2^{(u_2)}$ and fix $\mathbf{b}_2^{(u_2)} = \mathbf{b}_2^{(y_2)} = 1/2$. Note that in principle $\mathbf{y}_2$ can receive feedback from higher cortical areas, in which case the equations would look identical to Eq. 1 but the subscripts change as $1 \to 2$ and $2 \to 3$.

In addition to the membrane potential, the corresponding firing rates of V1 and V2 neurons are also simulated. These are modeled as low-pass filtered versions of their target rectified power-law transformations. For V1, as an example:

$$\tau_{y_1^+} \frac{\mathrm{d}\mathbf{y}_1^+}{\mathrm{d}t} = -\mathbf{y}_1^+ + \lfloor \mathbf{y}_1 \rfloor^2 \tag{68}$$

$$\tau_{u_1^+} \frac{\mathrm{d}\mathbf{u}_1^+}{\mathrm{d}t} = -\mathbf{u}_1^+ + \sqrt{\lfloor \mathbf{u}_1 \rfloor} \tag{69}$$

$$\tau_{a_1^+} \frac{\mathrm{d}\mathbf{a}_1^+}{\mathrm{d}t} = -\mathbf{a}_1^+ + \lfloor \mathbf{a}_1 \rfloor \tag{70}$$

$$\tau_{q_1^+} \frac{\mathrm{d}\mathbf{q}_1^+}{\mathrm{d}t} = -\mathbf{q}_1^+ + \lfloor \mathbf{q}_1 \rfloor \tag{71}$$

Intrinsic neural circuit noise

The influence of noise on the deterministic neural dynamics is treated analytically in this model. We consider two primary sources of noise: firing rate noise originating from the spiking process and synaptic noise.

i) Firing rate noise is added to the firing rate vector $\mathbf{k}^+$, where $\mathbf{k}^+$ represents various neural population firing rates (e.g., excitatory like $\mathbf{y}_1^+, \mathbf{u}_1^+$or inhibitory like $\mathbf{a}_1^+, \mathbf{q}_1^+$). This noise is attributed to the spiking mechanism, whose strength is found by the modified Gaussian Rectification (GR) model [186, 187]. Due to the high degree of redundancy in neural circuits, the strength of this noise is quenched due to the averaging of these redundant signals. The result is that the noise has a negligible impact on the firing rate dynamics [192, 193]. For a given firing rate vector $\mathbf{k}^+$, with an associated activation exponent β, the noise dynamics are described by:

$$\tau_{k^+} \frac{\mathrm{d}\mathbf{k}^+}{\mathrm{d}t} = -\mathbf{k}^+ + \lfloor \mathbf{k} \rfloor^\beta + \sigma_k \cdot \boldsymbol{\eta}(t), \tag{72}$$

where τ_{k^+} is the time constant for the firing rate dynamics of population $\mathbf{k}^+$, σ_k represents the noise strengths for different neuron types, and $\boldsymbol{\eta}(t)$ is a vector of independent Gaussian white noise process, each with zero mean and unit variance.

ii) Synaptic noise is modeled as exponentially low-pass filtered Gaussian white noise ($\mathbf{f}$), given by the following SDE:

$$\tau_f \frac{\mathrm{d}\mathbf{f}}{\mathrm{d}t} = -\mathbf{f} + \sigma_f \cdot \boldsymbol{\eta}(t), \tag{73}$$

where τ_f is the noise filtering time constant ($\tau_f = 1$ msec, approximating the time constant of synaptic transmission), and σ_f represents the noise strength. This filtered noise ($\mathbf{f}$) is then added to the dynamical equations for the membrane potentials of each neuron, Eq. 60-67, and it is the main source of intrinsic noise.

Upon incorporating these sources of noise, we get a high-dimensional nonlinear stochastic dynamical system with the following variables:

- Variables representing the membrane potential for different neuron types implementing normalization (Eq. 60-67).
- Firing rates (Eq. 72) for each membrane potential variable.
- Filtered synaptic noise (Eq. 73) for each membrane potential variable.

We then linearize this dynamical system about the fixed point found by simulating the deterministic part of the dynamical system, which gives us the Jacobian matrix **J**. We also construct the dispersion matrix **L** and the diffusion matrix **D** using the strength of the noise added in various dynamical variables. The definition of these matrices allows us to compute the covariance matrix and power spectra analytically as discussed in the Methods section of the main text.

Extrinsic noise in the local field potential

In addition to the noise sources detailed above, we also account for additional sources of noise that contribute to LFP power and coherence spectra. We consider two components of extrinsic noise: **i)** Correlated noise ($\boldsymbol{\zeta}_{\text{correlated}}$), assumed to stem from the volume conduction of electric fields generated by transmembrane current through the extracellular brain tissue, with negligible impact on spiking activity and firing rates [97–99]. **ii)** Uncorrelated noise ($\boldsymbol{\zeta}_{\text{uncorrelated}}$), that encompasses noise from various other sources including motion artifact, instrumentation noise, thermal noise, and biological sources of noise [194–196]. This extrinsic noise is added to the simulated neural responses (membrane potentials for different neuron types, denoted collectively by $\mathbf{x}$):

$$\mathbf{x}'(t) = \mathbf{x}(t) + \boldsymbol{\zeta}_{\text{uncorrelated}}(t) + \boldsymbol{\zeta}_{\text{correlated}}(t) \tag{74}$$

Since these noise sources are uncorrelated with the responses of the neurons, the power spectrum of the measured signals $\boldsymbol{\mathcal{S}}'(\omega)$ is the sum of the spectrum of the neurons' activity $\boldsymbol{\mathcal{S}}(\omega)$ and the spectra of these noise components:

$$\boldsymbol{\mathcal{S}}'(\omega) = \boldsymbol{\mathcal{S}}(\omega) + \boldsymbol{\mathcal{S}}_{\text{uncorrelated}}(\omega) + \boldsymbol{\mathcal{S}}_{\text{correlated}}(\omega) \tag{75}$$

The correlated noise is modeled to produce a $\frac{1}{\omega^4}$ power law scaling, which is achieved by doubly low-pass filtered white noise. Its power spectrum for each element is given by:

$$\boldsymbol{\mathcal{S}}_{\text{correlated, element}}(\omega) = \frac{\Sigma_{\text{corr}}}{(1 + \tau_{\text{noise}}^2 \omega^2)^2} \tag{76}$$

The uncorrelated power spectrum $\boldsymbol{\mathcal{S}}_{\text{uncorrelated}}(\omega)$ is diagonal, with each diagonal term having a similar filtered form:

$$\boldsymbol{\mathcal{S}}_{\text{uncorrelated}}(\omega) = \mathrm{D}\left(\frac{\Sigma_{\text{uncorr}}}{(1 + \tau_{\text{noise}}^2 \omega^2)^2}\right) \tag{77}$$

In our simulation, the ratio of power spectral densities of the correlated noise to the uncorrelated noise (i.e, $\Sigma_{\text{corr}}/\Sigma_{\text{uncorr}}$ for corresponding element) is set to 10 for all the results. The noise parameters used in the simulation are as follows,

τ_x	τ_f	τ_{noise}	σ_f	Σ_{corr}	Σ_{uncorr}
1 ms	1 ms	50 ms	0.01	9×10^{-4}	9×10^{-5}

Table 3: *Noise parameters*

Supplementary Material

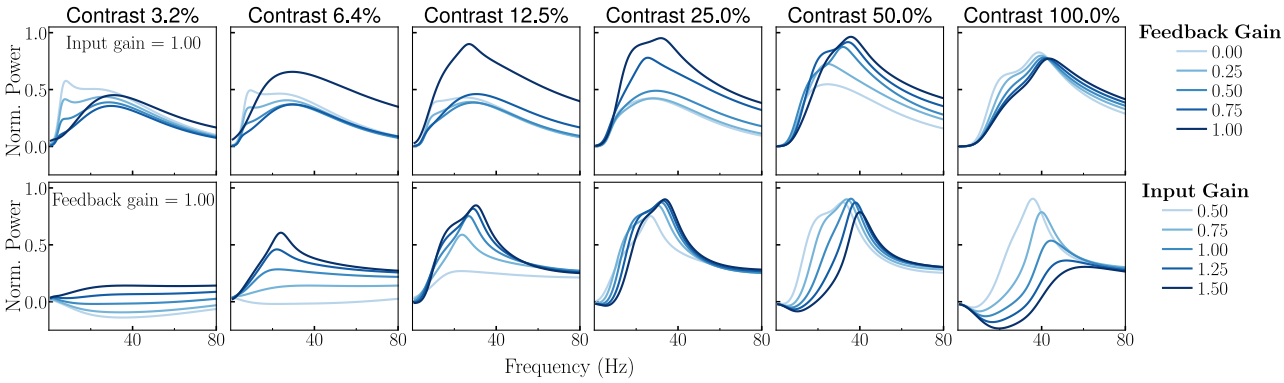

Figure 8: ***Theoretical predictions of varying feedback and input gain on power spectra across different contrast levels.*** *Each column displays results for a specific contrast, with contrast increasing from left to right across panels.* ***Top row,*** *Normalized power as a function of frequency, for varying feedback levels (indicated by the blue line color intensity), with input gain kept at baseline (1.0).* ***Bottom row,*** *Normalized power as a function of frequency for varying input gain levels, with feedback gain kept at baseline (1.0).*

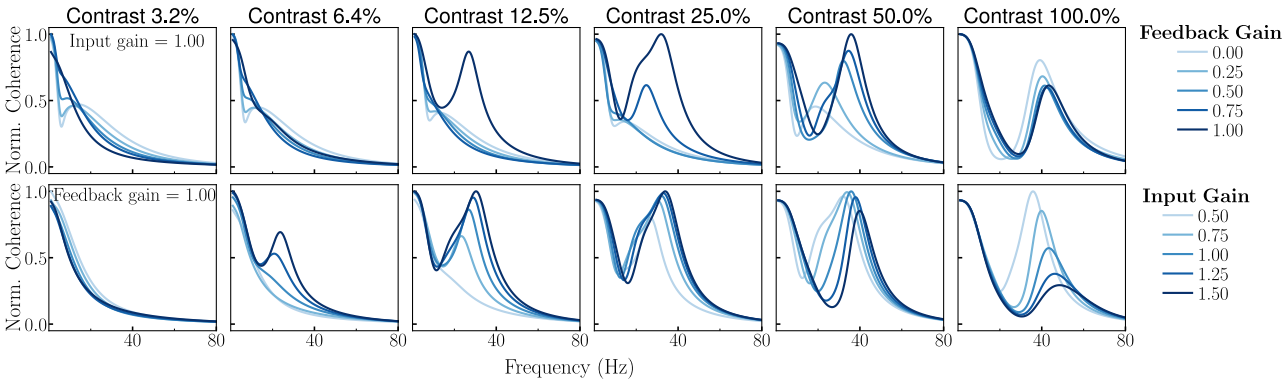

Figure 9: ***Theoretical predictions of varying feedback and input gain on V1-V2 coherence across different contrast levels.*** *Each column displays results for a specific contrast, with contrast increasing from left to right across panels.* ***Top row,*** *Normalized coherence as a function of frequency, for varying feedback levels (indicated by the blue line color intensity), with input gain kept at baseline (1.0).* ***Bottom row,*** *Normalized coherence as a function of frequency for varying input gain levels, with feedback gain kept at baseline (1.0).*

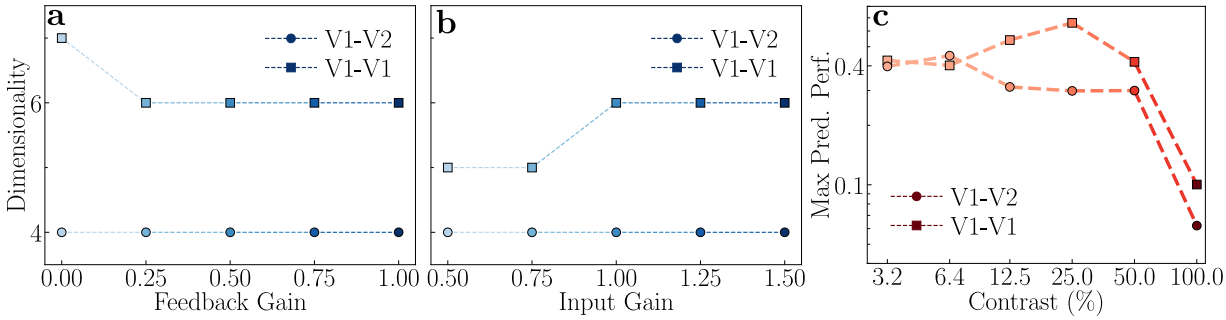

Figure 10: ***Theoretical predictions: Varying feedback gain, input gain, and contrast on communication subspaces.*** ***a,*** *Dimensionality of inter-areal (V1-V2, circles) and within-area (V1-V1, squares) communication subspaces as a function of feedback gain, at 100% stimulus contrast.* ***b,*** *Dimensionality of inter-areal (V1-V2, circles) and within-area (V1-V1, squares) communication subspaces as a function of input gain, at 100% stimulus contrast.* ***c,*** *Maximum prediction performance for inter-areal (V1-V2, circles) and within-area (V1-V1, squares) communication as a function of stimulus contrast.*

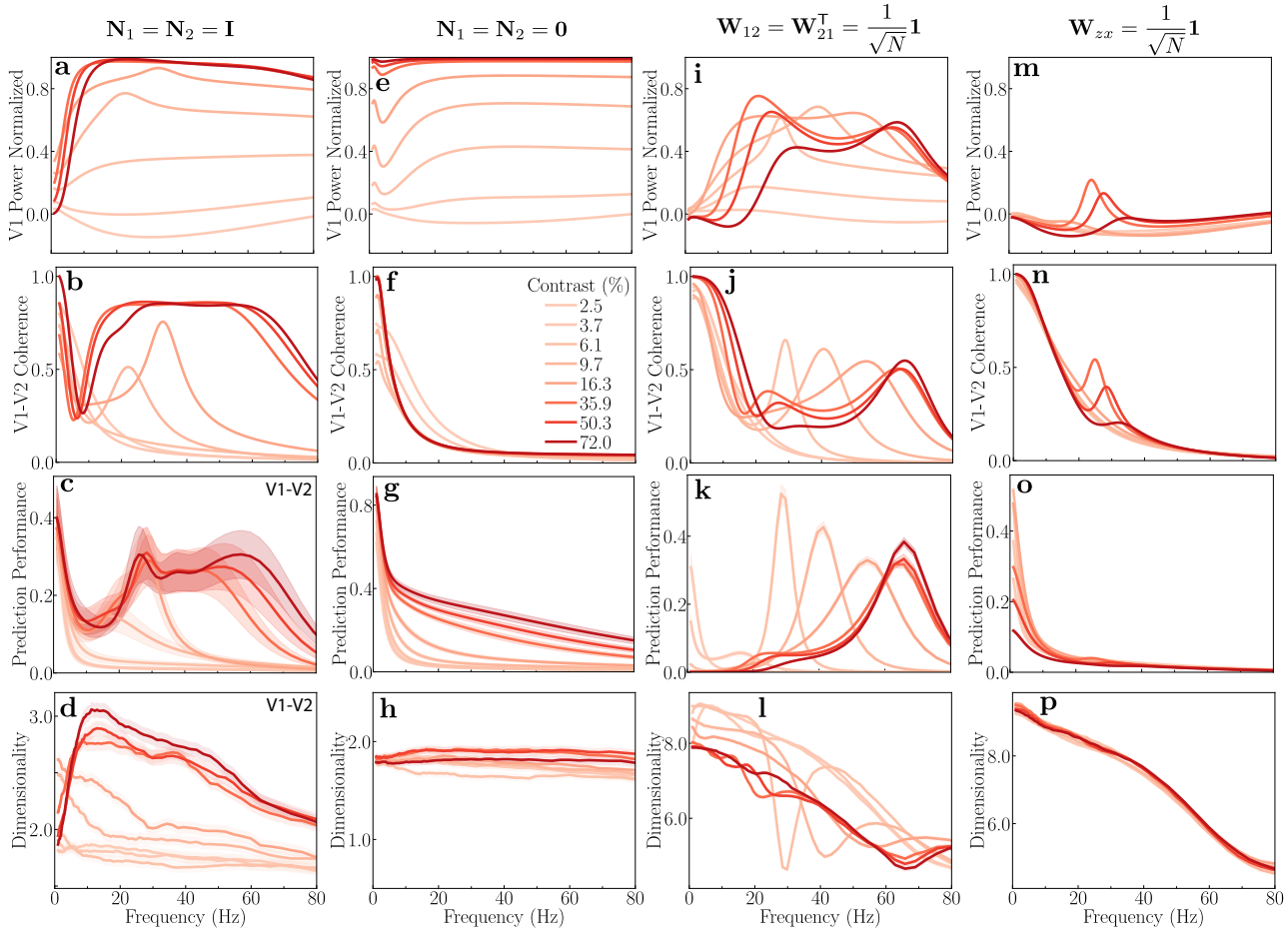

Figure 11: ***Effect of normalization and V1-V2 connectivity on oscillatory activity and communication subspaces.*** *The impact of divisive normalization and inter-area connectivity on network dynamics is explored across four model configurations (columns). For each model, we plot the V1 power spectrum, V1-V2 coherence, prediction performance, and the dimensionality of the communication subspace (rows). For communication subspace analysis, 18 out of 72 simulated neurons were randomly selected from each area, and the analysis was repeated multiple times; shaded areas represent the standard error of the mean (SEM). The number of neurons for the communication subspace analysis was selected for direct comparison with published experimental measurements of communication subspaces with broadband (natural image) input stimuli (see Suppl. Fig. 7 of Ref. [32]).* ***a-d, Self-normalization****: Model predictions when the normalization matrix is set to identity (*$\mathbf{I}$*), meaning each neuron only normalizes its own activity.* ***a,*** *Normalized V1 power spectra at different stimulus contrast values.* ***b,*** *The coherence peaks between V1 and V2 are narrow band at low contrast, and become broadband at high contrasts (*$>$ *16%).* ***c,*** *Prediction performance is highest at frequencies corresponding to peak coherence.* ***d,*** *The communication subspace dimensionality is low, but increases with contrast.* ***e-h, No-normalization****: Model predictions when the normalization matrix is a null matrix (i.e, normalization is removed).* ***e,f*** *Removing normalization eliminates prominent oscillatory activity in both the power and coherence spectra.* ***g,*** *Removing normalization also eliminates peaks in prediction performance.* ***f,*** *Subspace dimensionality remains low and shows little or no dependence on frequency or contrast.* ***i-l, Uniform inter-area connectivity****: Model prediction when the V2 to V1 connectivity matrix modified to* $\mathbf{I} + \frac{1}{\sqrt{N}}\mathbf{1}$*, where N is number of principal neruons and* $\mathbf{1}$ *is an* $(N \times N)$ *matrix of ones.* ***i,*** *V1 power shows distinct peaks in the low and high gamma frequency range.* ***j-k,*** *High coherence corresponds to high prediction performance.* ***l,*** *The dimensionality is higher compared to when inter-area connectivity is sparse, but still much lower than the number of neurons (18), and still markedly reduced at frequencies where coherence is strongest.* ***m-p, Uniform external input****: V1 tuning curves are modified to be flat, signifying an equal preference for all stimulus orientations (or equivalently, that the input stimulus is broadband, e.g., noise or natural images).* ***m,n,*** *Oscillatory activity is evident in both power and coherence spectrum.* ***o,*** *Prediction performance decreases with frequency.* ***p,*** *Communication subspace dimensionality is higher compared to when the tuning curves (or stimulus) is narrow band, but still much lower than the number of neurons (18).*