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

Abstract

The primate brain exhibits a hierarchical, modular architecture with conserved microcircuits executing canonical computations across reciprocally connected cortical areas. 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. 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. 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; iii) Normalization reduces the 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.

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] is one such fundamental canonical principle, playing a central role in regulating neural responses across sensory and higher cortical areas.

Feedback connections are ubiquitous throughout the primate brain, yet their precise computational role remains poorly understood. A wide range of functions has been attributed to feedback, including attentional modulation [2], inter-areal communication [3], learning [4], and generative modeling [5]. While divisive normalization has been extensively studied in feedforward recurrent networks [6–9], its implementation and effects in hierarchical networks with feedback 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) [10], proposes that synchronized oscillations (coherence) are crucial for effective inter-areal communication. Indeed, coherence between brain regions has been widely observed across cognitive, perceptual, and behavioral tasks [11, 12]. 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 emerging from underlying interactions [13, 14]. A complementary hypothesis suggests that brain areas communicate through a low-dimensional *communication subspace* (CS) embedded within the high-dimensional neural activity space [15]. This concept, also identified in artificial neural networks [16], posits that only activity within these shared subspaces between the source and target areas facilitates communication. Despite strong empirical evidence for both coherence-based and subspace-based communication, a unified theoretical framework capable of jointly addressing these mechanisms and explaining how they are dynam-

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

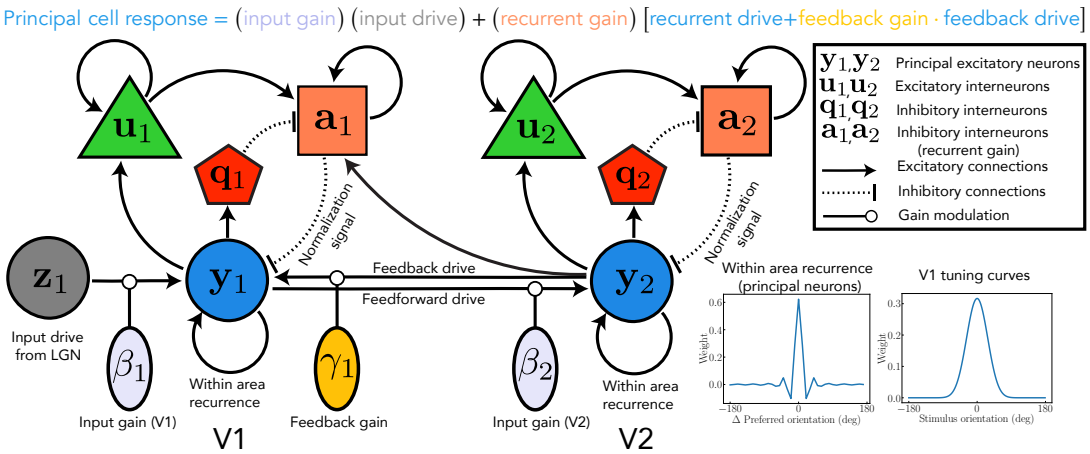

Figure 1: ***Schematic of the two-stage recurrent circuit model.*** *Solid and dashed lines represent excitatory and inhibitory connections, respectively. Principal excitatory neurons are denoted by* $\mathbf{y}$*, modulatory excitatory neurons by* $\mathbf{u}$*, and modulatory inhibitory neurons by* $\mathbf{a}$ *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.*

ically modulated by factors such as feedback and stimulus contrast is currently lacking.

To address these gaps, we introduce a hierarchical recurrent neural circuit theory (Fig. 1) built upon two *a priori* constraints: i) a hierarchy of brain areas with feedforward, recurrent, and feedback connections, and ii) recurrently implemented normalization within each brain area. Within this framework, we derive analytical expressions for inter-areal coherence and communication subspaces, offering experimentally testable predictions. Remarkably, these two constraints alone give rise to a rich set of emergent properties that explain a wide spectrum of neurophysiological phenomena, including modulation of contrast response functions, oscillatory dynamics evident in power and coherence spectra, the emergence of communication subspaces, and the dynamic control of functional connectivity. A central contribution of this work is a unified view of inter-areal communication. The theory simultaneously predicts empirical observations of both communication subspaces and inter-areal coherence, and further reveals that coherence actively shapes the communication subspace. In doing so, our framework reconciles previously competing hypotheses and provides a principled account of how coherence influences subspaces.

Finally, we emphasize 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. Additionally, we have verified that the key results hold across a wide range of parameters; for further details, see the *SI Appendix*.

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 [17], including higher level vision, olfaction, audition, somatosensation, attention, multi-sensory integration, working memory, and decision-making. For citations see [6]. 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.

In addition, normalization decorrelates population responses to reduce redundancy [18, 19] and prevents excessive recurrent excitation to maintain network stability [8, 9]. It is a computational-level description of neural responses that can be implemented using various mechanisms. Experimental evidence suggests that normalization operates via recurrent amplification [20, 21], amplifying weak inputs more than strong inputs, consonant with evidence that cortical circuits are dominated by recurrent excitation [22, 23].

Oscillatory **R**ecurrent **G**ated **N**eural **I**ntegrator **C**ircuits (ORGaNICS) is a theoretical framework that dynamically implements divisive normalization by modulating recurrent amplification [6, 7]. 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 connections 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: excitatory principal cells, excitatory modulator cells, and inhibitory modulator 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 (see equation at the top of Fig. 1). Unlike generic E-I models in which all neurons perform the same computation, ORGaNICs assign unique computations to each cell type (see Discussion and Methods). This 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 receives *input drive* defined as a weighted sum of responses from 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 and negative synaptic weights, which determine each neuron's stimulus preferences, such as orientation selectivity (Fig.1, V1 tuning curves), analogous to a standard receptive field model (see Methods). Input gain is determined independently by input modulator cells (β_1 and β_2 in Fig.1), whose gain may, in principle, depend on both afferent input and local population activity (e.g., responses of both LGN cells and V1 cells may contribute to the input gain in V1).

Each neuron also receives *recurrent drive* arising from weighted interactions among principal cells in the same brain area (Fig. 1, curved arrows). Recurrent connectivity follows a center-surround architecture, with local excitatory and distal inhibitory connections (Fig. 1, recurrent synaptic weights). The gain of this recurrent drive is dynamically controlled by recurrent modulator cells ($\mathbf{a}_1$ and $\mathbf{a}_2$), which combine excitatory ($\mathbf{u}_1$, $\mathbf{u}_2$) and inhibitory signals ($\mathbf{q}_1$, $\mathbf{q}_2$) from local interneuron populations as well as feedback from higher cortical areas (solid arrow from $\mathbf{y_2}$ to $\mathbf{a}_1$ in Fig. 1). Divisive normalization is achieved dynamically through the activity of these modulator cells, which carry the "normalization signal".

Finally, each area receives *feedback drive*, a weighted sum of responses from neurons at higher stages of the hierarchy (e.g., V2 $\rightarrow$ V1; Fig. 1). The feedback drive is specified by a feedback weight matrix, which, for simplicity, we take to be the transpose of the corresponding feedforward matrix from V1 to V2 [5]. The strength of this feedback is controlled by a feedback gain parameter (γ_1 in Fig. 1). More generally, feedback strength may be modulated either by changing the feedback gain or by changing the gain of the neural responses in the higher cortical areas providing feedback (see Discussion). Thus, the feedback gain may be a fixed property of the circuit or a dynamically regulated quantity that depends on both the inputs and outputs of the corresponding area. Importantly, even if the feedback gain is fixed, it may still be possible to manipulate the feedback strength experimentally to test predictions of the theory (elaborated in what follows).

The two-stage recurrent model with feedback implements divisive normalization exactly and simultaneously across multiple cortical areas. In particular, when the gains are set to $\gamma = \beta = 1$, and the recurrent weight matrix is the identity matrix (i.e., each neuron recurrently amplifies only itself), the steady-state responses of the model coincide precisely with the analytical steady-state solutions of the normalization equation (Methods, Eq. 2).

Contrast response functions

Both V1 (dashed) and V2 (solid) exhibit the characteristic saturating, sigmoidal dependence on contrast (As shown in Fig. 2a, b). 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) [24, 25]. 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 [27].

The theory offers a novel, experimentally-testable prediction that increasing either the feedback gain, γ_1, or input gain, β_1, enhances neural responses across cortical areas (Figs. 3a,b). Importantly, this response enhancement is evident primarily as a change in contrast gain, i.e., a shift of the contrast-response functions on the log contrast axis rather than a change in response amplitude.

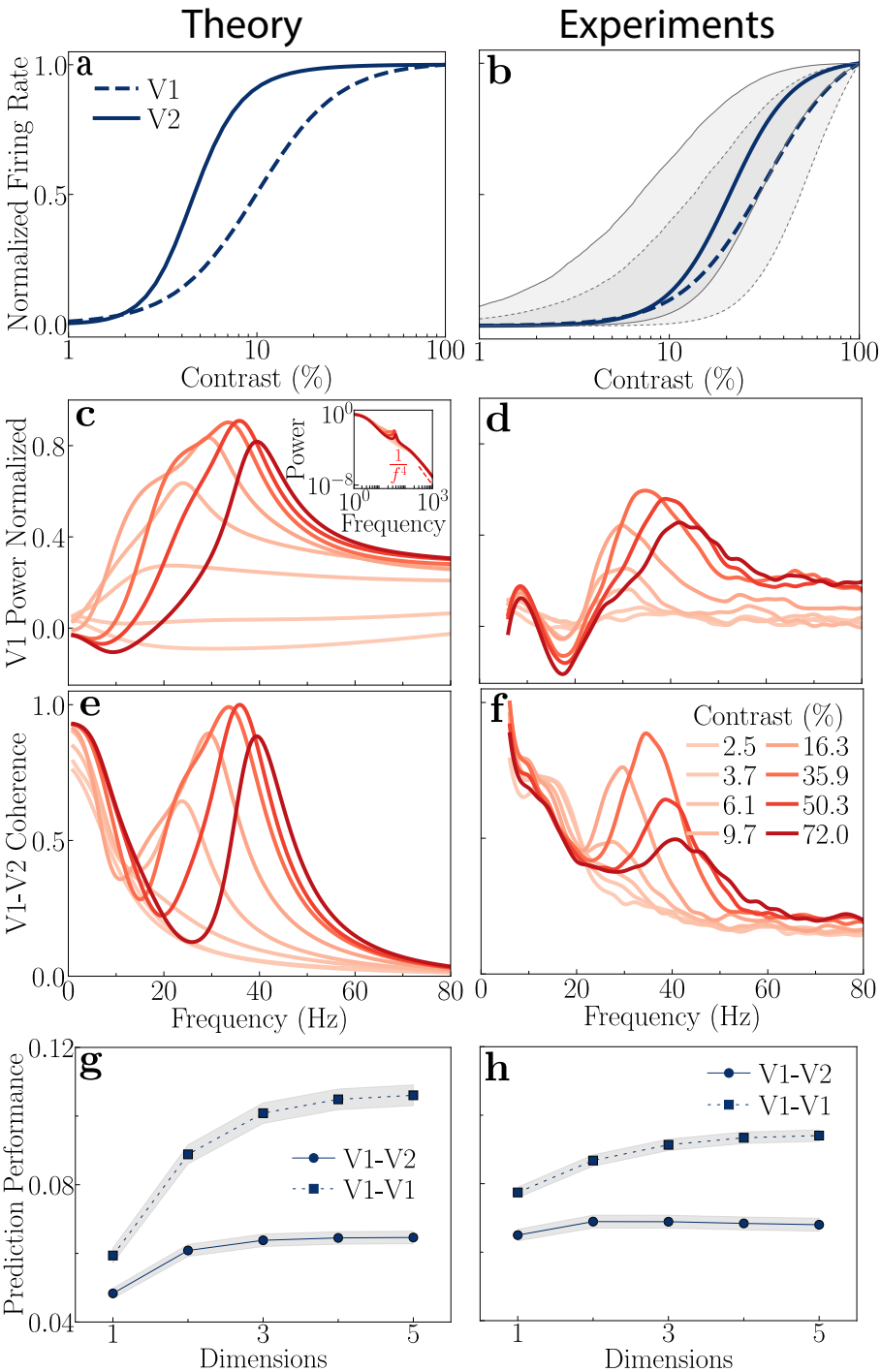

Figure 2: ***Comparison of theoretical predictions (left column) and experimental observations (right column) in V1 and V2.*** *All theoretical predictions are generated using the same baseline parameters (Table S1).* ***a,*** *Theoretical mean firing rates as a function of stimulus contrast (V1: dashed line, V2: solid line).* ***b,*** *Experimental mean firing rates of V1 and V2 (data from [24] and [25], 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 [26], replotted for comparison).* ***e,*** *Theoretical coherence spectra (normalized) 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 [26], 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 [15], 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).*

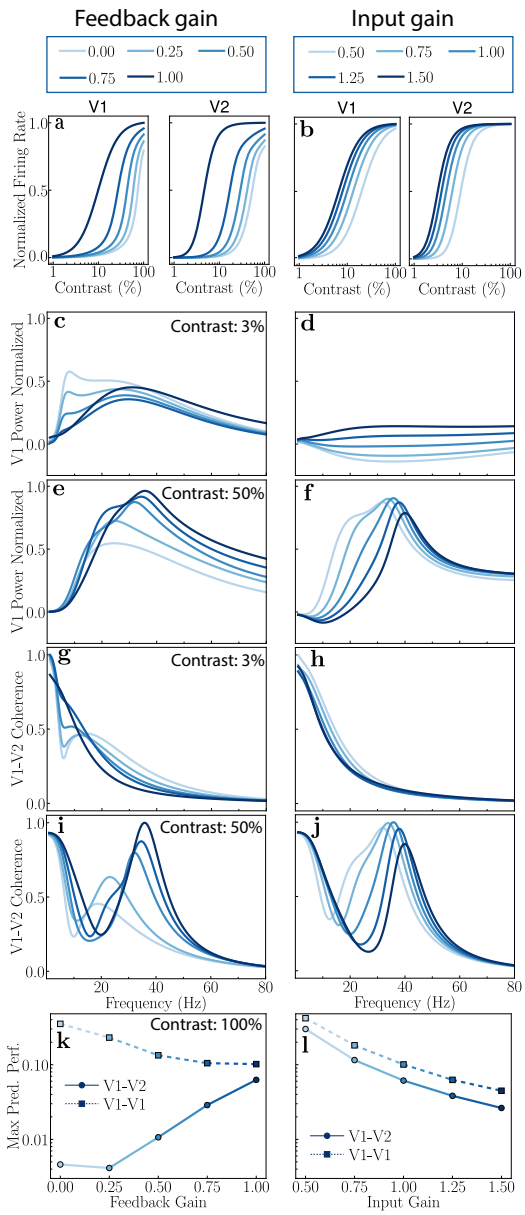

Figure 3: ***Theoretical predictions: Modulating feedback (left) and input gain (right). a,b,*** *Firing rates as a function of contrast. Increasing either feedback gain or input gain enhances neural responses, with feedback gain showing a greater impact in the higher cortical area V2.* ***c,d,*** *V1 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 feedback gain, 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.*

V1 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 [28]. Since our model has a known fixed point solution, we 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 [29]. Further details are provided in the *SI 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 [26]. 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. The oscillatory activity in the gamma-frequency range depends on normalization; removing normalization eliminates the peaks (see Suppl. Fig. S4e), commensurate with experimental observations that gamma oscillations are linked to normalization [30, 31]. The peak frequencies depend on the values of the intrinsic time constants; 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 [32], 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 recurrent processing in the circuit that also acts as a low-pass filter (each of which contributes a factor of $1/f^2$ to the power spectrum).

The theory offers a novel, experimentally-testable prediction that modulating feedback and input gains elicit distinct changes in the power spectra. Fig-

ures 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. S1).

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 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, regardless of the input gain and feedback gain values (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.

Inter-areal 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 [10, 33, 34].

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; [26]). The model accurately captures several crucial experimental observations: 1) high coherence at lower frequencies for low contrasts; 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. The coherence peaks in the gamma frequency range are due to normalization; removing normalization eliminates the peaks in coherence (see Suppl. Fig. S4f). The theory offers a novel, experimentally-testable prediction that feedback and input gain differentially modulate V1-V2 coherence across contrasts (Figs. 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)" [15]. 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 [15].

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 Figs. 2g,h, and Figs. 3k,l) 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; [15]).

The theory offers a novel, experimentally-testable prediction that feedback and input gain differentially modulate communication subspaces (Figs. 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 Suppl.

Figs. S3a,b). Conversely, increasing input gain reduced both the inter-area and within-area communication (Fig. 3l). These results are for 100% contrast; see Suppl. Fig. S3c 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 [35].

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 frequency component (see Methods). Using the Wiener-Khinchin theorem, we converted the power spectra of the source and target population into their noise correlations. To isolate noise correlations at a specific frequency, we applied an idealized narrow-band pass filter. This conceptual filter is implemented via a transfer function, which is non-zero only within a very narrow frequency band around the frequency of interest. Applying this filter to the power spectra and integrating across the frequency domain yields the frequency-specific correlations (see Methods, "Frequency decomposition of the communication subspace". 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. These communication peaks are due to normalization; removing normalization eliminates the peaks in both coherence and prediction performance (see Suppl. Figs. S4f,g). Under conditions of very low contrast, no distinct peak in communication is observed; instead, communication is mostly concentrated at very low frequen-

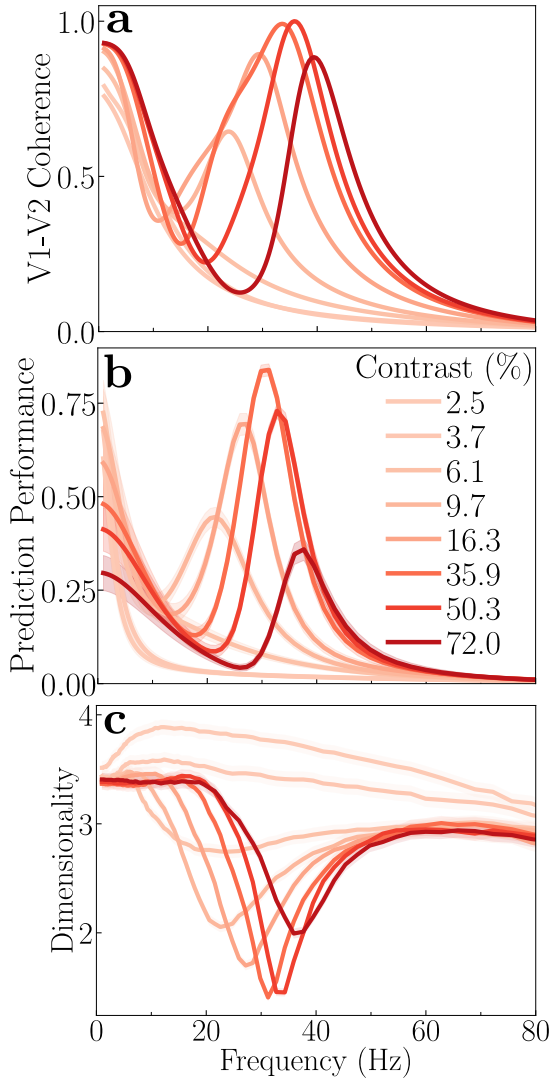

Figure 4: ***Theoretical prediction: Frequency-dependence of communication.*** ***a,*** *V1-V2 coherence spectra for different contrasts; same as in Fig. 2e.* ***b,*** *Prediction performance versus frequency at different contrasts, averaged across different subsets of the source and target populations. At low stimulus contrast ($\approx$ 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) across different subsets of the source and target populations.*

cies (< 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, "Unified view of inter-areal communication").

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 dictated by the rank of the covariance matrix between the source (V1) and target (V2) subpopulations (see Methods: Interpretation of the results, Suppl. Fig. S5). The dimensionality of the CS is influenced mainly by three factors. First, dimensionality depends on normalization. Although dimensionality is relatively low in the absence of normalization (Suppl. Fig. S4h), including normalization further reduces dimensionality at coherent frequencies (Fig. 4c). The intuition for why normalization reduces dimensionality is that when only a few neurons are strongly driven, and their responses are highly correlated across areas (V1 and V2), normalization suppresses the activity of other nearby neurons that are not strongly driven. This suppressive interaction, akin to a soft winner-take-all mechanism, diminishes the contribution of other neurons to the shared covariance, thereby reducing the overall dimensionality. Second, the dimensionality is increased for a fully interconnected (all-to-all) connectivity between V1 and V2 (Suppl. Fig. S4l). Third, delocalized inputs (e.g., noise images or naturalistic images) lead to a higher dimensionality than localized inputs (e.g., gratings) (Suppl. Fig. S4p), commensurate with experimental results [15].

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 [36, 37].

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 [38]. 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

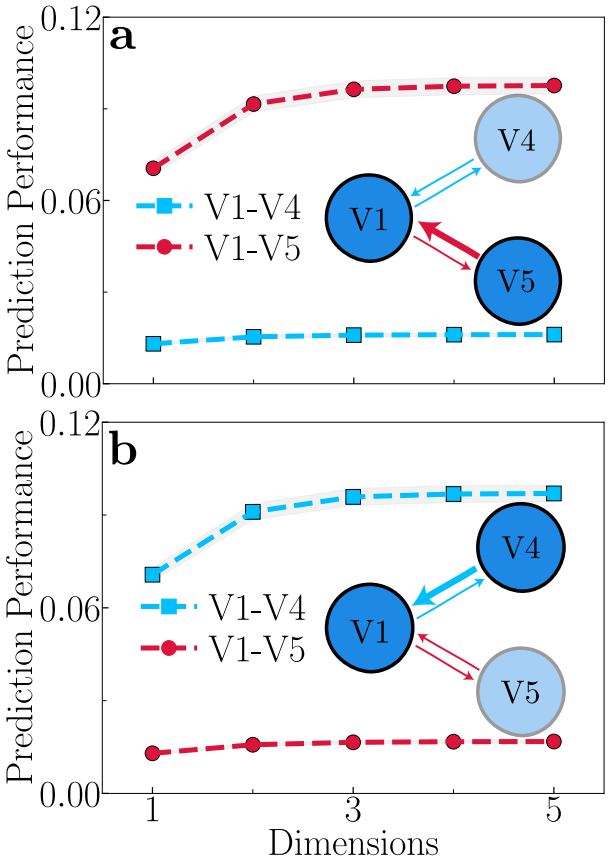

Figure 5: ***Theoretical prediction: Feedback-dependent modulation of functional connectivity.*** ***a,*** *Feedback from V5 → V1 is stronger than feedback from V4 → V1, and V1 preferentially enhances communication with V5.* ***b,*** *Feedback from V4 → V1 is stronger than feedback from V5 → V1, and 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 → V1 is greater than the feedback from V4 → 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 atten-

tion and task demands may selectively boost activity in one or another cortical area [39], and specifically that attending to motion increases the transfer of motion information (communication) between V1 and V5 [40]. 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.

Discussion

We present an analytically tractable theory of neural circuit function that implements divisive normalization dynamically across 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. Our theory yields two key, experimentally-testable predictions: first, high-coherence frequencies serve as a signature for both stronger communication and reduced subspace dimensionality, and second, that increasing feedback strength selectively enhances inter-areal communication and diminishes within-area communication, without altering the subspace dimensionality. A comprehensive list of further predictions is provided in *SI Appendix* section "Robustness with respect to model parameters". 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 mechanism for dynamically changing the functional connections between brain areas over time.

Neural circuit function

The original ORGaNICs model [6, 7], in its initial form without feedback connections, serves as a biophysically-plausible counterpart to Long Short Term Memory units (LSTMs) [41], a class of artificial recurrent neural networks (RNNs) prevalent in machine-learning applications [42]. Unlike LSTMs or any other previously proposed RNNs, ORGaNICs have built-in recurrent normalization, which offers several advantages, including stability even with exuberant recurrent connections [8, 9].

The hierarchical version of the ORGaNICs theory introduced here adds feedback connections, drawing a partial analogy to "hierarchical predictive coding" models of sensory and perceptual processing [43–48]. However, our hierarchical theory 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 [5] 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 [49], whereas predictive coding theories would suggest that the activity, representing prediction error, should be "explained away" as the stimulus becomes predictable.

Normalization

The proposed theory admits an analytical fixed point that follows the normalization equation (Eq. 2) *exactly*, for arbitrary non-negative normalization weights, when the input drive is constant over time (see *SI Appendix* for derivation) [8, 9]. Although several circuit mechanisms have been hypothesized to underlie normalization, including shunting inhibition, synaptic depression, and inhibition-stabilized networks [21, 50, 51], these models suffer from significant limitations. They either lack recurrent amplification or do not exhibit dynamics linked to normalization (e.g., gamma oscillations), or implement weighted normalization only approximately. Critically, none of them converge *exactly* to the normalization equation (Eq. 2). These limitations have significant practical consequences for generating testable predictions and fitting empirical data. Thus, despite the broad empirical success of the normalization model, it remains uncertain whether prior circuit models [51, 52] offer adequate mechanistic explanations.

Our 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 exhibit a small

response, leading to 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 [21]. 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 theory links normalization with oscillatory activity. Removing divisive normalization eliminates oscillatory peaks in power and coherence spectra (Suppl. Figs. S4e,f) and abolishes the link between coherence, prediction performance, and subspace dimensionality (Suppl. Figs. S4g,h), demonstrating that normalization is central to both the dynamics and function of the circuit.

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 Discussion section "Attention and functional connectivity"). Second, the feedback drive may contribute to selective attention (see Section "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 contributes to the normalization signal by providing excitatory drive to inhibitory neurons (see Methods). This mechanism is essential for maintaining network stability, and is commensurate with experimental evidence that normalization depends in part on loops through higher visual cortical areas [53].

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, creating sensory representations from higher-level abstract knowledge, akin to recalling an image from memory. When input and feedback are balanced, the framework may integrate prior beliefs with sensory evidence (see [5] for details), akin to models based on Bayesian inference [54, 55].

The theory 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. A similar conclusion was recently reached by Ni et al. [56] using a large-scale spiking neural circuit model. 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 [57]. 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 accounts for changes in functional connectivity [58] and feature-based attention [59]. 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 [60].

The theory may also provide insights into the neural processes underlying spatially-selective attention. The theory identifies two distinct pathways through which attention modulates neural gain. First, attention may modulate the input gain of the sensory neurons, consistent with established normalization model of attention [61]. Second, attention may modulate the activity of the sensory neurons via the top-down feedback drive. We propose that these mechanisms may map onto different cognitive processes: for instance, changes in input gain might primarily drive exogenous (bottom-up) attention, while the feedback drive facilitates endogenous (top-down) attention. Crucially, these two pathways are not just conceptually distinct but produce divergent spectral signatures. Our theory predicts differential effects of changing input gain versus feedback gain on the power and coherence spectra. These spectral metrics, therefore, offer an empirical means to distinguish between the two mechanisms.

Thalamo-cortical model

Alpha oscillations are a defining characteristic of thalamo-cortical dynamics, particularly prominent within the visual system during states of "relaxed wakefulness" [62, 63]. Specifically, alpha oscillations are hypothesized to be generated and modulated by thalamo-cortical loops [63, 64], suggesting that alpha rhythms signify a state of functional inhibition or "idling" in sensory pathways, potentially gated by corticothalamic feedback [64, 65]. While the hierarchical model presented in this paper was designed for cortico-cortico interactions, it may be adapted to model the thalamo-cortical interactions that contribute to the generation of alpha waves. 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 [65, 66]. It is, however, unclear if the current theory will suffice, on its own, to fully capture the phenomenology of 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 [67] and thalamocortical loops in other neural systems [68].

Limitations

Our model assumes instantaneous feedback, whereas biological feedback is delayed due to axonal conduction, synaptic transmission, and processing in higher cortical areas [69, 70]. Examining how inter-areal synchronization depends on feedback delay is an important direction for future work.

The feedforward, feedback, and recurrent weight matrices may contain both positive and negative weights. We therefore posit inhibitory interneurons to implement the sign inversion, corresponding to negative synaptic weights. 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.

Although most computations in the theory are biophysically plausible (see "Biological plausibility" and [7]), some mechanisms remain unresolved. In particular, while weighted sums and gain control may be implemented via balanced excitatory–inhibitory push–pull and push–push mechanisms [71, 72], we do not specify a cellular mechanism for the element-wise product of firing rates in Eq. 1b. The intrinsic time constants used in simulations are unrealistically short for individual neurons (see Table S1). Using more realistic values would shift gamma-band oscillations to lower frequencies. These parameters may therefore be better interpreted as effective population time constants rather than as the intrinsic time constants (capacitance/conductance) of individual neurons.

Finally, the model omits neural adaptation, a ubiquitous feature of cortical responses. Incorporating adaptation into the hierarchical framework remains an important direction for future studies.

Biological plausibility

We hypothesize a correspondence between the variables in the theory and cortical microcircuits composed of identified cell types. Unlike generic excitatory-inhibitory (E-I) models (e.g., [74]), our framework assigns unique computations and nonlinearities to distinct cell types (see *SI Appendix*). Principal cells $\mathbf{y}$ are hypothesized to correspond to pyramidal neurons; inhibitory modulatory cells $\mathbf{a}$ to a combination of parvalbumin-expressing (PV+) and somatostatin-expressing (SST+) interneurons; and inhibitory interneurons $\mathbf{q}$ to vasoactive intestinal polypeptide (VIP) neurons.

In the proposed circuit (Fig. 1), Principal cells $\mathbf{y}$ excite both inhibitory modulatory cells $\mathbf{a}$ (via $\mathbf{u}$) and

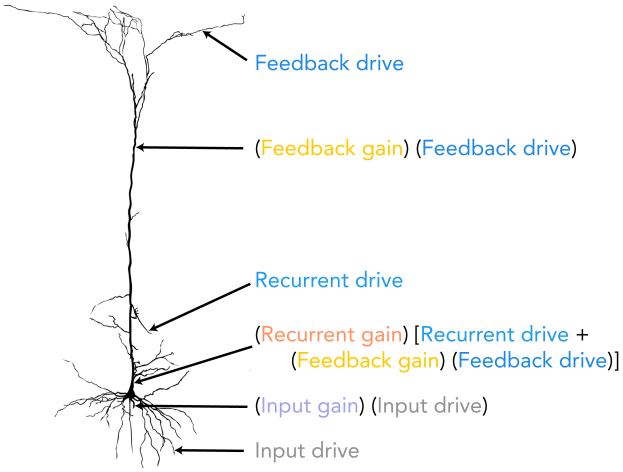

Figure 6: ***Hypothesized mapping of computational components in the model onto the dendritic compartments of a pyramidal cell.*** *The input drive is hypothesized to be the weighted sum of feedforward inputs arriving at the distal basal dendrites. This is modulated by an input gain, corresponding to 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 pyramidal neuron cell shown is from [73].*

inhibitory interneurons **q**. VIP-like cells **q** inhibit modulatory cells **a**, which in turn inhibit principal cells **y**. This architecture closely mirrors known cortical motifs: pyramidal neurons excite PV+, SST+, and VIP cells; VIP cells inhibit SST+ cells; and both PV+ and SST+ cells inhibit pyramidal neurons [75–79]. We further hypothesize that modulator cells **a** have large receptive fields and broad orientation-selectivity (reflecting properties of the normalization pool), consistent with the response properties of SST+ and PV+ neurons, respectively [76, 80, 81].

We also posit excitatory feedback from higher cortical areas to both principal cells **y** and modulatory cells **a** (Fig. 1). Feedback to principal cells is likely received by the distal apical dendrites of layer 1, while feedback to modulator cells **a** is commensurate with experimental evidence that feedback from higher cortical areas targets SST+ neurons [82–84], and that normalization depends in part on loops through higher visual cortical areas [53].

One potential discrepancy between the model and the cortical circuitry is that excitatory input from principal cells **y** to modulatory cells **a** is mediated by an intermediate excitatory population **u**, whereas pyramidal neurons directly excite PV+ and SST+ cells in cortex. However, the theory can be reformulated so that principal cells provide direct excitatory input to modulatory cells by substituting Eq. 1b into Eq. 1c (see Methods), without altering the model's functional behavior.

Finally, we hypothesize that the computations performed by principal cells correspond to processes distributed across dendritic compartments of pyramidal neurons (Fig. 6). 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 is computed in the distal basal dendrites, commensurate with evidence that feedforward synaptic inputs contact basal dendrites. The input gain could be controlled by inhibitory synapses near the basal trunk. The feedback drive is computed in 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 [85]. The recurrent drive is likely computed at more proximal apical dendrites, with recurrent gain shaped by inhibitory inputs near proximal apical trunk.

Methods

The Model

We propose a multistage hierarchical recurrent circuit that incorporates feedforward connections from the primary visual cortex (V1) to higher visual 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. [6, 7].

The response of principal excitatory neurons in the model is a sum of three key components: i) 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; ii) 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 iii) 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 modula-

tory inhibitory neurons ($\mathbf{a}$ and $\mathbf{q}$), with the following dynamics:

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

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

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

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

The above equations are for V1; V2 follows a similar set of equations (see *SI 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 [86, 87]. 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 (Fig. 7a illustrates the resulting orientation tuning curves). $\tau_{\mathbf{k}}$ 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 [7]. 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 [8, 9]. $\mathbf{W}_{12}$ 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 ($\mathbf{W}_{11}$) [9]. 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 *SI 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, γ_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 *SI Appendix*). 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). 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

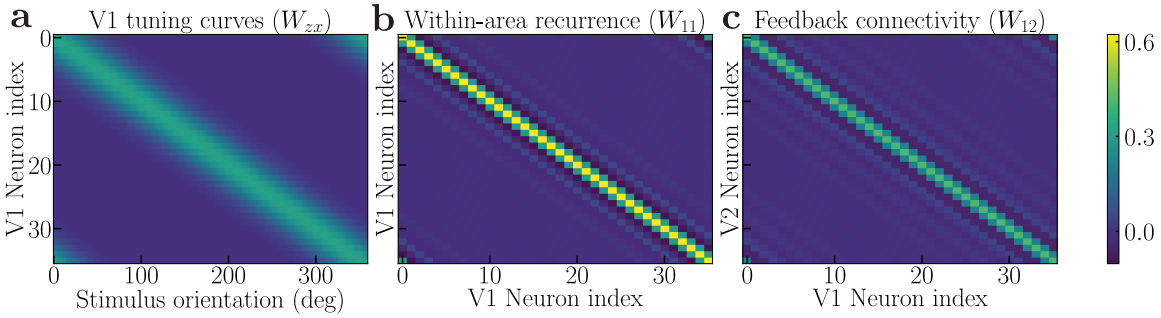

Figure 7: ***Connectivity matrices in the hierarchical recurrent circuit model.*** ***a****, Orientation tuning curves corresponding to the V1 encoding matrix ($\mathbf{W}_{zx}$). Each row represents the tuning curve of a V1 neuron. The weight matrix itself has positive and negative synaptic weights like standard models of orientation-selective receptive fields.* ***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 ($\mathbf{W}_{12} = \mathbf{W}_{21}$).*

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 $\mathbf{L}d\mathbf{W}$:

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

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); $d\mathbf{W}$ is a vector of independent Gaussian increments with correlation matrix $\mathbf{D}$ ($\mathbb{E}[d\mathbf{W} \cdot d\mathbf{W}^\mathsf{T}]$), 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 ($\mathbf{x}^*$):

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

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 [88]:

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

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 [15]. 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 [88]:

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

Once $\mathbf{C}(0)$ is obtained, we follow the method proposed by Semedo et al. [15] 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:

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

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 $\mathbf{B}_{\mathrm{opt}}$, we minimize ε with respect to $\mathbf{B}$, leading to the Ordinary Least Squares solution:

$$\mathbf{B}_{\mathrm{opt}} = \mathbf{C}_1^{-1}\mathbf{C}_3, \quad (10)$$

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

$$\hat{\mathbf{C}}_2 = \mathbb{E}[\hat{\mathbf{t}}\hat{\mathbf{t}}^\mathsf{T}] = \mathbf{B}_{\mathrm{opt}}^\mathsf{T}\mathbf{C}_1\mathbf{B}_{\mathrm{opt}} \quad (11)$$

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

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

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

$$\hat{\mathbf{C}}_2 = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\mathsf{T} \quad (13)$$

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, $\mathbf{B}_{\mathrm{RRR}}$, is constructed by projecting the optimal solution $\mathbf{B}_{\mathrm{opt}}$ onto the subspace spanned by the top i eigenvectors of $\hat{\mathbf{C}}_2$.

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

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:

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

where ε_i is the MSE for the rank-i approximation

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

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 ($\varepsilon_0 = \mathrm{Tr}(\mathbf{C}_2)$).

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

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

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

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

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 $\hat{\mathbf{C}}_2$.

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

Now the expression for prediction performance simplifies to:

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

Which can also be interpreted as:

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

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

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

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 governed by the rank of the cross-covariance matrix, $\mathbf{C}_3$. In the frequency domain, this means the dimensionality at a specific frequency is determined by the rank of the cross-power spectral density matrix (see Suppl. Fig. S5a). 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). Normalization reduces dimensionality because when just a small subset of neurons is strongly driven, and their responses are highly correlated across areas (V1 and V2), normalization suppresses the activity of surrounding neurons with weaker drive. This suppression, resembling a soft winner-take- all process, diminishes the contribution of those less active neurons to the population's shared variability, leading to a lower effective dimensionality.

In contrast, without this normalization, no such correlation between coherence and dimensionality is observed (Fig. S4h).

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.

Code Availability

The code to generate all the figures can be found in the following GitHub repository: `https://github.com/asit-pal/Hierarchical-ORGaNICs`

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.

Declaration of interests

David Heeger is co-founder and chief scientific officer of The Sequel Institute, which has licensed IP from NYU related to ORGaNICs. Shivang Rawat is currently a full-time employee at The Sequel Institute. Stefano Martiniani is a consultant for The Sequel Institute.

References

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

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

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

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

[13] 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.

[14] 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.

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

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

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

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

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

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

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

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

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

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

[30] Supratim Ray, Amy M Ni, and John HR Maunsell. "Strength of gamma rhythm depends on normalization". In: *PLoS Biology* 11.2 (2013), e1001477.

[31] Dora Hermes et al. "An image-computable model for the stimulus selectivity of gamma oscillations". In: *Elife* 8 (2019), e47035.

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

[33] 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.

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

[35] 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.

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

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

[38] 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.

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

[40] 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.

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

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

[43] 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.

[44] 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.

[45] 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.

[46] 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.

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

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

[49] 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.

[50] Matteo Carandini, David J Heeger, and J Anthony Movshon. "Linearity and normalization in simple cells of the macaque primary visual cortex". In: *Journal of Neuroscience* 17.21 (1997), pp. 8621–8644.

[51] Daniel B Rubin, Stephen D Van Hooser, and Kenneth D Miller. "The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex". In: *Neuron* 85.2 (2015), pp. 402–417.

[52] Guillaume Hennequin et al. "The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability". In: *Neuron* 98.4 (2018), pp. 846–860.

[53] 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.

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

[55] 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.

[56] Shencong Ni, Brendan Harris, and Pulin Gong. "Distributed and dynamical communication: a mechanism for flexible cortico-cortical interactions and its functional roles in visual attention". In: *Communications Biology* 7.1 (2024), p. 550.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

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

[62] 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.

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

[64] 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.

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

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

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

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

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

[70] 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.

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

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

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

[74] 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.

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

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

[77] 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.

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

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

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

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

[82] 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.

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

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

[85] 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.

[86] 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).

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

[88] 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.