



Structure of activity in multiregion recurrent neural networks

David G. Clark^{a,b,1} and Manuel Beiran^{a,b}

Edited by Ilya Nemenman, Emory University, Atlanta, GA; received February 26, 2024; accepted February 7, 2025 by Editorial Board Member Herbert Levine

Neural circuits comprise multiple interconnected regions, each with complex dynamics. The interplay between local and global activity is thought to underlie computational flexibility, yet the structure of multiregion neural activity and its origins in synaptic connectivity remain poorly understood. We investigate recurrent neural networks with multiple regions, each containing neurons with random and structured connections. Inspired by experimental evidence of communication subspaces, we use low-rank connectivity between regions to enable selective activity routing. These networks exhibit high-dimensional fluctuations within regions and low-dimensional signal transmission between them. Using dynamical mean-field theory, with cross-region currents as order parameters, we show that regions act as both generators and transmitters of activity—roles that are often in tension. Taming within-region activity can be crucial for effective signal routing. Unlike previous models that suppressed neural activity to control signal flow, our model achieves routing by exciting different high-dimensional activity patterns through connectivity structure and nonlinear dynamics. Our analysis of this disordered system offers insights into multiregion neural data and trained neural networks.

neural networks | modularity | disordered system | nonlinear dynamics

A striking example of convergent evolution in nervous systems is the emergence of well-defined anatomical regions that interact with one another (1–4). Recent advances in neural-recording technologies have enabled simultaneous monitoring of thousands of neurons across multiple brain areas in vivo (5–8). These studies reveal that neurons exhibit varying degrees of regional specialization in their activities (4, 9–11). This regional specialization, balanced with cross-region interactions, is believed to underlie the flexible, adaptive capabilities of neural circuits (12–14). Modern neural datasets thus reveal an intricate interplay between region-specific and broadly distributed signals.

These datasets raise fundamental questions about the origins and functions of multiregion neural activity (15–18). To address them, researchers have trained multiregion recurrent neural network models, either to perform cognitive tasks (19–22) or to generate recorded neural data (23, 24). These models have shed light on directed multiregion interactions involved in sensorimotor processing, context modulation, and changes in behavioral states (25).

However, in both real neural circuits and their artificial counterparts, the nature of multiregion interactions remains largely mysterious. In particular, we lack understanding of the connectivity supporting modular computations and the mechanisms of flexible signal routing. The coexistence and interaction of region-specific and network-wide dynamics are also unclear.

To address these challenges, we analyze a recurrent network model with multiple regions. Each region has a combination of random and low-rank connectivity, generating both high-dimensional fluctuations and specific low-dimensional patterns (26, 27). We connect regions using low-rank connectivity, enabling selective routing of low-dimensional signals between regions.

Due to its nonlinear dynamics and multiregion connectivity structure, this model produces an extremely rich and broad array of dynamic states depending on the connectivity. We develop an analytical theory of this multiregion activity structure by deriving and solving dynamical mean-field theory (DMFT) equations for the network in the limit where each region has infinitely many neurons for any finite number of regions. Given the complexity of the resulting DMFT equations, we solve them in stages of increasing complexity: first considering symmetric effective interactions leading to fixed-point solutions in the low-dimensional dynamics and then progressing to include disorder. Finally, we examine general effective interactions with the potential for limit-cycle solutions, requiring numerical solution.

Significance

The increasing availability of multiregion neural recordings underscores the challenge of understanding neural dynamics across interconnected brain regions. We propose and analyze a multiregion network model leveraging low-rank connectivity for selective activity routing. We develop a dynamical mean-field theory to analyze the model, revealing a competition between signal generation and transmission within regions. Our work provides an analytically tractable approach to multiregion interactions in high-dimensional, nonlinear neural systems, combining an experimentally motivated form of connectivity with recurrent network dynamics. The mean-field formalism offers a perspective on routing, establishes a theoretical foundation for machine-learning tools in neural data analysis, and advances our understanding of multiregion neural circuits.

Author affiliations: ^aZuckerman Institute, Columbia University, New York, NY 10027; and ^bKavli Institute for Brain Science, Columbia University, New York, NY 10027

Author contributions: D.G.C. and M.B. designed research; performed research; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. I.N. is a guest editor invited by the Editorial Board.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹To whom correspondence may be addressed. Email: dgc2138@cumc.columbia.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2404039122/-/DCSupplemental>.

Published March 7, 2025.

Our analysis of this disordered system reveals two key ideas, each supported by various specific results:

Key idea 1: Regions serve dual roles as generators and transmitters of activity, with an inherent tension between these functions. When the intrinsic dynamics within a region become too strong or complex, the region's ability to transmit signals is compromised. Our analysis characterizes this conflict and demonstrates how taming within-region dynamics is crucial for network-level communication.

Key idea 2: Signal routing throughout the network is achieved by shifting which subspaces of high-dimensional activity space are excited or unexcited through the interplay of connectivity statistics and nonlinear recurrent dynamics. The subset of subspaces that are excited depends on the geometric arrangement of low-rank connectivity patterns and the strength of disordered connectivity. Our approach complements earlier models of gating and routing in neural circuits, which emphasized single-neuron biophysical mechanisms such as neuromodulation, inhibition, or gain modulation (28), by developing a geometric, population-level view of information flow.

Overall, our work provides a theoretical framework for understanding the interplay between regional specialization and multiregion interactions in neural circuits, offering insights into the mechanisms underlying flexible signal routing and modular computations in the brain.

Multiregion Network Model

Here, we present the multiregion network model, first describing its dynamics and then its connectivity.

Dynamics. We study rate-based (nonspiking) recurrent neural networks comprising R regions, each containing N neurons. We consider a finite number of regions R and take the limit $N \rightarrow \infty$, corresponding to a small or moderate number of regions, each with a large number of neurons. The preactivations of the neurons, analogous to membrane potentials, are denoted by $x_i^\mu(t)$, where $\mu \in \{1, \dots, R\}$ specifies the region and $i \in \{1, \dots, N\}$ specifies the within-region neuron. The activations, analogous to firing rates, are given by $\phi_i^\mu(t) = \phi(x_i^\mu(t))$, where $\phi(x) = \text{erf}(\sqrt{\pi}x/2)$ is a pointwise nonlinearity that is linear for small $|x|$ and saturates at ± 1 for large $|x|$. Neurons interact through a synaptic coupling matrix $J_{ij}^{\mu\nu}$ according to

$$\frac{dx_i^\mu(t)}{dt} = -x_i^\mu(t) + \sum_{\nu=1}^R \sum_{j=1}^N J_{ij}^{\mu\nu} \phi_j^\nu(t). \quad [1]$$

Connectivity. The connections within each region μ are dense and consist of the sum of random disordered couplings, χ_{ij}^μ , and a rank-one matrix, as investigated by Mastrogiuseppe and Ostojic (26). This rank-one matrix is defined as the outer product of vectors $\mathbf{m}^{\mu\mu}$ and $\mathbf{n}^{\mu\mu}$. Connections between pairs of regions, such as from region ν to μ , are represented by additional rank-one matrices formed by outer products of vectors $\mathbf{m}^{\mu\nu}$ and $\mathbf{n}^{\nu\mu}$ (Fig. 1A). The synaptic coupling matrix is thus expressed as

$$J_{ij}^{\mu\nu} = \delta^{\mu\nu} \chi_{ij}^\mu + \frac{1}{N} m_i^{\mu\nu} n_j^{\nu\mu}. \quad [2]$$

Each element of χ_{ij}^μ is sampled independently from a zero-mean Gaussian with variance $(g^\mu)^2/N$. This $1/\sqrt{N}$ scaling of

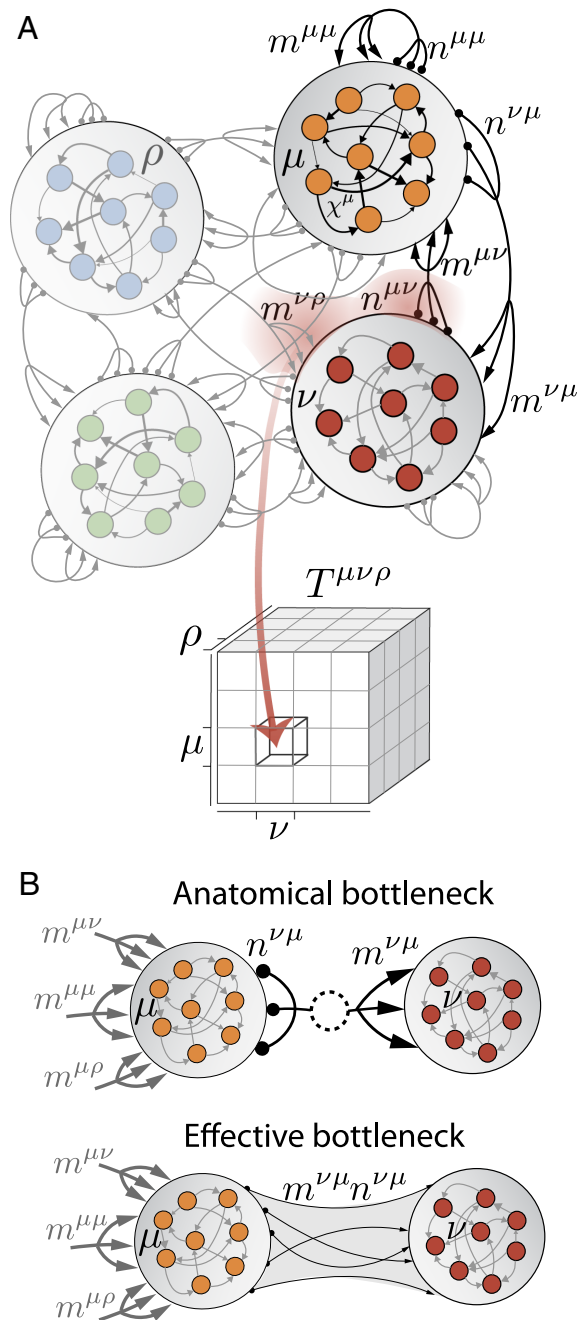


Fig. 1. (A) *Top:* Schematic of the synaptic connectivity model. Different regions, each with “random plus rank-one” connectivity, are linked via rank-one matrices representing communication subspaces. In this network of $R = 4$ regions, we highlight the rank-one and disordered couplings in region μ , as well as the structured couplings to and from region ν . Rank-one connections are defined through the outer product of vectors $\mathbf{m}^{\mu\nu}$ and $\mathbf{n}^{\nu\mu}$. *Bottom:* Tensor $T^{\mu\nu\rho}$, which encodes the geometric arrangement of the connectivity patterns and determines the dynamics of region-to-region currents in the mean-field picture. (B) Anatomical bottleneck or effective bottleneck implementing a rank-one connectivity matrix between regions ν and μ . The dashed circle represents a linear neuron with fast timescale.

the disordered couplings ensures that the eigenspectrum of χ_{ij}^μ remains independent of network size for large N .

For tractability, we assume that the components of the vectors $\mathbf{m}^{\mu\nu}$ and $\mathbf{n}^{\nu\mu}$ are zero-mean random variables drawn from a multivariate Gaussian. Specifically, for each neuron i in region ν , there are $2R$ jointly sampled components: $\{n_i^{\nu\rho}\}_{\rho=1}^R \cup \{m_i^{\nu\rho}\}_{\rho=1}^R$. To define

the second-order statistics of these components, we introduce the tensors:

$$T^{\mu\nu\rho} = \langle n_i^{\mu\nu} m_i^{\nu\rho} \rangle_{\mathbf{J}}, \quad [3a]$$

$$U^{\mu\nu\rho} = \langle m_i^{\mu\nu} m_i^{\nu\rho} \rangle_{\mathbf{J}}. \quad [3b]$$

Our analysis will demonstrate that specifying the remaining second-order statistics, $\langle n_i^{\mu\nu} n_i^{\nu\rho} \rangle_{\mathbf{J}}$, is not necessary to study the dynamics in the limit $N \rightarrow \infty$. However, to sample the vectors defining the low-rank part of the couplings, we must specify $\langle n_i^{\mu\nu} n_i^{\nu\rho} \rangle_{\mathbf{J}}$. We set this proportional to $\delta^{\mu\rho}$ with a scale factor large enough to ensure that the overall covariance matrix of vector components is positive-definite. As $N \rightarrow \infty$, these tensors can equivalently be expressed by the normalized overlaps or inner products:

$$T^{\mu\nu\rho} = \frac{1}{N} \sum_{i=1}^N n_i^{\mu\nu} m_i^{\nu\rho}, \quad [4a]$$

$$U^{\mu\nu\rho} = \frac{1}{N} \sum_{i=1}^N m_i^{\mu\nu} m_i^{\nu\rho}. \quad [4b]$$

Thus, $T^{\mu\nu\rho}$ and $U^{\mu\nu\rho}$ encode the geometric arrangement of connectivity patterns (Fig. 1 *A*, *Bottom*), providing a concise representation of the network's structure. When showing simulation results, we will consider only large networks where the particular realization of connectivity is not significant, and the system behavior is controlled by g^{μ} , $T^{\mu\nu\rho}$, and $U^{\mu\nu\rho}$.

Table 1 summarizes the variables and notation used throughout this article.

Biological Motivations and Assumptions

In constructing this model, we aimed to incorporate sufficient biological detail to capture nontrivial phenomena while maintaining analytical tractability. In this section, we elucidate the biological foundations of our model, outlining its underlying assumptions and limitations, first addressing the dynamics and then the connectivity.

Dynamics: Motivation and Assumptions. The complexity in our network model's dynamics, compared to linear networks that can simply be diagonalized, stems from the nonlinear activations of individual neurons. This nonlinearity is inspired by the transformation of input currents into spike trains by real neurons. While our model captures this crucial aspect, it does not account for other features of cortical circuits, such as distinct excitatory and inhibitory populations (i.e., Dale's law), sparse connectivity, and nonnegative firing rates.

This level of abstraction mirrors that used in the seminal work of Sompolinsky et al. (29), which described chaotic activity arising from strong random connectivity. Indeed, our multiregion model reduces to R independent samples of this model when the structured low-rank couplings are set to zero. In this special case, each disconnected region transitions from quiescence to high-dimensional chaos at a critical coupling variance, defined by $g^{\mu} = 1$.

Our use of this level of abstraction is supported by recent studies demonstrating that network models incorporating the biological features we omitted (i.e., nonnegative rates or spikes, sparse connections, and excitatory-inhibitory populations) can

Table 1. Summary of notation

Network variables	
$x_i^{\mu}(t)$	Preactivation ("membrane potential") of neuron i in region μ at time t (Eq. 1)
$\phi_i^{\mu}(t)$	Activation ("firing rate") of neuron i in region μ at time t (Eq. 1)
Network parameters	
N	Number of neurons in each region
R	Number of regions
J_{ij}^{ν}	Synaptic coupling from neuron j in region ν to neuron i in region μ (Eq. 2)
χ_{ij}^{μ}	Random component of within-region synaptic couplings in region μ (Eq. 2)
g^{μ}	SD (times \sqrt{N}) of random couplings in region μ
$\mathbf{m}^{\mu\nu}$	Vector with components $m_i^{\mu\nu}$; defines structured input pattern from region ν to neurons in region μ (Eq. 2)
$\mathbf{n}^{\mu\nu}$	Vector with components $n_i^{\mu\nu}$; defines structured readout pattern from neurons in region ν to region μ (Eq. 2)
DMFT variables	
$\Delta^{\mu}(t, t')$	Correlation function of preactivations in region μ (Eq. 5a)
$C^{\mu}(t, t')$	Correlation function of activations in region μ (Eq. 5b)
$S^{\mu\nu}(t)$	Current from region ν to region μ at time t (Eq. 6)
$H^{\mu\nu}(t)$	Drive to $S^{\mu\nu}(t)$ in the mean-field dynamics of the currents (Eq. 7)
$\psi^{\mu}(t)$	Neuronal gain in region μ at time t (Eq. 8)
A^{μ}	Sum of squared currents from all regions into region μ
$S_0^{\mu\nu}$	Fixed-point value of interregion current from region ν to region μ
$\sigma^{\mu\nu}(t)$	Perturbation to the interregion current from region ν to region μ
$\hat{\Delta}^{\mu}(\tau)$	Normalized stationary correlation function (Eq. 21)
DMFT parameters	
$T^{\mu\nu\rho}$	Normalized overlap between readout and input patterns, representing effective interaction from region ρ to region μ through region ν (Eq. 4a)
$\hat{T}^{\mu\nu, \rho\sigma}$	Matrix form of $T^{\mu\nu\rho}$ (Eq. 11)
$U^{\mu\nu\rho}$	Overlap between input vectors in region μ originating from regions ν and ρ (Eq. 4b)
$c^{\mu\nu}$	Symmetric parameterization of $T^{\mu\nu\rho}$ (Eq. 12)
u^{μ}	Rank-one contribution to "rank-one plus diagonal" parameterization of $c^{\mu\nu}$ (Eq. 13)
h^{μ}	Diagonal contribution to "rank-one plus diagonal" parameterization of $c^{\mu\nu}$ (Eq. 13)
a^{μ}	Strength of direct self-interaction (Eq. 14)
b^{μ}	Strength of indirect self-interaction (Eq. 14)

exhibit equivalent dynamical regimes. This equivalence has been observed both for disordered couplings, where the same transition to chaos occurs (30, 31), and for low-rank couplings (32, 33).

Connectivity: Motivation and Assumptions. We use rank-one matrices to model structured connectivity both within and between regions, based on separate experimental observations for each type of connectivity.

Within-region recordings show that neural activity during tasks often lies on a low-dimensional manifold (26, 34). Rank-one

connectivity can generate arbitrary one-dimensional dynamics (35), serving as a starting point for modeling structured low-dimensional activity. Many standard neural-network models, including Hopfield networks (36), ring attractors (37), and autoencoders (38), use low-rank connectivity. Furthermore, our model combines rank-one and disordered within-region connectivity. As shown by Mastrogiuseppe and Ostojic (26), such networks can produce chaotic activity, fixed points, or both, depending on the relative strengths of rank-one vs. disordered connectivity.

Cross-region rank-one connections are based on observed communication subspaces between cortical areas. In particular, Semedo et al. (39) found that only a low-dimensional subspace of V1 activity, distinct from the subspace capturing most V1 variance, correlates with activity in V2. Similar communication-subspace structure has been identified in visual processing (40), motor control (41, 42), attention (43), audition (22), and brain-wide activity (44). Low-rank cross-region connectivity offers a simple explanation for these subspaces, but of course is not the only explanation. Alternative hypotheses, such as global fluctuations or shared input, were considered less likely based on anatomy, spatial selectivity, and persistence under anesthesia by the authors of the original study (in the visual cortex). Here, we adopt low-rank connectivity for its simplicity, data compatibility, and, as we discuss in the next section, functional utility.

Biologically, low-rank cross-region connectivity, which acts as a type of bottleneck, can be implemented either anatomically or effectively [Fig. 1B; (26, 45)]. An anatomical bottleneck would involve a set of intermediary neurons between two areas (Fig. 1B, *Top*). These neurons, assumed to be linear with fast time constants, would read out activity from the source region and broadcast it to the target region (46). This framework also accommodates thalamocortical loops as anatomical bottlenecks between cortical regions (this complements existing models where thalamic nuclei create loops within a cortical area; such loops can be selectively modulated via basal-ganglia inhibition, controlling interregion communication). Alternatively, an effective bottleneck would arise from direct, monosynaptic connections between source and target regions with a low-rank structure (Fig. 1B, *Bottom*). A simple example of this occurs when all connections from a source to a target region have the same strength and sign, corresponding to a rank-one matrix that is sensitive only to the mean activity of the source region.

Under the interpretation of an effective bottleneck, the rank-one constraint results in a synaptic coupling from a neuron in region ν to a neuron in region μ that is proportional to the product of two scalar variables: $n_i^{\mu\nu}$ and $m_j^{\mu\nu}$. These variables are associated with the emitter and receiver populations, respectively. Such couplings, expressed as products of pre- and postsynaptic terms, arise naturally in neuroscience as a consequence of Hebbian plasticity.

Finally, while we use rank-one matrices, a more realistic model might involve higher-rank matrices, or matrices with smoothly decaying singular values. We find that even rank-one matrices induce rich multiregion activity structure, providing an adequate starting point.

Functional Significance of Low-Rank Cross-Region Connectivity. A rank-one connectivity matrix implements an activity-dependent bottleneck: the transmission of activity from source region ν to target region μ depends on the alignment of activity in ν with the row space of the connecting low-rank matrix. This row space, given by the span of $\mathbf{n}^{\mu\nu}$, represents the communication

subspace in our model. The bottleneck then projects this filtered activity into target region μ through the column space of the matrix, given by the span of $\mathbf{m}^{\mu\nu}$.

This connectivity structure allows selective communication between regions, controlled by the geometry encoded in $T^{\mu\nu\rho}$. To illustrate this mechanism, consider an activity pattern ϕ_i^ν in region ν . The activity communicated to region μ is proportional to the projection $N^{-1} \sum_{i=1}^N n_i^{\mu\nu} \phi_i^\nu$. For a generic pattern ϕ_i^ν (e.g., induced by the disordered connectivity χ_{ij}^ν), this projection is of order $1/\sqrt{N}$, vanishing as $N \rightarrow \infty$. However, if ϕ_i^ν has a component aligned with $\mathbf{n}^{\mu\nu}$, this projection remains of order unity.

For such alignment to occur, there must exist a region ρ such that $\mathbf{m}^{\nu\rho}$, which delivers input to region ν , has a component along $\mathbf{n}^{\mu\nu}$. This component is precisely $T^{\mu\nu\rho}$. Consequently, high-dimensional chaotic activity cannot propagate between regions as $N \rightarrow \infty$, ensuring that only structured, low-dimensional signals are transmitted.

DMFT

Mean-field theory is an analytical approach that describes large systems using a small set of summary statistics called order parameters. This method provides an exact description as $N \rightarrow \infty$ and a good approximation for large, finite N . DMFT extends this concept by introducing time-dependent order parameters to capture the temporal evolution of activity (29, 47). We now present the order parameters in the DMFT description of our multiregion network model and the equations governing their dynamics.

Order Parameters. Our multiregion model exhibits two types of dynamics: high-dimensional chaotic fluctuations from i.i.d. connectivity, and low-dimensional excitation within or between regions due to low-rank connectivity. These dynamics are described by distinct sets of order parameters.

High-dimensional fluctuations are characterized by correlation functions, which capture the temporal structure of chaotic fluctuations. For each region μ , we define correlation functions for the (pre)activations:

$$\Delta^\mu(t, t') = \frac{1}{N} \sum_{i=1}^N x_i^\mu(t) x_i^\mu(t'), \quad [5a]$$

$$C^\mu(t, t') = \frac{1}{N} \sum_{i=1}^N \phi_i^\mu(t) \phi_i^\mu(t'). \quad [5b]$$

Low-dimensional signal transmission within and between regions is described by currents, following the terminology of Perich et al. (12). These currents are consolidated in the matrix $S^{\mu\nu}(t)$, defined by

$$\frac{dS^{\mu\nu}(t)}{dt} = -S^{\mu\nu}(t) + \frac{1}{N} \sum_{i=1}^N n_i^{\mu\nu} \phi_i^\nu(t). \quad [6]$$

The current $S^{\mu\nu}(t)$ represents the activity in region ν that is transmitted to region μ (plus a low-pass filter).

Routing and Nonrouting Regions. The current matrix provides crucial information about activity flow between regions. We classify regions as routing or nonrouting based on their role

in signal transmission. We say that a region ν is routing if it transmits signals between other regions, indicated by at least one nonzero off-diagonal element in the ν -th column of the current matrix, $S^{\cdot\nu}(t)$; and at least one nonzero off-diagonal element in the ν -th row, $S^{\nu\cdot}(t)$. In contrast, we say that a region ν is nonrouting if all elements of its corresponding row and column in the current matrix are zero, except possibly for the diagonal element, $S^{\nu\nu}(t)$.

As we will demonstrate through exact solutions of the DMFT equations, a region may become nonrouting when its own activity is too strong, preventing signal flow. One way for this to occur is if the region's activity aligns with its internal structured connectivity, resulting in a nonzero diagonal element, $S^{\nu\nu} \neq 0$.

Experimentally, routing of this type could be detected through analyses similar to those used by Semedo et al. (39). By computing the communication subspace for a source region during spontaneous activity, one could see how activity patterns line up with that subspace during a task; the overlapping activity would be the routed signal.

Dynamical Mean-Field Equations. In the mean-field picture, currents interact according to

$$\frac{dS^{\mu\nu}(t)}{dt} = -S^{\mu\nu}(t) + H^{\mu\nu}(t), \text{ where} \quad [7a]$$

$$H^{\mu\nu}(t) = \psi^\nu(t) \sum_{\rho=1}^R T^{\mu\nu\rho} S^{\rho\rho}(t), \quad [7b]$$

where $\psi^\nu(t) = \psi(\Delta^\nu(t, t))$ is the average gain of neurons in region ν . The function $\psi(\Delta)$ performs a Gaussian average:

$$\psi(\Delta) = \langle \phi'(x) \rangle_x, \quad [8]$$

where $x \sim \mathcal{N}(0, \Delta)$. Thus, while standard neural networks have a vector dynamics shaped by a matrix, in our framework, region-to-region interactions, defined by the current order parameters, have a matrix dynamics shaped by a third-order tensor. Meanwhile, $\Delta^\mu(t, t')$ satisfies:

$$\begin{aligned} & \left(1 + \frac{d}{dt}\right) \left(1 + \frac{d}{dt'}\right) \Delta^\mu(t, t') \\ &= (g^\mu)^2 C^\mu(t, t') + \sum_{\nu, \rho=1}^R U^{\mu\nu\rho} H^{\mu\nu}(t) H^{\mu\rho}(t'), \quad [9] \end{aligned}$$

These equations are closed by expressing $C^\mu(t, t')$ in terms of $\Delta^\mu(t, t')$ via $C^\mu(t, t') = C(\Delta(t, t'), \Delta(t, t'), \Delta(t', t'))$, where $C(\Delta_{12}, \Delta_{11}, \Delta_{22})$ propagates preactivation correlations to activation correlations:

$$C(\Delta_{12}, \Delta_{11}, \Delta_{22}) = \langle \phi(x_1) \phi(x_2) \rangle_{x_1, x_2}, \quad [10]$$

where $(x_1, x_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Delta})$. $\psi(\Delta)$ and $C(\Delta_{12}, \Delta_{11}, \Delta_{22})$ can be evaluated analytically (SI Appendix).

Thus, the DMFT provides a set of deterministic, causal dynamic equations for the region-specific two-point functions and currents. While their derivation is relatively straightforward, solving them analytically is challenging due to their nonlinear and time-dependent structure, as well as the tensorial form of the interactions. In the next section, we show that by assuming certain symmetry properties of $T^{\mu\nu\rho}$, we can, remarkably, derive a rich

and instructive class of time-independent and time-dependent solutions.

For the remainder of the paper, we assume $U^{\mu\nu\rho} = \delta^{\nu\rho}$ for all μ , focusing on the role of $T^{\mu\nu\rho}$. Geometrically, this means that inputs from other regions into a target region μ are organized in orthogonal subspaces.

Symmetric Effective Interactions and Fixed Points

We now set out to derive exact solutions to the DMFT equations. In general, to simplify the analysis of many-body interactions, a natural choice is to assume symmetry. In standard neural networks, symmetric interactions ensure that the system converges to fixed points, precluding limit cycles and chaos. However, enforcing symmetry in the DMFT system is challenging because the effective interactions among the currents form a third-order tensor, $T^{\mu\nu\rho}$.

To clarify the structure of the interactions between currents in the DMFT, we rewrite the right-hand side of the current dynamics as $\psi^\nu(t) \sum_{\rho, \sigma=1}^R \hat{T}^{\mu\nu, \rho\sigma} S^{\rho\sigma}(t)$, where

$$\hat{T}^{\mu\nu, \rho\sigma} = \delta^{\nu\rho} T^{\mu\nu\sigma} \quad [11]$$

is a R^2 -by- R^2 dynamics matrix governing the linearized interaction of the R^2 currents (its spectrum is closely related to that of $J_{ij}^{\mu\nu}$; SI Appendix). We expect $\hat{T}^{\mu\nu, \rho\sigma}$ to influence the current dynamics similarly to how the synaptic weight matrix shapes neuronal dynamics in a standard neural network. Thus, a natural choice is to impose symmetry on the matrix $\hat{T}^{\mu\nu, \rho\sigma}$, i.e., $\hat{T}^{\mu\nu, \rho\sigma} = \hat{T}^{\rho\sigma, \mu\nu}$. This reduces the number of free parameters from $\mathcal{O}(R^3)$ to $\mathcal{O}(R^2)$ by requiring

$$T^{\mu\nu\rho} = \delta^{\mu\rho} c^{\mu\nu}, \text{ where } c^{\mu\nu} = c^{\nu\mu}. \quad [12]$$

The presence of $\delta^{\mu\rho}$ in $T^{\mu\nu\rho}$ implies that each region μ interacts either directly with itself ($\mu = \nu$) or indirectly with itself through an intermediate region, ν ($\mu \neq \nu$). Moreover, the symmetry of $c_{\mu\nu}$ implies that the coupling through which region μ interacts with itself via region ν is equivalent to that for region ν interacting with itself via region μ . This is illustrated in Fig. 2A.

To make analytical progress, we further constrain the symmetric matrix $c^{\mu\nu}$ to have a ‘‘rank-one plus diagonal’’ form, with only

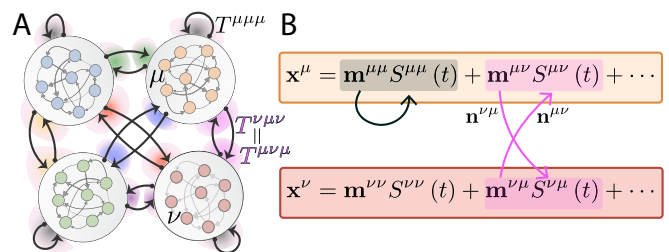


Fig. 2. (A) Restriction to the effective-interaction tensor $T^{\mu\nu\rho}$ corresponding to enforcing symmetry. This constraint sets $T^{\mu\nu\rho} = \delta^{\mu\rho} c^{\mu\nu}$, where $c^{\mu\nu}$ is a symmetric matrix. nonzero overlaps between connectivity patterns are indicated by colored auras, with equal colors indicating equal overlaps. In this scenario with $R = 4$ regions, the connectivity has 10 independent parameters: 4 for direct and 6 for indirect effective self-interactions. (B) Illustration of subspace-based routing in the case of symmetric effective interactions. When the activity subspace defined by the span of m_i^μ in region μ is excited, bidirectional communication between regions μ and ν is suppressed, and vice versa, due to the nonlinear dynamics of the network.

$\mathcal{O}(R)$ parameters,

$$c^{\mu\nu} = u^\mu u^\nu + \delta^{\mu\nu} b^\mu, \quad [13]$$

where u^μ and b^μ are arbitrary vectors. This form provides a minimal setting in which one has independent control over the strength of direct vs. indirect self-interactions, which are captured by the quantities

$$a^\mu = (u^\mu)^2 + b^\mu, \text{ and } b^\mu = (u^\mu)^2, \quad [14]$$

respectively. If $b^\mu = 0$, region μ is not connected to the rest of the network, and its dynamical repertoire is that of a rank-one network with disorder, studied in ref. 26.

Disorder-Free Case. We begin by examining the case without disorder in connectivity: $g^\mu = 0$ for all μ . Symmetric interactions typically lead to fixed points, which we find to be the case here (although we were unable to derive a global Lyapunov function). For the parameterization of $T^{\mu\nu\rho}$ defined above, the fixed points $S_0^{\mu\nu}$ of the currents satisfy:

$$S_0^{\mu\nu} = \psi_0^\nu (u^\mu u^\nu + \delta^{\mu\nu} b^\nu) S_0^{\nu\mu}, \quad [15a]$$

$$\psi_0^\nu = \psi(A^\nu), \text{ where } A^\nu = \sum_{\rho=1}^R (S_0^{\nu\rho})^2. \quad [15b]$$

Here, A^μ represents the squared L^2 -norm of row μ of the current matrix. In the absence of disorder, A^μ is the variance of preactivations in region μ . (Note that with a general form of $U^{\mu\nu\rho}$, this would become a Mahalanobis norm.) These equations yield a combinatorial family of stable and unstable fixed points, which can be categorized based on whether each region is routing or nonrouting. Notably, within this family of fixed points, a region is routing if, and only if, it produces no self-exciting activity, i.e., $S^{\mu\mu} = 0$. This directly illustrates Key Idea 1: the tension between signal generation and transmission.

For a given fixed point, let $\mathcal{S}_{\text{route}} \subseteq \{1, \dots, R\}$ be the subset of regions in routing mode. For a region $\mu \notin \mathcal{S}_{\text{route}}$, Eq. 15a simplifies to:

$$\psi_0^\mu = \frac{1}{a^\mu}, \quad [16a]$$

$$(S_0^{\mu\mu})^2 = A^\mu, \quad [16b]$$

$$S_0^{\mu\nu} = S_0^{\nu\mu} = 0 \text{ for all } \nu \neq \mu. \quad [16c]$$

On the other hand, for a region $\mu \in \mathcal{S}_{\text{route}}$, Eq. 15 implies:

$$\psi_0^\mu = \frac{1}{b^\mu}, \quad [17a]$$

$$S_0^{\mu\mu} = 0, \quad [17b]$$

$$S_0^{\mu\nu} u^\nu = S_0^{\nu\mu} u^\mu \text{ for all } \nu \in \mathcal{S}_{\text{route}} \setminus \{\mu\}. \quad [17c]$$

Additionally, for each region $\mu \in \mathcal{S}_{\text{route}}$:

$$A^\mu = \sum_{\nu \in \mathcal{S}_{\text{route}} \setminus \{\mu\}} (S_0^{\mu\nu})^2. \quad [18]$$

Combining these results, we have

$$A^\mu = \begin{cases} \psi^{-1}\left(\frac{1}{a^\mu}\right) & \text{for } \mu \notin \mathcal{S}_{\text{route}} \\ \psi^{-1}\left(\frac{1}{b^\mu}\right) & \text{for } \mu \in \mathcal{S}_{\text{route}}. \end{cases} \quad [19]$$

Here, $\psi^{-1}(1/x) = 2(x^2 - 1)/\pi$ is a monotonically increasing function of x , so A^μ increases with a^μ or b^μ . These equations determine the row norms A^μ for all μ and the pattern of (non)zero entries in the current matrix for a given bipartition of routing and nonrouting regions. For regions in routing mode, there is remaining freedom in choosing the current-matrix off-diagonal entries, resulting in a manifold of fixed points. We analyze the dimension and topology of this manifold in *SI Appendix*, finding that the set of stable fixed points (see below) forms multiple disconnected continuous attractors in current space, with the number depending on the values of A^μ .

Stability Analysis. There are 2^R possible ways to assign routing and nonrouting modes to regions, producing a combinatorial class of fixed points. To determine which states are stable, we perform a stability analysis, finding that region μ is in routing mode if, and only if, $a^\mu < b^\mu$. To demonstrate this, we consider a first-order perturbation $\sigma^{\mu\nu}(t)$ about a fixed point $S_0^{\mu\nu}$ and define a ‘‘local energy.’’

$$E[\sigma] = \frac{1}{2} \sum_{\mu, \nu=1}^R \left(\frac{\sigma^{\mu\nu}}{u^\mu} \right)^2. \quad [20]$$

We show in *SI Appendix* that $\partial_t E \leq 0$ for all $\sigma^{\mu\nu}$ if and only if $S_0^{\mu\nu}$ is in a configuration claimed to be stable. Moreover, when

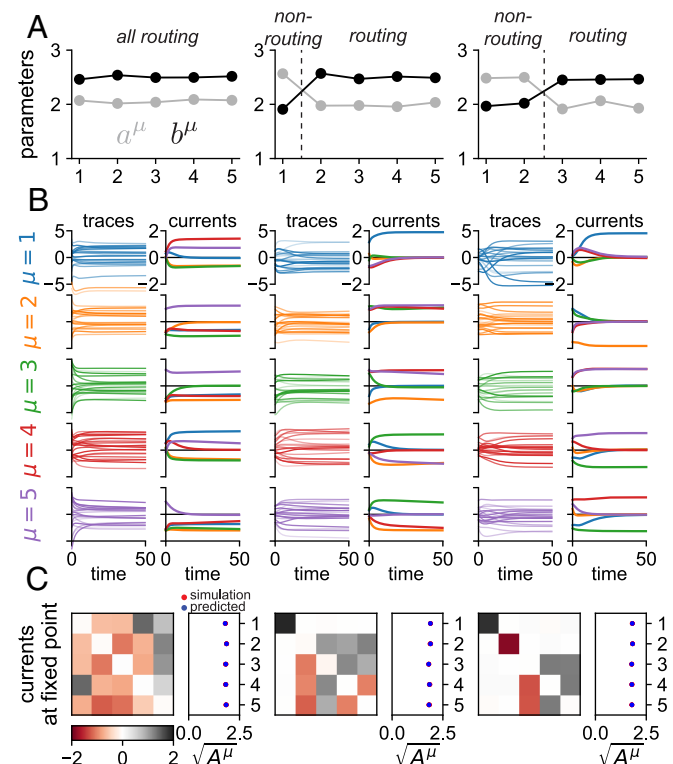


Fig. 3. Structure of fixed points in networks with symmetric effective interactions. The same information for three different cases is shown on the *Left*, *center*, and *Right*. (A) Values of a^μ and b^μ in the $R = 5$ regions. (B) Dynamics of sampled neurons (*Left*) and of incoming currents (*Right*) in large simulations for each region. (C) Visualization of the steady-state current matrix $S_0^{\mu\nu}$ (*Left*) and of the L^2 -norms of the rows of this matrix (*Right*). We show row-norms from the simulations (red dots) alongside analytical predictions (blue dot). In the leftmost plots, all regions are in nonrouting mode. In the *Middle* plots, region 1 is in nonrouting mode and regions 2 to 4 are in routing mode. In the rightmost plots, regions 1 and 2 are in nonrouting mode and regions 3 to 5 are in routing mode.

$S_0^{\mu\nu}$ is stable, there exists a family of choices for $\sigma^{\mu\nu}$ that lead to $\partial_t E = 0$. These directions correspond to translation along a continuous attractor manifold.

In this setup, a region μ can be toggled between routing and nonrouting modes by adjusting the relative magnitudes of a^μ and b^μ (Fig. 3). This approach to routing contrasts with traditional methods that manipulate individual neurons or synapses through neuromodulation, inhibition, or gain modulation. In particular, the gain ψ_0^μ is nonzero in both routing and nonrouting modes, unlike conventional gain-modulation methods that would be analogous to driving ψ_0^μ to zero to achieve a nonrouting state. Through the interplay between connectivity geometry and nonlinear recurrent dynamics, our model aligns neural activity with subspaces that either facilitate or inhibit cross-region communication, reflecting Key Idea 2.

Effect of Disorder. Maintaining the simplified parameterization of $T^{\mu\nu\rho}$, we now introduce disorder into the model by allowing nonzero values of g^μ . This addition potentially leads to high-dimensional chaotic fluctuations. While these fluctuations cannot propagate through the rank-one cross-region couplings (up to small, $O(1/\sqrt{N})$ fluctuations around the mean-field currents), they can disrupt low-dimensional signal transmission between regions, illustrating the tension between signal generation and transmission, Key Idea 1.

Despite the presence of disorder, the symmetric structure of the interactions ensures that the currents converge to fixed points, $S_0^{\mu\nu}$. However, the network's behavior is now controlled not just by the values of a^μ and b^μ , but also by the disorder strength g^μ . This richer dynamical landscape is naturally characterized by the correlation function $\Delta^\mu(t, t')$, which captures, for example, how quickly the network forgets its state at a given time through chaotic mixing. We focus on stationary solutions where $\Delta(t, t') = \Delta(\tau)$, with $\tau = t - t'$. Under these conditions, we can solve the DMFT equations analytically, determining $\Delta^\mu(0)$, $\Delta^\mu(\infty) = \lim_{\tau \rightarrow \infty} \Delta^\mu(\tau)$, and A^μ (Fig. 4 A and B and *SI Appendix*).

The solutions exhibit the following structure, as depicted in Fig. 4 C–E. For small g^μ , high-dimensional fluctuations are absent in region μ , resulting in $\Delta^\mu(\tau) = \Delta^\mu(0) = \Delta^\mu(\infty)$. This constant correlation function indicates that neural activity maintains perfect memory of its state, reflecting purely structured, nonchaotic dynamics. Routing and nonrouting modes behave as in the disorder-free case (Eqs. 16–18), with current stability determined by the relative magnitudes of a^μ and b^μ . Here, we assume that $b^\mu > a^\mu$ so that, without disorder, all regions are in routing mode (the behavior we will describe as disorder is increased is similar for $b^\mu < a^\mu$, but with changes to self-current rather than cross-region current).

This nonchaotic regime persists even for $g^\mu > 1$, demonstrating that currents from within the region (nonrouting mode) or from other regions (routing mode) can suppress chaos. However, compared to the disorder-free case, A^μ is reduced, indicating that disorder impedes currents. As g^μ increases further, a phase transition occurs. High-dimensional fluctuations begin to coexist with currents, characterized by $\Delta^\mu(\infty) < \Delta^\mu(0)$ and a decaying $\Delta^\mu(\tau)$. The decay of $\Delta^\mu(\tau)$ to a nonzero value $\Delta^\mu(\infty)$ indicates that the network partially forgets its state through chaotic mixing, while maintaining some structure through the persistent currents. In this regime, A^μ decreases even more.

At sufficiently large g^μ , another phase transition takes place, leading to a “disorder-dominated” nonrouting mode. Here, $\Delta^\mu(\tau)$ decays from $\Delta^\mu(0) > 0$ to $\Delta^\mu(\infty) = 0$, and $A^\mu = 0$.

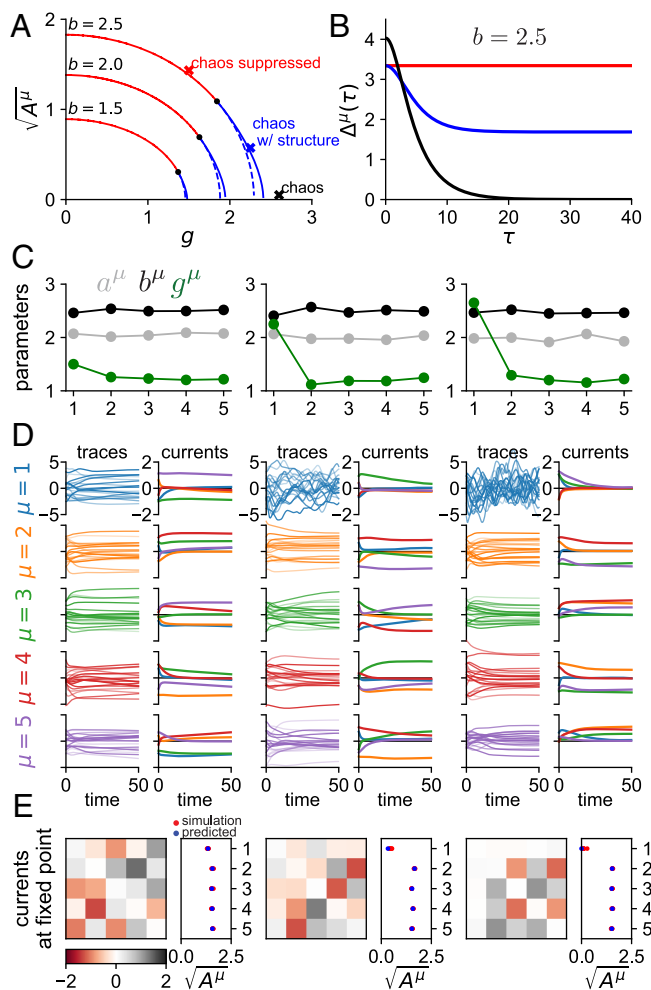


Fig. 4. Structure of activity in networks with disorder and symmetric effective interactions among regions. (A) Relationship between A^μ and g^μ for various values of b^μ in the DMFT. Dashed lines indicate nonphysical solutions of the DMFT equations corresponding to unstable fixed points. (B) Solutions for the two-point function $\Delta^\mu(\tau)$ for the parameter values indicated by the markers in (A). (C–E) are the same as (A–C) in Fig. 3, but with disorder, whose levels are shown in (A). All regions have $g^\mu > 1$, so regions produce high-dimensional fluctuations unless tamed by current-based activity. In the leftmost plots, chaos is suppressed in all regions, and all regions are in routing mode. In the *Middle* plots, all regions are in routing mode, and high-dimensional fluctuations exist alongside the structured current-based activity in region 1. In the rightmost plots, region 1 is in disorder-dominated nonrouting mode, and regions 2 to 5 are in routing mode. In chaotic regimes (*Middle* and *Right* columns), the interregion currents converge to steady values despite ongoing chaotic dynamics. This convergence occurs because the readout patterns project out the chaotic fluctuations, though small $O(1/\sqrt{N})$ fluctuations remain around the mean-field values.

The complete decay of the correlation function indicates that the network completely forgets its state at any given time, reflecting fully chaotic dynamics with no underlying structure. The values of ψ_0^μ and $\Delta^\mu(\tau)$ are no longer influenced by a^μ and b^μ . Instead, $\Delta^\mu(\tau)$ follows the solution described by Sompolinsky et al. (29), as if no structured connectivity were present. This disorder-dominated phase differs from the “structure-dominated” nonrouting mode of the disorder-free case in a crucial way: signal transmission from other regions is impeded by high-dimensional fluctuations rather than structured self-exciting activity, resulting in $S_0^{\mu\mu} = 0$.

Importantly, these disorder-induced phase transitions occur independently across regions, a consequence of the low-rank

structure of cross-region connectivity preventing the propagation of high-dimensional fluctuations.

To summarize, the behavior of $\Delta^\mu(\tau)$ reveals how network activity aligns with different subspaces: when $\Delta^\mu(\tau)$ is constant, activity lies in structured subspaces defined by currents; when it decays to a nonzero value, activity combines both current-based structure and chaotic components; and when it decays to zero, activity explores all dimensions chaotically. This progression illustrates Key Idea 2: signal routing is achieved not by silencing regions, but by controlling which subspaces of activity are excited or suppressed through the interplay of connectivity and dynamics.

Asymmetric Effective Interactions

We now relax all constraints on the effective interactions, including symmetry, allowing $T^{\mu\nu\rho}$ to have arbitrary elements. This can lead to a richer set of dynamic behaviors in the network. To analyze these dynamics, we focus on the spectrum of $\hat{T}^{\mu\nu,\rho\sigma}$, the matrix representation of $T^{\mu\nu\rho}$.

The leading eigenvalue of $\hat{T}^{\mu\nu,\rho\sigma}$ strongly influences the network's behavior. When this eigenvalue is real, the currents typically converge to fixed points. In contrast, a complex-conjugate pair of leading eigenvalues, especially with a large imaginary part, often results in limit cycles in the currents. We have not observed chaotic attractors in the currents.

To characterize the interplay between current dynamics, within-region high-dimensional fluctuations, and the leading eigenvalue of $\hat{T}^{\mu\nu,\rho\sigma}$, we conducted a comprehensive analysis. We focused on networks with $R = 2$ regions, setting disorder levels $g^1 = g^2 = 1.5$. For each complex number λ on a grid in the upper half-plane, we generated 50 random effective-interaction tensors $T^{\mu\nu\rho}$ whose associated matrix $\hat{T}^{\mu\nu,\rho\sigma}$ had λ as its leading eigenvalue. For each tensor, we numerically solved the DMFT equations to obtain the two-point functions $\Delta^\mu(t, t')$ and currents $S^{\mu\nu}(t)$. We then analyzed the normalized two-point function:

$$\hat{\Delta}^\mu(\tau) = \frac{\Delta^\mu(t, t + \tau)}{\sqrt{\Delta^\mu(t)\Delta^\mu(t + \tau)}}, \quad t \gg 1, \quad [21]$$

where t is large enough to disregard transients. The behavior of $\hat{\Delta}^\mu(\tau)$ indicates the presence and nature of high-dimensional fluctuations in region μ . In particular, similar to the interpre-

tation of $\Delta^\mu(\tau)$ in the previous section, when $\hat{\Delta}^\mu(\tau)$ decays to a nonzero value, region μ displays chaotic fluctuations with underlying structure due to currents providing order-one mean activity. This structure can also be seen in the currents themselves. Conversely, $\hat{\Delta}^\mu(\tau)$ decaying to zero indicates that there are only chaotic fluctuations in region μ .

Fig. 5 summarizes our findings. As the real part of λ increases with a small imaginary part, we observe a progression from pure chaos, to fixed points coexisting with chaos, to pure fixed points (Fig. 5A and C). Strikingly, when the imaginary part of λ is larger, we see a parallel series of transitions: from chaos, to limit cycles coexisting with chaos, to pure limit cycles. The coexistence of limit cycles with high-dimensional fluctuations is particularly intriguing, as it demonstrates that reliable, time-dependent routing can occur beneath apparently noisy activity.

The dashed circle in Fig. 5A indicates the support of the bulk spectrum of $J_{ij}^{\mu\nu}$. For nontrivial current dynamics to emerge, the leading eigenvalue of $\hat{T}^{\mu\nu,\rho\sigma}$ must lie outside this circle. This illustrates how high-dimensional fluctuations within regions (the bulk) can impede structured cross-region communication (the outlier), highlighting the tension between signal generation and transmission (Key Idea 1).

To assess the predictive power of the leading eigenvalue, we computed the entropy of the empirical distribution over the five possible dynamic states at each λ (Fig. 5B). For large imaginary parts of λ , we observe a reliable transition from chaos to limit cycles coexisting with high-dimensional fluctuations as the real part increases, with a critical value near $\text{Re}\lambda = 1.5$. In regions where pure fixed points or limit cycles dominate, the behavior becomes more variable, especially where different states intermingle.

We next explored how modulating disorder can shape multiregion dynamics and signal routing. Fig. 6 shows two cases with fixed $T^{\mu\nu\rho}$ in networks of $R = 3$ regions. In both cases, introducing disorder in region 1 switched the current dynamics from fixed points to limit cycles. Importantly, this transition did not occur by silencing region 1; instead, the gains of all regions remained of order unity throughout the transition (Fig. 6C). This supports Key Idea 2, demonstrating that signal routing is achieved by shaping the alignment of neural activity with particular subspaces, rather than through traditional gain modulation methods.

To further understand time-dependent signal routing, we analyzed the spectrum of $\psi^\nu(t)\hat{T}^{\mu\nu,\rho\sigma}$ across time (Fig. 6D).

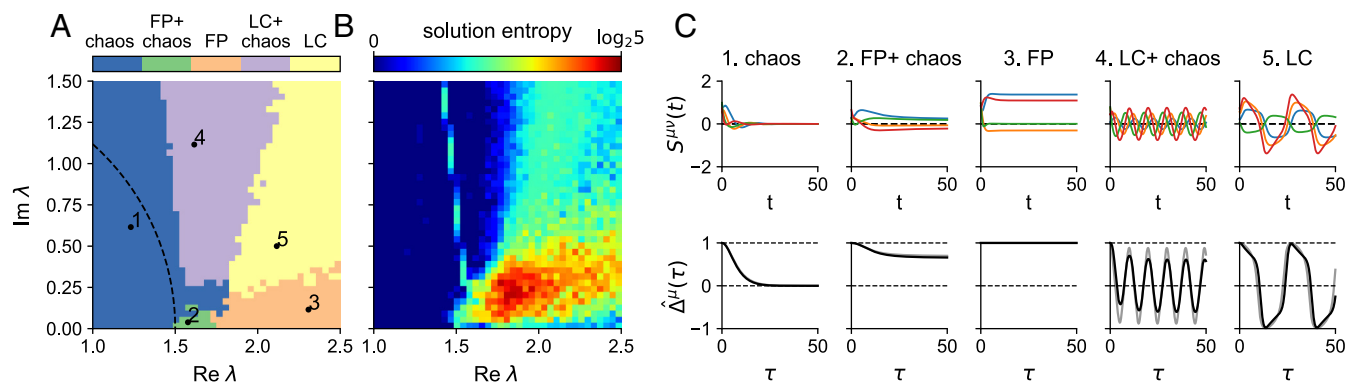


Fig. 5. Dynamic behaviors in networks with asymmetric effective interactions ($R = 2$ regions). (A) Most common dynamic behavior across 50 realizations of $T^{\mu\nu\rho}$, as a function of the leading eigenvalue λ of $\hat{T}^{\mu\nu,\rho\sigma}$. (B) Entropy of the distribution over dynamic behaviors at each λ . (C) Example time series of currents $S^{\mu\nu}$ (Top) and two-point functions $\hat{\Delta}^\mu(\tau)$ (Bottom) for each dynamic behavior. In the Top row, colors represent different currents; in the Bottom row, black and gray lines represent the two regions.

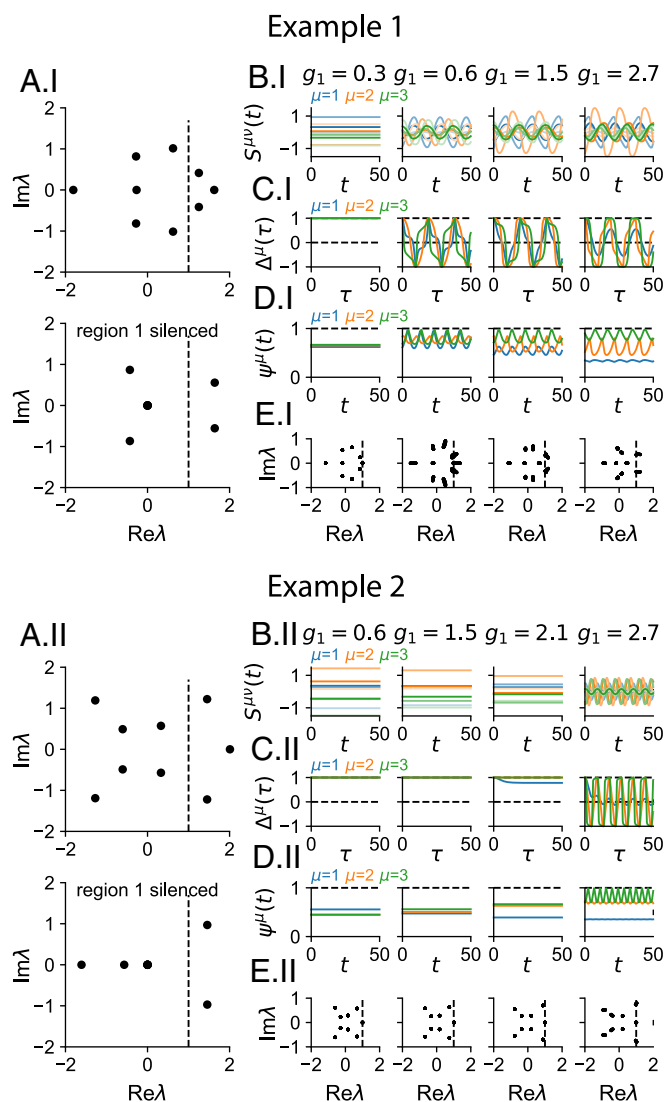


Fig. 6. Modulating multiregion dynamics through disorder in a 3-region network. Two examples (1 and 2) show how introducing disorder in region 1 switches current dynamics from fixed points to limit cycles. (A) Spectra of $\hat{\gamma}^{\mu\nu,\rho\sigma}$ before (I) and after (II) silencing region 1. The resulting switch from real to complex-conjugate pair of the leading eigenvalue suggests that introducing disorder in region 1 will generate limit cycles. (B) Time evolution of currents $S^{\mu\nu}$, with colors indicating the target region μ . (C) Normalized two-point functions $\hat{\Delta}^{\mu}(\tau)$ for increasing disorder g^1 in region 1. (D) Time-dependent gains $\psi^{\mu}(t)$. (E) Time-evolving spectra of $\psi^{\nu}(t)\hat{\gamma}^{\mu\nu,\rho\sigma}$, showing how the eigenvalue distribution changes throughout the limit cycle.

During limit cycles, the leading eigenvalues hover around unity, indicating that current dynamics are regulated through sequential subspace activation and subtle gain adjustments.

These findings demonstrate that in both fixed-point and dynamic attractor scenarios, adjusting effective interactions or disorder levels can shift signal routing through the network. This routing occurs not by silencing entire regions, but by altering which subspaces are active, leading to phase transitions in current dynamics while maintaining nonzero gains. This mechanism aligns with both Key Ideas 1 and 2, highlighting the tension between signal generation and transmission and emphasizing the role of subspace activation in controlling signal flow.

Input-Driven Switches

Our model shows that a region's ability to transmit signals depends on the balance between its within-region activity and

cross-region communication, as described in Key Idea 1. While this balance can be modified by adjusting synaptic couplings, as demonstrated in the previous sections, external inputs offer an alternative method for controlling routing that is more amenable to experimental probing (22).

We extended the DMFT to incorporate inputs, introducing effective interactions that capture overlaps between recurrent connectivity and input vectors (*SI Appendix*). To illustrate this, we examined a simple example with 5 regions. Initially, region 1 exhibits strong self-exciting activity and does not route signals. When we add input to region 1 that other regions can read out and feed back, it transitions to a state where region 1 communicates with the network and its self-exciting activity is suppressed. This input-driven switch mirrors the connectivity-based switches studied earlier and exemplifies one of many possible scenarios for input-based activity modulation.

The specific effects of inputs depend on the multiregion connectivity geometry encoded in $T^{\mu\nu\rho}$. Experimentally, inputs could be provided to a region using techniques like optogenetics. Given knowledge of cross-region subspace geometry, one could predict resulting network-level activity changes. This geometry could be estimated using methods similar to those developed by Semedo et al. (39).

Discussion

In this work, we focused on rank-one communication subspaces with jointly Gaussian loadings. This connectivity provides a starting point for studying more complicated forms of communication between areas. For example, we can extend our rank-one connectivity model to rank- K subspaces, facilitating richer, higher-dimensional communication. Maintaining the ranks of these subspaces as intensive prevents high-dimensional chaotic fluctuations from propagating between regions, preserving the modularity of the disorder-based gating mechanism. While increasing the rank increases the number of dynamic variables in the mean-field picture (namely, by a factor of K), the Gaussian distribution determining the loadings restricts the complexity of their effective interactions. An alternative is to use a mixture-of-Gaussians distribution with C components, allowing for more complex interactions, such as chaotic dynamics among the currents (35, 48). Together, these extensions expand the effective-interaction tensor by three indices, detailed in a tensor diagram in *SI Appendix*. Finally, an important future direction will be to incorporate biological constraints, such as excitatory and inhibitory neurons and nonnegative firing rates. The work of (30) is a promising starting point.

How might the connectivity geometry defining $T^{\mu\nu\rho}$ be established? We propose that this structure could emerge through the pressures of a learning process. Consider a region μ that needs to perform a computation based on a one-dimensional signal from region ν . In this case, establishing a rank-one cross-region coupling matrix $\mathbf{m}^{\mu\nu}(\mathbf{n}^{\mu\nu})^T$, which could occur through Hebbian plasticity, is sufficient. The preactivations in ν lie within the subspace $S^{\nu} = \text{span}\{\mathbf{n}^{\nu\rho}\}_{\rho=1}^R$. For μ to use a signal from ν , the row space spanned by $\mathbf{n}^{\mu\nu}$ must then overlap with S^{ν} . This overlap implies that $T^{\mu\nu\rho} = N^{-1}(\mathbf{n}^{\mu\nu})^T \mathbf{m}^{\nu\rho} \neq 0$ for at least one ρ . This simplified picture of learning neglects the fact that regions are connected in loops. Future research is required to explore how regions learn tasks in a recurrently connected network, addressing the "multiregion credit assignment" problem.

The question "What defines a brain region?" is, at its essence, about how within-region connectivity differs from

cross-region connectivity. Previous work, such as that by Aljadeff et al. (49), studied networks with disordered couplings both within and between regions, but found that chaotic activity is globally distributed, undermining the notion of distinct regions. In contrast, our model, which uses low-rank cross-region connectivity, leads to rich functional consequences and modular activity states, making it a more interesting candidate framework for regional organization.

The symmetric connectivity geometry we studied, characterized by $c^{\mu\nu}$, has not yet been observed in functional communication-subspace analyses or current connectomics data. However, as larger-scale mammalian connectomes become available in the coming years, it would be valuable to compute observables like $T^{\mu\nu\rho}$. Given its interesting functional consequences, our symmetry-constrained version would be a natural structure to look for, analogous to how researchers have examined correlations between reciprocal synapses in existing datasets.

A notable aspect of our model and theoretical approach is its alignment with existing methods for neural-data analysis. Specifically, the technique developed by Perich et al. (12) for analyzing multiregion neural recordings involves training a recurrent network to mimic the data and then decomposing the activity in terms of cross-region currents. Intriguingly, our model's low-dimensional mean-field dynamics offer a closed description in terms of these currents, rather than relying solely on single-region quantities such as two-point functions. This alignment strongly supports the use of current-based analyses in neural data interpretation.

Furthermore, our model could be adapted to fit multiregion neural data using approaches akin to those of Valente et al. (50). Subsequently reducing the model to the mean-field description we derived could provide insights into the dynamics of the fitted model. This positions our work as a bridge connecting practical recurrent network-based data analysis methods to a deeper analytical understanding of network dynamics.

Another data-driven application of our framework lies in analyzing connectome data (51). Large-scale reconstructions of neurons and their connections are now available for flies (52, 53), parts of the mammalian cortex (54), and other organisms (55). For connectome datasets where regions are identified, the cross-region connectivity could be approximated as having a low-rank structure, allowing for a reduction using our mean-field

framework. This enables a comparison of predicted neuronal dynamics with recorded activity.

In scenarios where regions are not already defined, our framework suggests solving the “inverse problem:” determining a partitioning of neurons into regions such that the cross-region connectivity is well approximated by low-rank matrices. Developing a specialized clustering algorithm for this purpose and applying it to connectome data, such as from the fly, would be interesting. Even in cases where anatomical knowledge suggests certain region definitions, identifying “unsupervised regions” based on the assumption of low-rank cross-region interactions could offer an interesting functional perspective on regional delineation.

Materials and Methods

Additional methods and results are provided in *SI Appendix*. *SI Appendix, section 1* describes the analytical approach to characterizing the spectrum of the connectivity matrix. *SI Appendix, section 2* provides closed-form expressions for key functions in the DMFT analysis. *SI Appendix, section 3* extends the model to include disorder in cross-region couplings. *SI Appendix, section 4* details the mathematical analysis of fixed point stability using a local energy function. *SI Appendix, section 5* characterizes the structure of fixed points using convex geometry. *SI Appendix, section 6* generalizes the DMFT equations to include the effects of disorder. *SI Appendix, section 7* further examines the dynamics of networks with unconstrained effective interactions. *SI Appendix, section 8* extends the analysis to include external inputs and demonstrates input-driven switching behavior. *SI Appendix, section 9* describes a generalization of the model to more complex connectivity structures. *SI Appendix, section 10* situates our model within a broader class of low-rank mixture-of-Gaussians networks, demonstrating its relationship to and distinctions from more general low-rank models.

Data, Materials, and Software Availability. Code for simulations and analysis have been deposited in <https://github.com/davidclark1/MultiregionDMFT> (56).

ACKNOWLEDGMENTS. We are extremely grateful to L. F. Abbott for his advice on this work. We thank Albert J. Wakhloo for comments on the manuscript, as well as Rainer Engelken, Haim Sompolinsky, Ashok Litwin-Kumar, and members of the Litwin-Kumar and Xiao-Jing Wang groups for helpful discussions. D.G.C. was supported by the Kavli Foundation. M.B. was supported by NIH award R01EB029858. The authors were additionally supported by the Gatsby Charitable Foundation GAT3708.

- D. J. Felleman, D. C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
- K. Ito et al., A systematic nomenclature for the insect brain. *Neuron* **81**, 755–765 (2014).
- O. Randlett et al., Whole-brain activity mapping onto a zebrafish brain atlas. *Nat. Methods* **12**, 1039–1046 (2015).
- Q. Wang et al., The Allen mouse brain common coordinate framework: A 3D reference atlas. *Cell* **181**, 936–953 (2020).
- J. J. Jun et al., Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
- T. A. Machado, I. V. Kauvar, K. Deisseroth, Multiregion neuronal activity: The forest and the trees. *Nat. Rev. Neurosci.* **23**, 683–704 (2022).
- J. Manley et al., Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron* **112**, 1694–1709 (2024).
- S. Chen et al., Brain-wide neural activity underlying memory-guided movement. *Cell* **187**, 676–691 (2024).
- N. T. Markov et al., Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex* **21**, 1254–1272 (2011).
- A. S. Ecker et al., State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014).
- I. C. Lin, M. Okun, M. Carandini, K. D. Harris, The nature of shared cortical variability. *Neuron* **87**, 644–656 (2015).
- M. G. Perich et al., Inferring brain-wide interactions using data-constrained recurrent neural network models. *bioRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.12.18.423348> (Accessed 20 February 2025).
- G. Okazawa, R. Kiani, Neural mechanisms that make perceptual decisions flexible. *Annu. Rev. Physiol.* **85**, 191–215 (2023).
- C. Fang, K. L. Stachenfeld, “Predictive auxiliary objectives in deep RL mimic learning in the brain” in *The Twelfth International Conference on Learning Representations* (2024).
- S. Musall, M. T. Kaufman, A. L. Juavinett, S. Gluf, A. K. Churchland, Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
- N. A. Steinmetz, P. Zarka-Haas, M. Carandini, K. D. Harris, Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
- I. Brain Laboratory et al., A brain-wide map of neural activity during complex behaviour. *bioRxiv* [Preprint] (2023). <https://doi.org/10.1101/2023.07.04.547681> (Accessed 20 February 2025).
- E. S. Schaffer et al., The spatial and temporal structure of neural activity across the fly brain. *Nat. Commun.* **14**, 5572 (2023).
- L. Pinto et al., Task-dependent changes in the large-scale dynamics and necessity of cortical regions. *Neuron* **104**, 810–824 (2019).
- J. A. Michaels, S. Schaffelhofer, A. Agudelo-Toro, H. Scherberger, A goal-driven modular neural network predicts parietofrontal neural dynamics during grasping. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32124–32135 (2020).
- G. Chen, B. Kang, J. Lindsey, S. Druckmann, N. Li, Modularity and robustness of frontal cortical networks. *Cell* **184**, 3717–3730 (2021).
- J. Barbosa et al., Early selection of task-relevant features through population gating. *Nat. Commun.* **14**, 6837 (2023).
- A. S. Andalman et al., Neuronal dynamics regulating brain and behavioral state transitions. *Cell* **177**, 970–985.e20 (2019).
- A. Nair et al., An approximate line attractor in the hypothalamus encodes an aggressive state. *Cell* **186**, 178–193 (2023).
- M. G. Perich, K. Rajan, Rethinking brain-wide interactions through multi-region ‘network of networks’ models. *Curr. Opin. Neurobiol.* **65**, 146–151 (2020).

26. F. Mastrogiuseppe, S. Ostoic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**, 609–623 (2018).
27. U. Pereira-Obilinovic, J. Aljadeff, N. Brunel, Forgetting leads to chaos in attractor networks. *Phys. Rev. X* **13**, 011009 (2023).
28. L. Abbott, "Where are the switches on this thing" in *23 Problems in Systems Neuroscience*, J. L. van Hemmen, T. J. Sejnowski, Eds. (Oxford University Press, USA, 2006), pp. 423–431.
29. H. Sompolinsky, A. Crisanti, H. J. Sommers, Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259 (1988).
30. J. Kadmon, H. Sompolinsky, Transition to chaos in random neuronal networks. *Phys. Rev. X* **5**, 041030 (2015).
31. F. Mastrogiuseppe, S. Ostoic, Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *PLoS Comput. Biol.* **13**, e1005498 (2017).
32. E. Herbert, S. Ostoic, The impact of sparsity in low-rank recurrent neural networks. *PLoS Comput. Biol.* **18**, e1010426 (2022).
33. Y. Shao, S. Ostoic, Relating local connectivity and global dynamics in recurrent excitatory-inhibitory networks. *PLoS Comput. Biol.* **19**, e1010855 (2023).
34. J. A. Gallego, M. G. Perich, L. E. Miller, S. A. Solla, Neural manifolds for the control of movement. *Neuron* **94**, 978–984 (2017).
35. M. Beiran, A. Dubreuil, A. Valente, F. Mastrogiuseppe, S. Ostoic, Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Comput.* **33**, 1572–1615 (2021).
36. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982).
37. R. Ben-Yishai, R. L. Bar-Or, H. Sompolinsky, Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3844–3848 (1995).
38. S. Denève, A. Alemi, R. Bourdoukan, The brain as an efficient and robust adaptive learner. *Neuron* **94**, 969–977 (2017).
39. J. D. Smedo, A. Zandvakili, C. K. Machens, M. Y. Byron, A. Kohn, Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259 (2019).
40. J. D. Smedo *et al.*, Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat. Commun.* **13**, 1099 (2022).
41. M. G. Perich *et al.*, Motor cortical dynamics are shaped by multiple distinct subspaces during naturalistic behavior. bioRxiv [Preprint] (2020). <https://doi.org/10.1101/2020.07.30.228767> (Accessed 20 February 2025).
42. S. Kondapavulur *et al.*, Transition from predictable to variable motor cortex and striatal ensemble patterning during behavioral exploration. *Nat. Commun.* **13**, 2450 (2022).
43. R. Srinath, D. A. Ruff, M. R. Cohen, Attention improves information flow between neuronal populations without changing the communication subspace. *Curr. Biol.* **31**, 5299–5313 (2021).
44. C. J. MacDowell, A. Libby, C. I. Jahn, S. Tafazoli, T. J. Buschman, Multiplexed subspaces route neural activity across brain-wide networks. bioRxiv [Preprint] (2023). <https://doi.org/10.1101/2023.02.08.527772> (Accessed 20 February 2025).
45. D. Sussillo, L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**, 544–557 (2009).
46. L. Logiaco, L. Abbott, S. Escola, Thalamic control of cortical dynamics in a model of flexible motor sequencing. *Cell Rep.* **35**, 109090 (2021).
47. D. Hansel, H. Sompolinsky, Solvable model of spatiotemporal chaos. *Phys. Rev. Lett.* **71**, 2710 (1993).
48. A. Dubreuil, A. Valente, M. Beiran, F. Mastrogiuseppe, S. Ostoic, The role of population structure in computations through neural dynamics. *Nat. Neurosci.* **25**, 783–794 (2022).
49. J. Aljadeff, M. Stern, T. Sharpee, Transition to chaos in random networks with cell-type-specific connectivity. *Phys. Rev. Lett.* **114**, 088101 (2015).
50. A. Valente, J. W. Pillow, S. Ostoic, Extracting computational mechanisms from neural data using low-rank RNNs. *Adv. Neural Inf. Process. Syst.* **35**, 24072–24086 (2022).
51. L. F. Abbott *et al.*, The mind of a mouse. *Cell* **182**, 1372–1376 (2020).
52. Z. Zheng *et al.*, A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell* **174**, 730–743 (2018).
53. L. K. Scheffer *et al.*, A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, e57443 (2020).
54. J. Winnubst *et al.*, Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell* **179**, 268–281 (2019).
55. D. G. C. Hildebrand *et al.*, Whole-brain serial-section electron microscopy in larval zebrafish. *Nature* **545**, 345–349 (2017).
56. D. Clark *et al.*, MultiregionDMFT. Github. <https://github.com/davidclark1/MultiregionDMFT>. Deposited 20 February 2024.