
Symmetries and continuous attractors in disordered neural circuits

In the format provided by the
authors and unedited

Supplementary Information

Symmetries and continuous attractors in disordered neural circuits

David G. Clark, L. F. Abbott, Haim Sompolinsky

Contents

Notation and conventions	2
1 List of symbols	2
2 Fourier conventions and identities	6
Supplementary Note 1: General theory of low-rank networks	10
3 Symmetries, degeneracies, and dynamics in low-rank recurrent neural networks	10
4 Existence of quasicontinuous attractors with arbitrary manifolds via universal approximation	15
Supplementary Note 2: Large-N theory of the model	17
5 Details of large- N analysis of optimized weights	17
6 Details of dynamical mean-field theory	20
7 Fourier coefficients of correlation functions	29
Supplementary Note 3: Model construction and stability	31
8 Details of data-derived network model	31
9 Structural and global stability	35
10 Additional figures	38
Supplementary Note 4: Dynamics of learned attractors	40
11 Introduction	40
12 Problem formulation	41
13 Order parameters	42
14 Single-site dynamics	42
15 Differential equations for the order parameters	43
16 Linearization around fixed points	43
17 Gaussian feedback weights	46
18 Taxonomy of ring-attractor models	47
19 Path-integral derivation	49

Notation and conventions

1	List of symbols	2
2	Fourier conventions and identities	6
2.1	Fourier series on the circle	6
2.2	Higher dimensions	7
2.3	Fourier transform on the real line	7
2.4	Discretization	7
2.5	Convolution	8
2.6	Real sine-cosine basis	8
2.7	Wrapped functions	9

1 List of symbols

The tables below list the principal symbols used in the main text and this Supplementary Information, together with the principal places each one appears. Main-text and Methods sections, which are unnumbered, are cited by title. Fourier-transform conventions follow Sec. 2. Supplementary Note 4 uses its own notation, described at its start.

Network model and target manifold

Symbol	Meaning	Introduced in
$x_i(t), \phi_i(t)$	Input current and firing rate of neuron i at time t	main-text Eq. 1, ‘Constructing a data-derived network model’
$\phi(x)$	Single-neuron activation function, softplus or error function	Methods Eq. 18, ‘Model settings’
τ	Network time constant	main-text Eq. 1, ‘Constructing a data-derived network model’; ‘Simulation details’
N	Number of neurons	main-text Eq. 1, ‘Constructing a data-derived network model’
$J_{ij}, \mathbf{J}$	Weight from neuron j to neuron i , and the full $N \times N$ matrix	main-text Eq. 1, ‘Constructing a data-derived network model’; ‘Computing the weight matrix’, ‘Weights in the large-network limit’
b	Bias current in network dynamics; in Setting 2, the soft threshold of the activation	main-text Eq. 1, ‘Constructing a data-derived network model’; ‘Model settings’, ‘Statistical generative process’
$Z[x^\star]$	Normalization functional fixing unit mean firing rate over θ	‘Statistical generative process’, ‘Model settings’
c	Uniform inhibition strength for mean-mode stabilization	‘Model settings’; Secs. 8.3, 8.4

continued on the next page

Network model and target manifold (continued)

Symbol	Meaning	Introduced in
θ	Encoded angle (head direction), parameterizing the target manifold	‘Constructing a data-derived network model’, ‘Statistical generative process’
$x_i^*(\theta), \phi_i^*(\theta)$	Target input current and target firing rate of neuron i	‘Constructing a data-derived network model’, ‘Computing the weight matrix’
ψ	Angular center of the activity bump	main-text Eq. 12, ‘Dynamical mean-field theory’; Secs. 3.3, 6.7
$\omega(t)$	Angular velocity input	‘Angular velocity integration’; Sec. 6.2

Data, optimization, and model–data comparison

Symbol	Meaning	Introduced in
N_{data}	Number of head-direction cells in the dataset, $N_{\text{data}} = 1,533$	‘Statistical generative process’; Sec. 8.1
N_θ	Number of angular bins	‘Weights in the large-network limit’; Secs. 2, 5.1
σ_{smooth}	Width of the Gaussian filter used to smooth tuning curves	Methods Eq. 13, ‘Data processing’
$\mathcal{L}_{\text{Pois},i}$	Cross-validated Poisson log-likelihood used to select the smoothing width	Methods Eq. 13, ‘Data processing’
θ_i^{COM}	Circular center of mass of neuron i ’s tuning curve	‘Tuning-curve statistics’; Sec. 5.4
z_{thresh}	Threshold, in standard deviations, for counting tuning-curve peaks	‘Tuning-curve statistics’, ‘Statistical generative process’
I_{HD}	Head-direction information content	‘Tuning-curve statistics’, ‘Statistical generative process’
ρ_i^{flip}	Flip-symmetry correlation of neuron i ’s tuning curve	‘Tuning-curve statistics’, ‘Statistical generative process’
$\mathbf{X}, \Phi$	Matrices of target currents and target firing rates, $N_\theta \times N$	‘Weights in the large-network limit’, ‘Computing the weight matrix’; Sec. 5.1
$\mathcal{L}(\mathbf{J})$	Loss function for the weight optimization	main-text Eq. 2, ‘Constructing a data-derived network model’; ‘Computing the weight matrix’
λ	Regularization parameter in the loss, $\lambda = 10^{-8}$ unless noted	main-text Eq. 2, ‘Constructing a data-derived network model’; ‘Computing the weight matrix’; Sec. 8.4

continued on the next page

Data, optimization, and model–data comparison (continued)

Symbol	Meaning	Introduced in
$C^\phi(\theta, \theta')$	Two-point correlation function of the target firing rates, the finite- N estimate of Γ^ϕ	main-text Eq. 4, ‘Constructing a data-derived network model’
$K_{i,\lambda}(\theta, \theta')$	Optimization kernel, omitting the i -th regressor to enforce $J_{ii} = 0$	main-text Eq. 3, ‘Constructing a data-derived network model’; ‘Computing the weight matrix’
σ_{obs}	Observation-noise standard deviation in the maximum <i>a posteriori</i> decoder	Sec. 8.5

Generative process, order parameters, and mean-field theory

Symbol	Meaning	Introduced in
$\Gamma^x(\Delta\theta), \Gamma^\phi(\Delta\theta)$	Two-point correlation functions of the input currents and of the firing rates; discretized forms $\mathbf{\Gamma}^x, \mathbf{\Gamma}^\phi$ are $N_\theta \times N_\theta$	‘Statistical generative process’; Secs. 6.1, 7
σ	Fourier decay scale of $\Gamma^x(\Delta\theta)$	‘Statistical generative process’, ‘Model settings’
β	Sharpness parameter of the activation function	Methods Eq. 18, ‘Model settings’
$m(\theta, t), \tilde{m}(\theta, t)$	Overlap order parameter, and its low-pass filtered form	main-text Eq. 5, ‘Dynamics of the data-derived network model’; main-text Eq. 10, ‘Dynamical mean-field theory’; Secs. 6.1, 6.2
$\mathcal{J}(\Delta\theta)$	Effective interactions in the mean-field theory, with Fourier coefficients $\hat{\mathcal{J}}_n$	main-text Eq. 10, ‘Dynamical mean-field theory’; main-text Eq. 9, ‘Weights in the large-network limit’; Sec. 6.1
$Q(t)$	Activity-dependent gain, a quadratic functional of $\tilde{m}(\theta, t)$	main-text Eq. 11, ‘Dynamical mean-field theory’; Secs. 6.1, 6.3
$\mathcal{U}(\Delta\theta)$	Gain kernel entering $Q(t)$; velocity-modified form $\mathcal{U}_{\omega(t)}(\Delta\theta)$	Secs. 6.1, 6.2
$\mathcal{G}(C_{11}, C_{22}, C_{12})$	Gaussian expectation of $\phi(x_1)\phi(x_2)$ for a zero-mean pair with the stated covariance	Eq. 87, Sec. 6.1; Secs. 6.3, 6.6
$G_Q(h)$	Effective activation function of the mean-field equations	Eq. 90, Sec. 6.1; ‘Dynamical mean-field theory’
α	Average slope of the activation function over the input-current distribution	Eq. 69, Sec. 5.2
$M(\theta, \theta')$	Jacobian for the stability analysis	‘Dynamical mean-field theory’; Sec. 6.3

Fourier and spectral structure

Symbol	Meaning	Introduced in
n	Fourier mode index, $n \in \mathbb{Z}$, truncated to $n = -D, \dots, D$	Sec. 2
D	Frequency cutoff, giving $2D + 1$ retained modes	‘Weights in the large-network limit’; Sec. 3.3
$\hat{\Gamma}_n^x, \hat{\Gamma}_n^\phi$	Fourier coefficients of $\Gamma^x(\Delta\theta)$ and $\Gamma^\phi(\Delta\theta)$, doublet-degenerate	main-text Eq. 9, ‘Weights in the large-network limit’; Eq. 27, Sec. 2; Sec. 5.2
$\hat{\mathcal{J}}_n$	Fourier coefficients of the effective interactions	main-text Eq. 9, ‘Weights in the large-network limit’
λ_n^J	Eigenvalues of $\mathbf{J}$	main-text Eq. 9, ‘Weights in the large-network limit’; Secs. 3.3, 3.4
λ_1^M, λ_2^M	Leading and second-largest eigenvalues of the Jacobian $M(\theta, \theta')$	‘Dynamical mean-field theory’; Sec. 6.3
$\mathbf{F}$	Matrix of Fourier modes in the real sine-cosine basis, $N_\theta \times (2D + 1)$	Secs. 5.1, 5.2
$\mathbf{U}^x, \mathbf{U}^\phi$	Left singular vector matrices of $\mathbf{X}$ and $\mathbf{\Phi}$	main-text Eq. 7, ‘Weights in the large-network limit’; Secs. 5.1, 5.2
$\mathbf{V}^x, \mathbf{V}^\phi$	Right singular vector matrices of $\mathbf{X}$ and $\mathbf{\Phi}$; rows hold one neuron’s tuning-curve Fourier coefficients	main-text Eq. 7, ‘Weights in the large-network limit’; Secs. 5.1, 5.2
$\mathbf{S}^x, \mathbf{S}^\phi$	Singular value vectors of $\mathbf{X}$ and $\mathbf{\Phi}$	Eq. 55, Sec. 5.1; Sec. 5.2
r_{eff}	Effective rank of the null-model singular-value envelope	Sec. 5.4

Low-rank representation

Symbol	Meaning	Introduced in
$\mathbf{U}, \mathbf{V}$	Loading matrices of the low-rank form, both $N \times K$	Eq. 29, Sec. 3; Secs. 3.3, 4
K	Rank of the low-rank form, equal to $2D + 1$ for our model	Eq. 29, Sec. 3
$\mathbf{u}, \mathbf{v}$	One neuron’s rows of $\mathbf{U}$ and $\mathbf{V}$, drawn from the joint distribution $P(\mathbf{u}, \mathbf{v})$	Eq. 29, Sec. 3; Secs. 3.1, 3.3
$\mathbf{C}$	Cross-covariance of the loadings $\mathbf{u}$ and $\mathbf{v}$	Eq. 43, Sec. 3.3; Sec. 3.2
$\kappa(t), \tilde{\kappa}(t)$	Overlap order parameters of the low-rank representation, and their low-pass filtered forms	Eq. 31, Sec. 3; Secs. 3.1, 3.4
$\mathbf{R}(\psi)$	Rotation acting on the loadings, under which $P(\mathbf{u}, \mathbf{v})$ is invariant	Secs. 3.1, 3.3

Grid-cell extension

Symbol	Meaning	Introduced in
$\boldsymbol{\theta} = (\theta_1, \theta_2)$	Two-dimensional angular (toroidal) coordinate	Eq. 128, Sec. 7; ‘Grid cells’
$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$	Two-dimensional cartesian coordinate	Sec. 7
$\mathbf{H}$	Hexagonal-lattice matrix, whose columns are the lattice axes	Eq. 129, Sec. 7
$\mathbf{v}(t)$	Two-dimensional cartesian velocity	Sec. 7
$\mathbf{p} = (p_1, p_2)$	Lattice wrapping vector of the two-dimensional wrapped Gaussian	Eq. 128, Sec. 7
$\mathbf{n} = (n_1, n_2)$	Two-dimensional Fourier index	Secs. 2, 7
η	Periodicity strength, $\eta \in [0, 1]$, exactly periodic at $\eta = 1$	Eq. 130, Sec. 7; ‘Grid cells’
ν	Number of grid periods spanning 2π , giving grid period $2\pi/\nu$	Eq. 128, Sec. 7; ‘Grid cells’
$\mathcal{J}(\mathbf{s})$	Two-dimensional effective interactions, with Fourier coefficients $\hat{\mathcal{J}}_{\mathbf{n}}$	Sec. 7

2 Fourier conventions and identities

This section fixes our Fourier and discretization conventions and states the identities used elsewhere in the paper. Throughout this section, $F(\theta)$ is a generic function, and $K(\theta, \theta')$ is a generic kernel, where $\theta, \theta' \in [0, 2\pi)$. For a translation-invariant kernel, we write $K(\Delta\theta)$ where $\Delta\theta = \theta - \theta'$. For a discretized kernel, we write $\mathbf{K}$, an $N_\theta \times N_\theta$ matrix.

2.1 Fourier series on the circle

A function $F(\theta)$ of an angle θ has Fourier coefficients $\hat{F}_n$ indexed by the integers n . We use the convention in which the factor $1/2\pi$ appears in the forward transform,

$$\hat{F}_n = \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{-in\theta} F(\theta), \quad (1)$$

$$F(\theta) = \sum_{n=-\infty}^{\infty} e^{in\theta} \hat{F}_n. \quad (2)$$

These two relations are inverses of one another due to the identities

$$\frac{1}{2\pi} \int_0^{2\pi} d\theta e^{-in\theta} = \delta_{n,0}, \quad (3)$$

$$\frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{inx} = \sum_{p=-\infty}^{\infty} \delta(x - 2\pi p). \quad (4)$$

The right-hand side of Eq. 4 is an evenly spaced train of delta functions called a Dirac comb.

With $1/2\pi$ in the forward transform, the coefficients of a delta function are

$$\hat{\delta}_n = \frac{1}{2\pi} \quad \text{for every } n. \quad (5)$$

2.2 Higher dimensions

For a function of d angles $\boldsymbol{\theta}$, the coefficients are indexed by a vector of d integers $\mathbf{n}$, and the forward transform carries $1/(2\pi)^d$,

$$\hat{F}_{\mathbf{n}} = \frac{1}{(2\pi)^d} \int_{[0,2\pi)^d} d^d \boldsymbol{\theta} e^{-i\mathbf{n} \cdot \boldsymbol{\theta}} F(\boldsymbol{\theta}), \quad (6)$$

$$F(\boldsymbol{\theta}) = \sum_{\mathbf{n} \in \mathbb{Z}^d} e^{i\mathbf{n} \cdot \boldsymbol{\theta}} \hat{F}_{\mathbf{n}}. \quad (7)$$

Continuum statements below generally hold in d dimensions with each normalizing factor $1/2\pi$ replaced by $1/(2\pi)^d$, and in the discretized statements each factor of N_θ replaced by N_θ^d . The paper uses $d = 2$ for the grid-cell analysis (main text, ‘Grid cells’; Sec. 7).

2.3 Fourier transform on the real line

For a function $f(x)$ of a real variable x , we use the same normalization convention as above, namely with the factor $1/2\pi$ in the forward transform,

$$\hat{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{-ikx} f(x), \quad (8)$$

$$f(x) = \int_{-\infty}^{\infty} dk e^{ikx} \hat{f}(k). \quad (9)$$

These are inverses of one another due to the identity

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ikx} = \delta(x). \quad (10)$$

2.4 Discretization

It will often be useful to regard two-point correlation functions as matrices, written in bold, once their arguments are discretized. For example, $K(\theta, \theta')$ is a function of angles θ and θ' , and $\mathbf{K}$ is the matrix with elements $K(\theta_a, \theta_b)$, where

$$\theta_a = \frac{2\pi(a-1)}{N_\theta}, \quad a = 1, \dots, N_\theta. \quad (11)$$

We assume throughout that N_θ is large relative to the smoothness scale of tuning curves and correlation functions.

Integrals are discretized in the usual way,

$$\int_0^{2\pi} d\theta F(\theta) \longrightarrow \frac{2\pi}{N_\theta} \sum_{a=1}^{N_\theta} F(\theta_a). \quad (12)$$

Applying Eq. 12 to Eq. 1 gives the discrete form of the coefficients, in which the factors of 2π cancel,

$$\hat{F}_{\mathbf{n}} = \frac{1}{N_\theta} \sum_{a=1}^{N_\theta} e^{-i\mathbf{n} \cdot \boldsymbol{\theta}_a} F(\boldsymbol{\theta}_a). \quad (13)$$

Eq. 12 also fixes how delta functions and inverse kernels are converted, namely,

$$\delta(\theta_a - \theta_b) \longrightarrow \frac{N_\theta}{2\pi} \delta_{ab}, \quad (14)$$

$$K^{-1}(\theta_a, \theta_b) \longrightarrow \left(\frac{N_\theta}{2\pi} \right)^2 [\mathbf{K}^{-1}]_{ab}. \quad (15)$$

The former follows from the definition $\int_0^{2\pi} d\theta \delta(\theta) = 1$, and the latter from the definition of the inverse as satisfying

$$\int_0^{2\pi} d\theta' K^{-1}(\theta, \theta') K(\theta', \theta'') = \delta(\theta - \theta''). \quad (16)$$

2.5 Convolution

Convolution via a kernel corresponds to multiplication of Fourier coefficients; in the discretized and continuum forms, we have

$$G_a = \frac{1}{N_\theta} \sum_{b=1}^{N_\theta} K(\theta_a - \theta_b) F_b \implies \hat{G}_n = \hat{K}_n \hat{F}_n, \quad (17)$$

$$G(\theta) = \frac{1}{2\pi} \int_0^{2\pi} d\theta' K(\theta - \theta') F(\theta') \implies \hat{G}_n = \hat{K}_n \hat{F}_n. \quad (18)$$

Fourier coefficients of an inverse kernel. The integral in Eq. 16 carries no factor of $1/2\pi$, and so equals 2π times the convolution of Eq. 18. Taking coefficients of both sides of Eq. 16, and using Eq. 5 on the right, gives

$$2\pi \widehat{K^{-1}}_n \hat{K}_n = \frac{1}{2\pi}, \quad (19)$$

and thus

$$\widehat{K^{-1}}_n = \frac{1}{(2\pi)^2 \hat{K}_n}. \quad (20)$$

Eigenvalues of a translation-invariant kernel. The $N_\theta \times N_\theta$ matrix $\mathbf{K}$ with elements $K(\theta_a - \theta_b)$ is circulant, and its eigenvectors are the Fourier modes. Applying $\mathbf{K}$ to one such mode gives

$$\sum_{b=1}^{N_\theta} K(\theta_a - \theta_b) e^{in\theta_b} = e^{in\theta_a} \sum_{b=1}^{N_\theta} K(\theta_b) e^{-in\theta_b} \quad (21)$$

$$= N_\theta \hat{K}_n e^{in\theta_a}, \quad (22)$$

the last step being Eq. 13. The eigenvalues of $\mathbf{K}$ are therefore

$$\lambda_n^K = N_\theta \hat{K}_n. \quad (23)$$

2.6 Real sine-cosine basis

For a real function $F(\theta)$, the Fourier coefficients have Hermitian symmetry, $\hat{F}_{-n} = \hat{F}_n^*$. Collecting the two terms of each pair in Eq. 2 gives the equivalent expansion in real modes,

$$F(\theta) = \hat{F}_0 + \sum_{n \geq 1} [a_n \cos(n\theta) + b_n \sin(n\theta)], \quad (24)$$

$$a_n = 2 \operatorname{Re} \hat{F}_n \quad (n \geq 1), \quad (25)$$

$$b_n = -2 \operatorname{Im} \hat{F}_n \quad (n \geq 1). \quad (26)$$

If $F(\theta)$ is also even, its coefficients are real and equal in pairs,

$$\hat{F}_n = \hat{F}_{-n}, \quad (27)$$

so that $b_n = 0$ and $a_n = 2\hat{F}_n$ for every $n \geq 1$.

For a kernel $K(\Delta\theta)$ that is translation-invariant and even, as correlation functions are, the eigenvectors $e^{in\theta}$ and $e^{-in\theta}$ have the same eigenvalue, so every eigenvalue with $n \neq 0$ is doubly degenerate. Any basis of the corresponding two-dimensional space is then a valid eigenbasis, including $(\cos(n\theta), \sin(n\theta))$. In this paper, we generally adopt this basis, one motivation being that when Fourier modes arise as singular vectors of a matrix, they must be real.

The real expansion of $K(\Delta\theta)$ in Eq. 24, treating it as a function of a single variable $\Delta\theta$ rather than as a translation-invariant operator, contains cosines only. This should not be taken to mean that half of the nonconstant eigenvalues vanish; indeed, the evenness is what produces the doublet degeneracy in the first place.

2.7 Wrapped functions

Multiplying Eq. 4 by $f(x + \theta)$ and integrating over x gives the Poisson summation formula,

$$\sum_{n=-\infty}^{\infty} e^{in\theta} \hat{f}(n) = \sum_{p=-\infty}^{\infty} f(\theta + 2\pi p), \quad (28)$$

whose right-hand side is the “wrapped” version of $f(x)$. Comparing with Eq. 2, the Fourier coefficients of a wrapped function are $\hat{f}(n)$, the real-line transform $\hat{f}(k)$ of Eq. 8 evaluated at integer frequencies $k = n$.

Supplementary Note 1: General theory of low-rank networks

This Note collects results about low-rank recurrent networks in general, of which our model is a special case. They concern how eigenvalue degeneracy, continuous symmetry, and continuous attractors are related.

3	Symmetries, degeneracies, and dynamics in low-rank recurrent neural networks	10
3.1	Connectivity and dynamical symmetries	11
3.2	Eigenvalue degeneracies from continuous symmetries	11
3.3	Application to our model	12
3.4	Do eigenvalue degeneracies imply dynamical symmetries?	13
3.5	Eigenvalue degeneracies without continuous symmetry	15
4	Existence of quasicontinuous attractors with arbitrary manifolds via universal approximation	15

3 Symmetries, degeneracies, and dynamics in low-rank recurrent neural networks

The optimal weights (main-text Eq. 3) exhibit approximate low-rank structure, where the number of large eigenvalues is much smaller than the number of neurons N (indeed, with $\lambda \rightarrow 0$ and without the constraint $J_{ii} = 0$, the weights are exactly rank $2D + 1$ if we truncate at D Fourier modes). This situates our model within the framework of low-rank recurrent neural networks, a well-studied class of models whose dynamics admit low-dimensional mean-field descriptions [1–4]. This framework is a natural setting in which to study the relationship between distributional symmetry, eigenvalue degeneracy, and the symmetry of the dynamics.

In a general low-rank recurrent neural network, the weight matrix takes the form

$$\mathbf{J} = \frac{1}{N} \mathbf{U} \mathbf{V}^T, \quad (29)$$

where $\mathbf{U}$ and $\mathbf{V}$ are $N \times K$ matrices with K fixed as $N \rightarrow \infty$. Corresponding rows of $\mathbf{U}$ and $\mathbf{V}$, denoted by $\mathbf{u}$ and $\mathbf{v}$, are sampled i.i.d. across rows from a joint distribution $P(\mathbf{u}, \mathbf{v})$. The model studied in this paper is of this form, as we make explicit in Sec. 3.3.

Low-rank weight matrices yield network dynamics that can be analyzed through a K -dimensional vector of overlap order parameters,

$$\boldsymbol{\kappa}(t) = \frac{1}{N} \mathbf{V}^T \boldsymbol{\phi}(t). \quad (30)$$

Throughout this section the network dynamics are the ones studied in the main text (Eq. 1, with $b = 0$), now driven by a weight matrix of the form of Eq. 29. The vector of input currents is confined to the K -dimensional subspace spanned by the columns of $\mathbf{U}$ (any initial component outside of this subspace decays away). The dynamics close in this subspace and are described by $\boldsymbol{\kappa}(t)$.

For low-pass filtered overlaps $\tilde{\boldsymbol{\kappa}}(t)$ satisfying $\tau \partial_t \tilde{\boldsymbol{\kappa}}(t) = -\tilde{\boldsymbol{\kappa}}(t) + \boldsymbol{\kappa}(t)$, the effective dynamics as $N \rightarrow \infty$ are

$$\tau \partial_t \tilde{\boldsymbol{\kappa}}(t) = -\tilde{\boldsymbol{\kappa}}(t) + \langle \mathbf{v} \phi(\mathbf{u}^T \tilde{\boldsymbol{\kappa}}(t)) \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})}. \quad (31)$$

When specialized to our ring-attractor model, the components of $\tilde{\boldsymbol{\kappa}}(t)$ correspond to the Fourier coefficients of the filtered overlap order parameter $\tilde{m}(\theta, t)$ (main-text Eq. 5), and Eq. 31 is the Fourier-space representation of the mean-field dynamics of main-text Eq. 10, which treats these same dynamics in θ -space (the relationship between these descriptions is discussed further in Sec. 6).

3.1 Connectivity and dynamical symmetries

We now show that invariances in the weight matrix statistics generate corresponding invariances in the effective dynamics.

Suppose the joint distribution $P(\mathbf{u}, \mathbf{v})$ is invariant under transformation by a $K \times K$ orthogonal matrix $\mathbf{R} \in O(K)$

$$P(\mathbf{R}\mathbf{u}, \mathbf{R}\mathbf{v}) = P(\mathbf{u}, \mathbf{v}). \quad (32)$$

To see that this distributional symmetry translates to a symmetry of the overlap dynamics, use the invariance and change variables $(\mathbf{u}', \mathbf{v}') = (\mathbf{R}\mathbf{u}, \mathbf{R}\mathbf{v})$ in the expectation

$$\tau \partial_t \tilde{\boldsymbol{\kappa}}(t) = -\tilde{\boldsymbol{\kappa}}(t) + \langle \mathbf{v} \phi(\mathbf{u}^T \tilde{\boldsymbol{\kappa}}(t)) \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})} \quad (33)$$

$$= -\tilde{\boldsymbol{\kappa}}(t) + \langle \mathbf{v} \phi(\mathbf{u}^T \tilde{\boldsymbol{\kappa}}(t)) \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{R}\mathbf{u}, \mathbf{R}\mathbf{v})} \quad (34)$$

$$= -\tilde{\boldsymbol{\kappa}}(t) + \mathbf{R}^T \langle \mathbf{v} \phi(\mathbf{u}^T \mathbf{R} \tilde{\boldsymbol{\kappa}}(t)) \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})} \quad (35)$$

$$\implies \tau \mathbf{R} \partial_t \tilde{\boldsymbol{\kappa}}(t) = -\mathbf{R} \tilde{\boldsymbol{\kappa}}(t) + \langle \mathbf{v} \phi(\mathbf{u}^T \mathbf{R} \tilde{\boldsymbol{\kappa}}(t)) \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})}. \quad (36)$$

Thus, if $\tilde{\boldsymbol{\kappa}}(t)$ satisfies the effective dynamics Eq. 31, so does $\mathbf{R} \tilde{\boldsymbol{\kappa}}(t)$. The symmetry of the dynamics requires the full joint distribution of loading vectors to be invariant, not merely its second-order statistics.

A particularly interesting structure occurs when the set of orthogonal matrices $\mathbf{R}$ under which $P(\mathbf{u}, \mathbf{v})$ is invariant forms a continuous symmetry, specifically, a representation of some Lie group G . In this case, the dynamics admit the same continuous symmetry. For example, if the system has a stable fixed point $\boldsymbol{\kappa}^*$, then its entire orbit $\{\mathbf{R} \boldsymbol{\kappa}^* \mid \mathbf{R} \in G\}$ forms an attractor manifold.

Pezon et al. [5] established a related result within a framework where neurons have assigned coordinates in a low-dimensional feature space and weight-matrix elements measure feature similarity through a kernel. Our analysis, in contrast, operates directly at the level of loading-vector statistics.

3.2 Eigenvalue degeneracies from continuous symmetries

We now show how continuous symmetries of the weight statistics lead to eigenvalue degeneracy, and why eigenvalue degeneracy does not uniquely determine the symmetry of the dynamics.

The nonzero eigenvalues of the $N \times N$ weight matrix are the same as those of the $K \times K$ loading cross-covariance $\frac{1}{N} \mathbf{U}^T \mathbf{V}$, which as $N \rightarrow \infty$ becomes

$$\frac{1}{N} \mathbf{U}^T \mathbf{V} \xrightarrow{N \rightarrow \infty} \mathbf{C} = \langle \mathbf{u} \mathbf{v}^T \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})}. \quad (37)$$

Note that the spectrum of $\mathbf{J}$ at large N therefore depends only on second-order statistics of the loadings.

The invariance of $P(\mathbf{u}, \mathbf{v})$ under $\mathbf{R}$ implies, by changing variables $(\mathbf{u}', \mathbf{v}') = (\mathbf{R}\mathbf{u}, \mathbf{R}\mathbf{v})$ in the expectation,

$$\mathbf{C} = \langle \mathbf{u} \mathbf{v}^T \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})} \quad (38)$$

$$= \langle \mathbf{u} \mathbf{v}^T \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{R}\mathbf{u}, \mathbf{R}\mathbf{v})} \quad (39)$$

$$= \mathbf{R}^T \langle \mathbf{u} \mathbf{v}^T \rangle_{\mathbf{u}, \mathbf{v} \sim P(\mathbf{u}, \mathbf{v})} \mathbf{R} \quad (40)$$

$$\implies \mathbf{C} = \mathbf{R}^T \mathbf{C} \mathbf{R}. \quad (41)$$

That is, $\mathbf{C}$ commutes with $\mathbf{R}$.

Now suppose that $\mathbf{R}$ is an orthogonal representation of a compact Lie group G (compactness is needed to ensure complete reducibility of the representation in what follows). Its action on $\mathbb{R}^K$ can be decomposed into

invariant subspaces in which it acts with an irreducible representation. In the following statement we assume for simplicity that each irreducible representation is non-repeated, but this is not essential. Labeling the irreducible representations by dimensions $d_1, \dots, d_L$, the relation $\mathbf{C} = \mathbf{R}^T \mathbf{C} \mathbf{R}$ implies that the matrix $\mathbf{C}$ (and hence $\mathbf{J}$) has eigenvalue degeneracies, with each irreducible representation ℓ contributing a set of d_ℓ eigenvalues of equal modulus. Thus, the eigenvalue degeneracies are determined by the dimensions of the irreducible representations that constitute the representation of the Lie group under which the connectivity transforms.

Whether the d_ℓ eigenvalues are equal or only equal in modulus depends on the type of the irreducible representation. For a real-type representation, such as the three-dimensional representation of $SO(3)$ that would describe head direction in three dimensions, the only matrices commuting with the representation are multiples of the identity, so $\mathbf{C}$ acts as a scalar on that subspace and the d_ℓ eigenvalues are real and equal. In contrast, the two-dimensional representations of the circle group $SO(2)$ commute with every 2×2 rotation matrix (it is a complex-type group), not just with the identity, and the eigenvalues can be a complex-conjugate pair rather than a real doublet. This complex case corresponds to a bump traveling at constant angular velocity rather than a stationary one (main text, ‘Dynamical mean-field theory’). In our construction, $\mathbf{C}$ is diagonal (Sec. 3.3), and remains diagonal in generalizations of the construction to other groups, such as the two-torus. More generally, for a stationary attractor, the leading part of the spectrum should be real, so its degenerate eigenvalues should be equal, not just in modulus.

3.3 Application to our model

The ring-attractor model studied in this paper instantiates the above principles. Its loadings, at large N under the generative process (Setting 3; Methods, ‘Model settings’), are

$$\mathbf{U} = \sqrt{N} \mathbf{V}^x, \quad \mathbf{V} = \sqrt{N} \mathbf{V}^\phi \text{diag} \left(\sqrt{\hat{\mathcal{J}}} \right), \quad (42)$$

where the factors of $\sqrt{N}$ make the components of $\mathbf{u}$ unit-variance. The columns of $\mathbf{V}^x$ and $\mathbf{V}^\phi$ run over the sine-cosine basis, namely the constant mode followed by $(\cos(n\theta), \sin(n\theta))$ pairs for $n = 1, \dots, D$, and their rows hold the corresponding Fourier coefficients of a single neuron’s input-current and firing-rate tuning curves. The components of $\hat{\mathcal{J}}$ (main-text Eq. 9) are the $2D + 1$ Fourier coefficients that exhibit doublet degeneracy.

The pairwise covariances are all diagonal, namely

$$\langle \mathbf{u} \mathbf{u}^T \rangle = \mathbf{I}_{2D+1}, \quad \langle \mathbf{v} \mathbf{v}^T \rangle = \text{diag}(\hat{\mathcal{J}}), \quad \mathbf{C} = \langle \mathbf{u} \mathbf{v}^T \rangle = \alpha \text{diag}(\hat{\mathcal{J}}), \quad (43)$$

where α is the average slope of the activation function (Eq. 69). The first two follow from the orthonormality of the columns of $\mathbf{V}^x$ and $\mathbf{V}^\phi$, and the third from Eq. 71. Each doublet block of $\mathbf{C}$ is therefore $\lambda_n^J \mathbf{I}_2$ with $\lambda_n^J = \alpha \hat{\mathcal{J}}_n$, the eigenvalues of $\mathbf{J}$, as expected (main-text Eq. 9).

Components of $\mathbf{u}$ are jointly Gaussian due to the Gaussian process construction. In contrast, $\mathbf{v}$ exhibits higher-order correlations across modes and with modes of $\mathbf{u}$ due to the nonlinear transformation $\phi(x)$ applied in θ -space, making the joint distribution $P(\mathbf{u}, \mathbf{v})$ non-Gaussian. It nonetheless retains a circular $SO(2)$ symmetry inherited from the rotation-invariant input-current Gaussian process.

To define this symmetry concretely, let $\mathbf{R}(\psi)$ be the $(2D + 1) \times (2D + 1)$ matrix

$$\mathbf{R}(\psi) = \begin{pmatrix} 1 & & & \\ & \mathbf{r}(\psi) & & \\ & & \mathbf{r}(2\psi) & \\ & & & \ddots \\ & & & & \mathbf{r}(D\psi) \end{pmatrix}, \quad \mathbf{r}(\psi) = \begin{pmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix}. \quad (44)$$

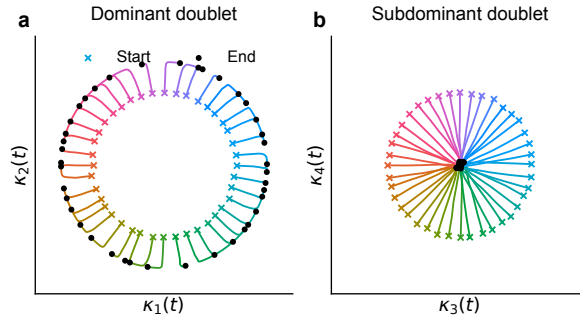
For each angle ψ , the transformation $\mathbf{R}(\psi)$ leaves the distribution invariant

$$P(\mathbf{R}(\psi)\mathbf{u}, \mathbf{R}(\psi)\mathbf{v}) = P(\mathbf{u}, \mathbf{v}). \quad (45)$$

This invariance reflects the fact that rotating the tuning curves $x_i^*(\theta)$ and $\phi_i^*(\theta)$ by an angle ψ does not change the likelihood of observing them. The same ψ must appear in every block. Independent angles in different blocks would give a larger symmetry group, which we return to below.

The matrices $\mathbf{R}(\psi)$ form a representation of $SO(2)$ acting on the $(2D + 1)$ -dimensional space of Fourier coefficients. This representation is reducible and decomposes into a direct sum of $D+1$ irreducible representations, one of them trivial (the constant mode) and D of them two-dimensional. Each two-dimensional irreducible representation contributes a doublet of eigenvalues to the spectrum of $\mathbf{J}$. Having more than one two-dimensional irreducible representation, that is, $D > 1$, is what generates heterogeneity in the tuning curves. With a single one, each neuron would carry one angle and the curves would be shifted copies of one another (see also Sec. 5.3). By the argument of Sec. 3.1, this distributional symmetry generates $SO(2)$ symmetry in the overlap dynamics.

3.4 Do eigenvalue degeneracies imply dynamical symmetries?



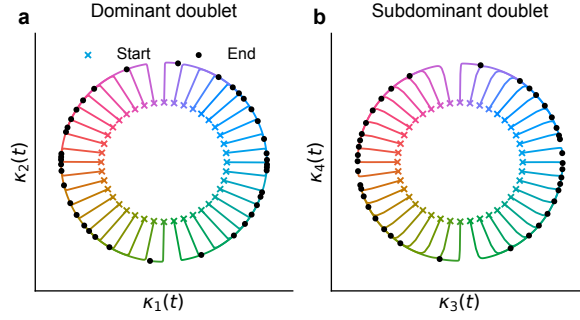
Supplementary Figure 1: Rank-four network with Gaussian loadings, whose distribution has $(SO(2))^2$ symmetry and whose attractor is a single ring ($N = 10,000$, $K = 4$). Forty initial conditions carry varying angles and equal amplitude in both doublets. Trajectories of $\kappa(t)$ run over 10τ , colored by initial angle in the doublet plotted. **a**, Projection onto the dominant doublet, $\kappa_1(t)$ and $\kappa_2(t)$. **b**, Projection onto the subdominant doublet, $\kappa_3(t)$ and $\kappa_4(t)$. Final states form a ring in **a** and collapse to the origin in **b**.

Eigenvalue degeneracies, and eigenvalues in general, depend only on the second-order statistics of the loadings, whereas the symmetry of the dynamics depends on their full joint distribution. Here we show that the spectrum therefore cannot determine the symmetry group. What's more, the symmetry group does not in turn determine the attractor manifold. We establish each point with a modification of our model that preserves the second-order statistics, and therefore the spectrum, while changing the dynamics. Throughout we drop the constant mode, so that $K = 2D$.

Example 1. For the first modification, let $P(\mathbf{u}, \mathbf{v})$ be Gaussian with the same second-order moments as in our actual model (Eq. 43). This Gaussianization preserves the weight-matrix eigenvalues, and in particular their doublet degeneracy. The symmetry group of the dynamics, however, expands from $SO(2)$ to the torus $(SO(2))^D$, which rotates each doublet independently,

$$\mathbf{R}(\psi_1, \dots, \psi_D) = \begin{pmatrix} \mathbf{r}(\psi_1) & & & \\ & \mathbf{r}(\psi_2) & & \\ & & \ddots & \\ & & & \mathbf{r}(\psi_D) \end{pmatrix}. \quad (46)$$

What happens to the dynamics? Because the loadings are Gaussian, the average in Eq. 31 can be evaluated in closed form with Stein's lemma. Writing $\tilde{\kappa}_n(t)$ for the pair of overlaps associated with doublet n , and



Supplementary Figure 2: Rank-four network with mixture-of-Gaussians loadings, whose distribution has the same $(SO(2))^2$ symmetry as Supplementary Fig. 1 but whose attractor is the full torus ($N = 10,000$, $K = 4$). Forty initial conditions carry varying angles and equal amplitude in both doublets. Trajectories of $\kappa(t)$ run over 10τ , colored by initial angle in the doublet plotted. **a**, Projection onto the dominant doublet, $\kappa_1(t)$ and $\kappa_2(t)$. **b**, Projection onto the subdominant doublet, $\kappa_3(t)$ and $\kappa_4(t)$. Final states form a ring in both **a** and **b**.

$\lambda_n^J = \alpha \hat{J}_n$ (Eq. 43) for the associated eigenvalue,

$$\tau \partial_t \tilde{\kappa}_n(t) = -\tilde{\kappa}_n(t) + \lambda_n^J \left\langle \phi'(\sqrt{q(t)}z) \right\rangle_{z \sim \mathcal{N}(0,1)} \tilde{\kappa}_n(t), \quad (47)$$

where $q(t) = \|\tilde{\kappa}(t)\|_2^2$ is the variance of $\mathbf{u}^T \tilde{\kappa}(t)$, a squared norm because $\langle \mathbf{u} \mathbf{u}^T \rangle = \mathbf{I}$.

While the symmetry group of the dynamics is the torus, these effective dynamics show that the attractor manifold is not. In particular, all doublets are linked through a single shared gain, set by $q(t)$, and that shared gain confines the attractor to a single ring. More specifically, at a fixed point each doublet requires either $\tilde{\kappa}_n = \mathbf{0}$ or $\lambda_n^J \langle \phi'(\sqrt{q}z) \rangle_{z \sim \mathcal{N}(0,1)} = 1$. The second condition asks one number, shared by every doublet, to equal $1/\lambda_n^J$, so for generic, distinct λ_n^J it can hold in at most one doublet, and all the others are pinned to zero. A fixed point thus lives in one doublet at a time. Rotating a silent doublet moves nothing, so the orbit of such a point under $(SO(2))^D$ is a single ring rather than a torus, and the fixed points form a union of D disjoint rings, not a product.

To analyze stability, we linearize Eq. 47 about a fixed point active in doublet n . A perturbation in a silent doublet n' obeys $\tau \partial_t \delta \tilde{\kappa}_{n'} = -(1 - \lambda_{n'}^J \langle \phi'(\sqrt{q}z) \rangle_{z \sim \mathcal{N}(0,1)}) \delta \tilde{\kappa}_{n'}$, and the fixed point itself sets $\langle \phi'(\sqrt{q}z) \rangle_{z \sim \mathcal{N}(0,1)} = 1/\lambda_n^J$, so the decay rate is $1 - \lambda_{n'}^J/\lambda_n^J$, which is positive for all $n' \neq n$ if and only if λ_n^J is the largest eigenvalue. Thus, only the ring carrying the largest eigenvalue is stable. Supplementary Fig. 1 shows this in a rank-four network ($K = 4$, $\mathbf{U} = \mathbf{V}$ for simplicity) whose two doublets have loading standard deviations 10 and 5. Initialized with varying angles and equal amplitude in both doublets, $\kappa(t)$ ends up with only its dominant-doublet components. This is a winner-take-all effect of the kind found in networks with Gaussian loadings storing discrete fixed points [2], here generalized to continuous manifolds.

In our actual model, the activation function $\phi(x)$ generates higher-order correlations in $P(\mathbf{u}, \mathbf{v})$ that reduce the symmetry group from $(SO(2))^D$ to $SO(2)$, and the single ring is then the attractor by virtue of this symmetry rather than by competition.

In this example, winner-take-all confines the ring to one doublet, so only a single Fourier mode of the overlap order parameter is ever active. In our model, the $SO(2)$ orbit engages all D doublets at once, which is what makes the tuning curves heterogeneous rather than shifted copies of one sinusoid.

Example 2. In Example 1, a shared gain arising from Gaussianity confined the attractor to a single ring, despite the symmetry group of the dynamics being $(SO(2))^D$. We now depart from Gaussianity so that each

doublet has a gain of its own. The spectrum and the symmetry group are untouched, yet the attractor grows from one ring to the full torus $(SO(2))^D$, the symmetry group of the dynamics.

Let $P(\mathbf{u}, \mathbf{v})$ be a mixture of D Gaussians in which the n -th component carries weight $1/D$ and is supported on doublet n alone, so that each neuron is specialized for a single doublet. The network is simply D non-interacting subnetworks. Within component n the second-order statistics are D times the doublet- n block of Eq. 43, namely $\langle \mathbf{u}_n \mathbf{u}_n^T \rangle = D \mathbf{I}_2$, $\langle \mathbf{v}_n \mathbf{v}_n^T \rangle = D \hat{\mathcal{J}}_n \mathbf{I}_2$ and $\langle \mathbf{u}_n \mathbf{v}_n^T \rangle = D \lambda_n^J \mathbf{I}_2$, with every other element zero. The factor of D compensates the mixture weight, so averaging over components recovers Eq. 43 exactly, and with it the eigenvalues of $\mathbf{J}$. The symmetry group is $(SO(2))^D$.

In place of the shared-gain dynamics of Eq. 47 we obtain

$$\tau \partial_t \tilde{\boldsymbol{\kappa}}_n(t) = -\tilde{\boldsymbol{\kappa}}_n(t) + \lambda_n^J \left\langle \phi'(\sqrt{q_n(t)}z) \right\rangle_{z \sim \mathcal{N}(0,1)} \tilde{\boldsymbol{\kappa}}_n(t), \quad (48)$$

with mode-specific gain $q_n(t) = D \|\tilde{\boldsymbol{\kappa}}_n(t)\|_2^2$. The fixed-point conditions $\lambda_n^J \langle \phi'(\sqrt{q_n}z) \rangle_{z \sim \mathcal{N}(0,1)} = 1$ operate independently, one per doublet. A doublet turns on when $\lambda_n^J \phi'(0) > 1$, and the resulting ring is stable. The attractor manifold is therefore the product of those rings, which is the full D -torus. Supplementary Fig. 2 shows this in a network built to match the one in Supplementary Fig. 1 in rank and in loading covariance. Initialized identically, $\boldsymbol{\kappa}(t)$ retains both angles and relaxes to a ring in each, so the final states fill a ring in both planes.

3.5 Eigenvalue degeneracies without continuous symmetry

In every example so far, the degeneracy was rooted in a continuous symmetry. However, degeneracy can also arise with no continuous symmetry at all. As an example, consider classical Hopfield networks storing P random patterns with i.i.d. components in the limit $N \rightarrow \infty$ [6]. When the patterns are statistically equivalent, the weight matrix exhibits a P -fold eigenvalue degeneracy reflecting the exchangeability of patterns. This degeneracy reflects a functional purpose, though not a continuous attractor: under an appropriate pattern distribution, the P patterns are multistable. However, the spectral structure differs qualitatively from that of continuous-attractor models. In the Hopfield case, there is a single cluster of P degenerate eigenvalues. In continuous-attractor models of the type we have constructed, the spectrum exhibits multiple levels of doublet degeneracy, one for each irreducible representation of the symmetry group. This multilevel doublet structure, rather than degeneracy per se, is the distinguishing spectral signature.

In summary, eigenvalue degeneracies with a multilevel structure are non-generic and point to dynamical symmetry, but do not uniquely specify the symmetry group of the dynamics or the attractor manifold without additional information about higher-order structure.

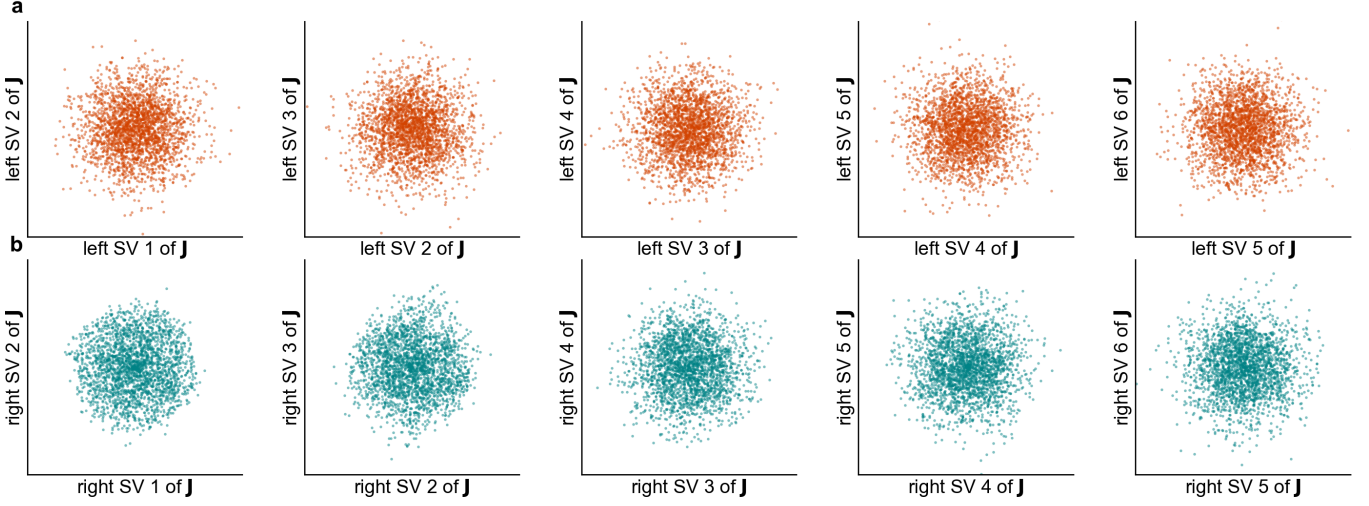
Although the eigenvectors $\mathbf{V}^x$ and $\mathbf{V}^\phi$ of $\mathbf{J}$ appear disordered in scatter plots (Fig. 6f and Supplementary Fig. 3), and thus naive techniques such as spectral embedding do not work, they contain higher-order correlations that encode the $SO(2)$ symmetry. Whether this structure can be reliably extracted from data remains an open question.

4 Existence of quasicontinuous attractors with arbitrary manifolds via universal approximation

Consider dynamics of a K -dimensional vector $\boldsymbol{\kappa}(t)$ governed by

$$\tau \partial_t \boldsymbol{\kappa}(t) = -\boldsymbol{\kappa}(t) + \mathbf{f}(\boldsymbol{\kappa}(t)), \quad (49)$$

where $\mathbf{f}(\boldsymbol{\kappa})$ is an evolution function (including the decay term $-\boldsymbol{\kappa}(t)$ imposes no loss of generality). Suppose these dynamics possess an attractor manifold of dimension $\leq K$, of arbitrary shape and with no assumed symmetry.



Supplementary Figure 3: Left and right singular vectors of the optimal weight matrix $\mathbf{J}$ ($N = 2,500$, generative-process Fourier decay scale $\sigma = 1.42$, sharpness $\beta = 2.77$, error-function activation). **a**, Left singular vectors of $\mathbf{J}$, equal (at large N) to the right singular vectors $\mathbf{V}^x$ of the input-current matrix $\mathbf{X}$. **b**, Right singular vectors of $\mathbf{J}$, equal (at large N) to the right singular vectors $\mathbf{V}^\phi$ of the target firing-rate matrix Φ .

Universal approximation (see, e.g., Chong [7]) guarantees that, for sufficiently large N , one can find $N \times K$ matrices $\mathbf{U}$ and $\mathbf{V}$ and an N -dimensional bias $\mathbf{b}$ such that

$$\mathbf{f}(\boldsymbol{\kappa}) \approx \mathbf{V}^T \phi(\mathbf{U}\boldsymbol{\kappa} + \mathbf{b}) \quad (50)$$

to arbitrary accuracy, uniformly over $\boldsymbol{\kappa}$ on a compact set encompassing the explored state space, provided $\phi(x)$ is non-polynomial. Defining $\mathbf{x}(t) = \mathbf{U}\boldsymbol{\kappa}(t)$ yields

$$\tau \partial_t \mathbf{x}(t) = -\mathbf{x}(t) + \mathbf{U}\mathbf{V}^T \phi(\mathbf{x}(t) + \mathbf{b}), \quad (51)$$

a rank- K recurrent neural network that embeds the original dynamical system in the K -dimensional subspace spanned by the columns of $\mathbf{U}$.

At finite N the approximation is inexact, so states on the resulting manifold drift slowly rather than remaining fixed. The manifold is therefore a quasicontinuous attractor, and the drift vanishes as the approximation improves.

Supplementary Note 2: Large- N theory of the model

This Note develops the theory of our paper’s network model at large N , including both its connectivity and its dynamics. Everything in this Note starts from the generative process of the main text (‘Statistical generative process’).

5	Details of large-N analysis of optimized weights	17
5.1	Singular value decompositions of the tuning-curve matrices	17
5.2	Weight-matrix analysis details	18
5.3	Comparison to cosine-based connectivity	19
5.4	Attempting to extract information about the latent ring structure from the right singular vectors despite their disordered nature	20
6	Details of dynamical mean-field theory	20
6.1	Derivation of dynamical mean-field theory equations	21
6.2	Velocity-dependent dynamics	23
6.3	Jacobian analysis of the mean-field dynamics	24
6.4	Finite domain of effective activation function	25
6.5	Reciprocal symmetry of $\mathbf{J}$ at large N	26
6.6	Emergence of Mexican hat interactions from tuning-curve statistics	26
6.7	Continuous attractors without continuous symmetry	26
7	Fourier coefficients of correlation functions	29

5 Details of large- N analysis of optimized weights

Here we derive the large- N limit of the tuning-curve singular value decompositions and of the least-squares optimal weights, whose results are summarized in the main-text section ‘Weights in the large-network limit’.

5.1 Singular value decompositions of the tuning-curve matrices

At large N , the singular value decompositions of $\mathbf{X}$ and $\mathbf{\Phi}$ converge to Fourier decompositions. The left singular vectors $\mathbf{U}^x$ and $\mathbf{U}^\phi$ are the eigenvectors of the empirical two-point correlation matrices $\frac{1}{N}\mathbf{X}\mathbf{X}^T$ and $\frac{1}{N}\mathbf{\Phi}\mathbf{\Phi}^T$. As $N \rightarrow \infty$ these converge to $\mathbf{\Gamma}^x$ and $\mathbf{\Gamma}^\phi$, the $N_\theta \times N_\theta$ matrices with elements $\Gamma^x(\theta - \theta')$ and $\Gamma^\phi(\theta - \theta')$. Both matrices are circulant, so their eigenvectors are Fourier modes, which we collect in the orthonormal columns of the $N_\theta \times (2D + 1)$ matrix $\mathbf{F}$,

$$F_{a,0} = \frac{1}{\sqrt{N_\theta}}, \quad F_{a,\cos n} = \sqrt{\frac{2}{N_\theta}} \cos(n\theta_a), \quad F_{a,\sin n} = \sqrt{\frac{2}{N_\theta}} \sin(n\theta_a). \quad (52)$$

Thus

$$\mathbf{U}^x \rightarrow \mathbf{F}, \quad \mathbf{U}^\phi \rightarrow \mathbf{F}, \quad (53)$$

with the caveat that the limit is fixed only up to a rotation within each doublet of columns, applied jointly to $\mathbf{U}^x$ and $\mathbf{V}^x$, and likewise to $\mathbf{U}^\phi$ and $\mathbf{V}^\phi$. We adopt the sine-cosine basis above in both cases. Any other basis leaves $\mathbf{J}$ unchanged, and so this sine-cosine basis is simply a convention. This gives

$$(\mathbf{U}^x)^T \mathbf{U}^\phi \rightarrow \mathbf{F}^T \mathbf{F} = \mathbf{I}_{2D+1}. \quad (54)$$

The singular values of $\mathbf{X}$ and $\mathbf{\Phi}$ are $\sqrt{N}$ times the square roots of the eigenvalues of those same matrices. Those eigenvalues converge to $N_\theta \hat{\Gamma}^x$ and $N_\theta \hat{\Gamma}^\phi$ (Sec. 2), so

$$\mathbf{S}^x \rightarrow \sqrt{N N_\theta \hat{\Gamma}^x}, \quad \mathbf{S}^\phi \rightarrow \sqrt{N N_\theta \hat{\Gamma}^\phi}. \quad (55)$$

The right singular vectors are the tuning-curve Fourier coefficients, rescaled mode by mode. Rearranging the decomposition (for any N) gives

$$\mathbf{V}^x = \mathbf{X}^T \mathbf{U}^x (\text{diag}(\mathbf{S}^x))^{-1}, \quad \mathbf{V}^\phi = \mathbf{\Phi}^T \mathbf{U}^\phi (\text{diag}(\mathbf{S}^\phi))^{-1}. \quad (56)$$

Let us consider the x case. With $\mathbf{U}^x \rightarrow \mathbf{F}$, the element of $\mathbf{V}^x$ in row i and the column belonging to $\cos(n\theta)$ is $\sqrt{N_\theta/2} a_{i,n}^x / S_n^x$, where

$$a_{i,0}^x = \frac{1}{2\pi} \int_0^{2\pi} d\theta x_i^*(\theta) = \hat{x}_{i,0}^*, \quad (57)$$

$$a_{i,n}^x = \frac{1}{\pi} \int_0^{2\pi} d\theta \cos(n\theta) x_i^*(\theta) = 2 \text{Re } \hat{x}_{i,n}^* \quad (n \geq 1). \quad (58)$$

With Eq. 55 this gives

$$V_{i,0}^x \rightarrow \frac{a_{i,0}^x}{\sqrt{N \hat{\Gamma}_0^x}}, \quad (59)$$

$$V_{i,\cos n}^x \rightarrow \frac{a_{i,n}^x}{\sqrt{2N \hat{\Gamma}_n^x}} \quad (n \geq 1) \quad (60)$$

and likewise for the sine columns and for $\mathbf{V}^\phi$.

The coefficient $a_{i,n}^x$ has variance $\langle (a_{i,n}^x)^2 \rangle = 2\hat{\Gamma}_n^x$ for $n \geq 1$ and $\langle (a_{i,0}^x)^2 \rangle = \hat{\Gamma}_0^x$ for $n = 0$, so the elements of $\mathbf{V}^x$ have variance $1/N$, consistent with the orthonormality of its columns.

5.2 Weight-matrix analysis details

We now evaluate the weights. At finite N , Methods Eqs. 26 and 27 give

$$\mathbf{J} = \mathbf{X}^T (\mathbf{\Phi} \mathbf{\Phi}^T + N N_\theta \lambda \mathbf{I}_{N_\theta})^{-1} \mathbf{\Phi} \quad (61)$$

$$= \mathbf{V}^x \text{diag}(\mathbf{S}^x) (\mathbf{U}^x)^T \left(\mathbf{U}^\phi \text{diag}(\mathbf{S}^\phi)^2 (\mathbf{U}^\phi)^T + N N_\theta \lambda \mathbf{I}_{N_\theta} \right)^{-1} \mathbf{U}^\phi \text{diag}(\mathbf{S}^\phi) (\mathbf{V}^\phi)^T \quad (62)$$

$$= \mathbf{V}^x \text{diag}(\mathbf{S}^x) (\mathbf{U}^x)^T \mathbf{U}^\phi \text{diag}(\mathbf{S}^\phi) \left(\text{diag}(\mathbf{S}^\phi) (\mathbf{U}^\phi)^T \mathbf{U}^\phi \text{diag}(\mathbf{S}^\phi) + N N_\theta \lambda \mathbf{I}_{2D+1} \right)^{-1} (\mathbf{V}^\phi)^T \quad (63)$$

$$= \mathbf{V}^x \text{diag}(\mathbf{S}^x) (\mathbf{U}^x)^T \mathbf{U}^\phi \text{diag}(\mathbf{S}^\phi) \left(\text{diag}(\mathbf{S}^\phi)^2 + N N_\theta \lambda \mathbf{I}_{2D+1} \right)^{-1} (\mathbf{V}^\phi)^T \quad (64)$$

$$= \mathbf{V}^x \text{diag}(\mathbf{S}^x) (\mathbf{U}^x)^T \mathbf{U}^\phi \text{diag} \left(\frac{\mathbf{S}^\phi}{(\mathbf{S}^\phi)^2 + N N_\theta \lambda} \right) (\mathbf{V}^\phi)^T \quad (65)$$

The steps applied here are (a→b) substitution of SVDs; (b→c) use of the push-through identity (Methods Eq. 25 with $\mathbf{Z} = \mathbf{U}^\phi \text{diag}(\mathbf{S}^\phi)$); (c→d) use of the orthonormality of $\mathbf{U}^\phi$ to simplify $(\mathbf{U}^\phi)^T \mathbf{U}^\phi = \mathbf{I}_{2D+1}$; and (d→e) a cosmetic rearrangement. This expression is exact at any N .

Taking $N \rightarrow \infty$ and inserting the limits established in Sec. 5.1, namely $\mathbf{U}^x, \mathbf{U}^\phi \rightarrow \mathbf{F}$ so that $(\mathbf{U}^x)^T \mathbf{U}^\phi \rightarrow \mathbf{I}_{2D+1}$ (Eq. 54), together with the singular values of Eq. 55, we obtain

$$\mathbf{J} = \mathbf{V}^x \text{diag} \left(\frac{\sqrt{\hat{\Gamma}^x \hat{\Gamma}^\phi}}{\hat{\Gamma}^\phi + \lambda} \right) (\mathbf{V}^\phi)^T. \quad (66)$$

The regularizer drops out as soon as $\lambda \ll \hat{\Gamma}_n^\phi$ for each retained mode, and in particular in the minimum-norm limit $\lambda \rightarrow 0$. This yields the singular value decomposition of $\mathbf{J}$ (same as Eq. 8 in the main text),

$$\mathbf{J} = \mathbf{V}^x \text{diag} \left(\sqrt{\frac{\hat{\Gamma}^x}{\hat{\Gamma}^\phi}} \right) (\mathbf{V}^\phi)^T.$$

To establish the right singular vector overlaps, start from Eq. 56, from which it follows that

$$(\mathbf{V}^\phi)^T \mathbf{V}^x = N \left(\text{diag}(\mathbf{S}^\phi) \right)^{-1} (\mathbf{U}^\phi)^T \left[\frac{1}{N} \Phi \mathbf{X}^T \right] \mathbf{U}^x (\text{diag}(\mathbf{S}^x))^{-1}. \quad (67)$$

As $N \rightarrow \infty$, the empirical ϕ - x correlation function (the bracketed term) converges to a limiting $N_\theta \times N_\theta$ circulant matrix whose (θ, θ') element is

$$\langle \phi(x^*(\theta)) x^*(\theta') \rangle_{x^* \sim \mathcal{N}(0, \Gamma^x)} = \langle \phi'(x^*(\theta)) \rangle_{x^* \sim \mathcal{N}(0, \Gamma^x)} \langle x^*(\theta) x^*(\theta') \rangle_{x^* \sim \mathcal{N}(0, \Gamma^x)} = \alpha \Gamma^x(\theta - \theta'). \quad (68)$$

Here we applied Stein's lemma, as in Sec. 3.4, and defined

$$\alpha = \langle \phi'(x^*(0)) \rangle_{x^* \sim \mathcal{N}(0, \Gamma^x)} \quad (69)$$

to be the average slope of the activation function with respect to the Gaussian distribution of input currents. The matrices $\mathbf{U}^x$ and $\mathbf{U}^\phi$ flanking the bracket in Eq. 67 converge to $\mathbf{F}$, and $\mathbf{F}$ diagonalizes any circulant matrix, so

$$\mathbf{F}^T \Gamma^x \mathbf{F} = N_\theta \text{diag}(\hat{\Gamma}^x). \quad (70)$$

Combining these results gives

$$(\mathbf{V}^\phi)^T \mathbf{V}^x = \alpha \text{diag} \left(\sqrt{\frac{\hat{\Gamma}^x}{\hat{\Gamma}^\phi}} \right). \quad (71)$$

At large N , therefore, non-corresponding columns of $\mathbf{V}^x$ and $\mathbf{V}^\phi$ are exactly orthogonal, while corresponding columns overlap by $\alpha \sqrt{\hat{\Gamma}_n^x / \hat{\Gamma}_n^\phi}$, which is in general neither zero nor one. The two embeddings of the ring, one in input-current space and one in firing-rate space, are thus aligned mode by mode but not identical, and the activation function is what separates them. Using the singular value decomposition of $\mathbf{J}$ above and Eq. 71, one can verify that columns of $\mathbf{V}^x$ are right eigenvectors of $\mathbf{J}$ and columns of $\mathbf{V}^\phi$ are left eigenvectors of $\mathbf{J}$, with eigenvalues (same as Eq. 9 in the main text):

$$\lambda_n^J = \alpha \hat{\mathcal{J}}_n, \quad \hat{\mathcal{J}}_n = \frac{\hat{\Gamma}_n^x}{\hat{\Gamma}_n^\phi}.$$

5.3 Comparison to cosine-based connectivity

The weights of the simplest classical model are $J_{ij} = \frac{1}{N} [J_0 + J_1 \cos(\theta_i - \theta_j)]$ [8]. Written in the analogous form, our model's weights at large N (main-text Eq. 8) are

$$J_{ij} = \frac{1}{N} \sum_{n=0}^D \sqrt{\hat{\mathcal{J}}_n} r_{i,n}^x r_{j,n}^\phi \cos \left(n \left(\theta_{i,n}^x - \theta_{j,n}^\phi \right) \right),$$

where $r_{i,n}^x$ and $\theta_{i,n}^x$ are the amplitude and phase (divided by n) of Fourier mode n in row i of $\sqrt{N} \mathbf{V}^x$, and $r_{j,n}^\phi$ and $\theta_{j,n}^\phi$ are the same quantities in row j of $\sqrt{N} \mathbf{V}^\phi$. The classical matrix has a single nonconstant Fourier mode, whereas ours has D . On its own, this difference would modify only the shape of the tuning curves, for example by sharpening them, since with a single angle per neuron, $\theta_{i,n}^x = \theta_{i,n}^\phi = \theta_i$, the tuning curves remain shifted copies of one another. Heterogeneity instead arises when the angles differ across modes (and, in our model, between x and ϕ), so that each neuron carries $2D$ angles rather than one. Furthermore, whereas classical models assign each mode a single amplitude, in ours the amplitude $r_{i,n}^x r_{j,n}^\phi$ varies across neurons as well as modes.

In summary, compared to classical cosine connectivity, our model has

1. higher-order Fourier modes,
2. random, Fourier mode-specific phases, and
3. random, neuron and Fourier mode-specific amplitudes.

Together, these features amount to a more complex embedding of the system into the neuronal space, equivalent to $\mathbf{V}^x$ and $\mathbf{V}^\phi$ being disordered rather than Fourier (main text, ‘Weights in the large-network limit’).

5.4 Attempting to extract information about the latent ring structure from the right singular vectors despite their disordered nature

We took the top two columns of $\mathbf{V}^x$ and assigned each neuron the angle of its resulting two-dimensional coordinate [9]. Despite the independent and identically distributed nature of these coordinates (which form an isotropic Gaussian cloud, as expected), the spectral-embedding angle reveals the noisy circulant structure, shown for the generative-process network in Supplementary Fig. 4b.

In our model, the SVD-derived angular coordinate θ_i^{SVD} defined above is correlated with the tuning-curve center of mass θ_i^{COM} , with Fisher–Lee circular correlation $|r|$ between 0.77 and 0.85 across 10 realizations of the generative-process network (Setting 3; Methods, ‘Model settings’). To measure this, we used $N = 5,000$, larger than the $N = 2,500$ used elsewhere because at that size the spectrum is sometimes too noisy to resolve the leading doublet, and missing this doublet produces very small $|r|$. The leading singular-vector doublet therefore carries information about the tuning of neurons with respect to the encoded variable.

This fact does not, however, provide a method for detecting continuous-attractor organization from connectivity data alone. To test the specificity of the spectral-embedding angle, we constructed null models that lack doublet degeneracy in their singular-value spectra and have variable approximate rank (Supplementary Fig. 4). Each null-model spectrum is a Gaussian envelope in the singular-value index,

$$S_n = S_1 \exp\left(-\frac{(n-1)^2}{2r_{\text{eff}}^2}\right), \quad n = 1, \dots, N, \quad (72)$$

anchored at the leading singular value S_1 of the optimal weight matrix. The parameter r_{eff} sets the approximate rank. Each null model is

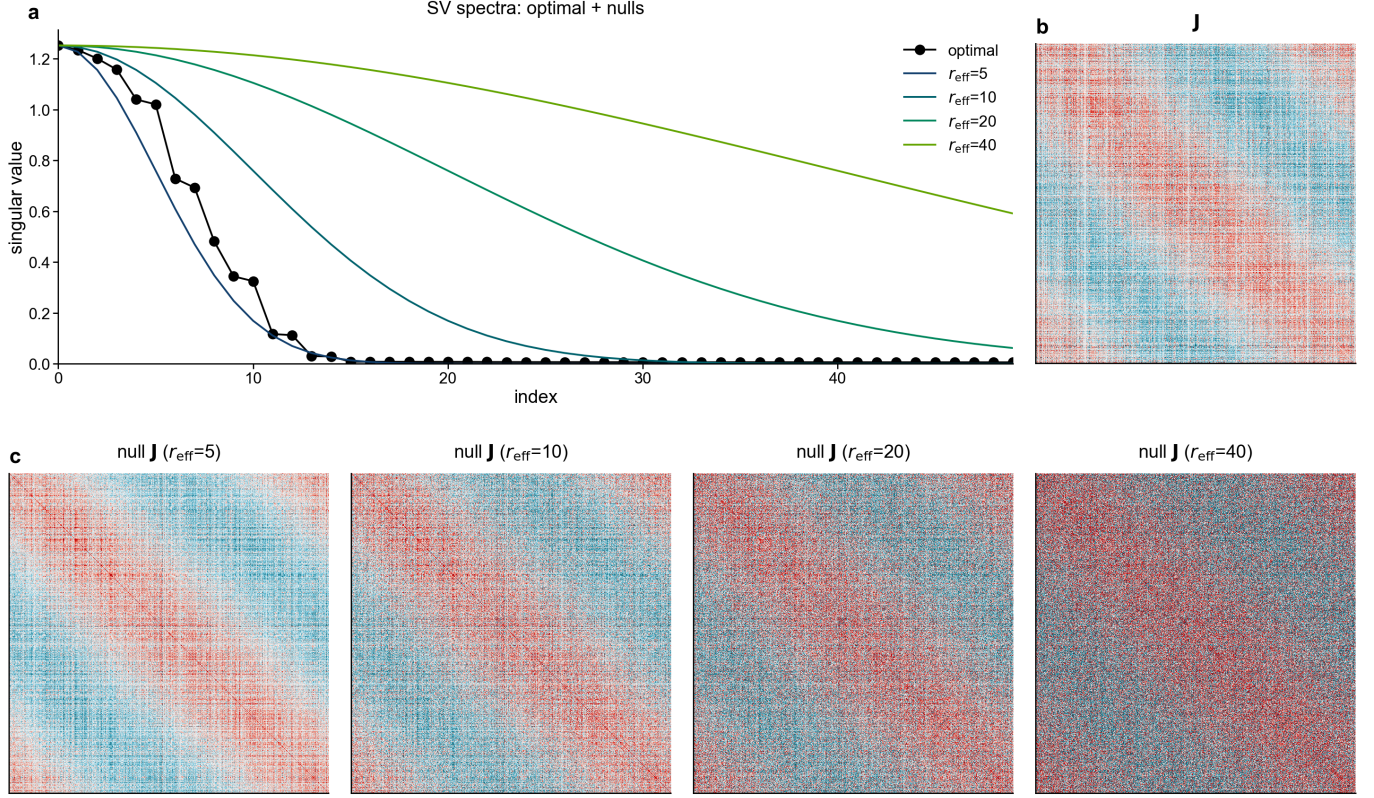
$$\mathbf{J}_{\text{null}} = \mathbf{O} \text{diag}(\mathbf{S}_{\text{null}}) \mathbf{O}^T \quad (73)$$

with $\mathbf{O}$ a Haar-random orthogonal $N \times N$ matrix. These matrices serve as null models for the specificity of the data-analysis method rather than as functional null models for continuous-attractor networks (since they do not produce continuous-attractor dynamics). Sorting the elements of these matrices by the same spectral-embedding angle reveals noisy circulant structure indistinguishable from that of the true weight matrix. Thus, being approximately low-rank is sufficient to produce this spurious circulant structure; doublet degeneracy is not necessary. The effect is strongest for the most nearly low-rank null models and fades as r_{eff} grows.

6 Details of dynamical mean-field theory

Here we derive the dynamical mean-field theory (DMFT) of the main text. It is one of three treatments in this Supplementary Information that arrive at the same dynamics.

Sec. 3 gives the mean-field dynamics of a general low-rank network in terms of its K overlaps $\tilde{\mathbf{K}}(t)$ (Eq. 31). As noted below that equation, casting our model in the general low-rank form makes the overlaps correspond to the Fourier coefficients of the filtered overlap order parameter $\tilde{m}(\theta, t)$. The main-text DMFT is therefore what Eq. 31 becomes in a θ basis rather than a Fourier basis.



Supplementary Figure 4: Optimal weight matrix and four null models of low effective rank, each sorted by the spectral-embedding angle ($N = 2,500$, generative-process Fourier decay scale $\sigma = 1.42$, sharpness $\beta = 2.77$, error-function activation). **a**, Singular-value spectra of the optimal weight matrix and of the four null models, with effective ranks $r_{\text{eff}} \in \{5, 10, 20, 40\}$. **b**, Optimal weight matrix $\mathbf{J}$, with neurons sorted by the spectral-embedding angle. **c**, The four null models, sorted by the same procedure.

Supplementary Note 4 derives a strictly more general theory, extending Rivkind and Barak [10] and Darshan and Rivkind [11] to networks that combine learned low-rank structure with random reservoir weights of strength g . Our model is its $g \rightarrow 0$ limit, with the learned patterns forming a continuum indexed by θ and their embedding random.

The present section takes the most direct route, working in θ space from the start and carrying no extra generality. This route is preferable for two reasons. First, compared to the Fourier-coefficient space, the Gaussian average is more naturally carried out in θ space, using Stein’s lemma as shown below. Second, working in θ space permits comparison with classical models, since $m(\theta, t)$ is the activity bump and $\mathcal{J}(\Delta\theta)$ is the Mexican hat interaction profile.

6.1 Derivation of dynamical mean-field theory equations

We begin by deriving the dynamical mean-field equations for zero velocity input ($\omega(t) = 0$), which capture the essential features of the derivation while simplifying the presentation. This allows us to describe the transient evolution of the overlap order parameter from arbitrary initial conditions. We then extend the analysis to incorporate velocity inputs.

Starting with the network dynamics (main-text Eq. 1) with $b = 0$ and the solution for the weights $\mathbf{J}$

(main-text Eq. 3), we have

$$(1 + \tau \partial_t) x_i(t) = \sum_{j=1}^N \underbrace{\left[\frac{1}{N} \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' K_{i,\lambda}^{-1}(\theta, \theta') x_i^*(\theta) \phi_j^*(\theta') \right]}_{J_{ij}} \phi_j(t). \quad (74)$$

In the limit $N \rightarrow \infty$ with $\lambda = 0$, the inverse optimization kernel $K_{i,\lambda}^{-1}(\theta, \theta')$ converges to $[\Gamma^\phi]^{-1}(\theta - \theta')$, where $\Gamma^\phi(\Delta\theta)$ is invertible due to all Fourier components being nonzero (we need not truncate them at a finite cutoff in this derivation). Moving the sum over j inside the integrals and using the definition of $m(\theta, t)$ from main-text Eq. 5, we obtain

$$(1 + \tau \partial_t) x_i(t) = \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' [\Gamma^\phi]^{-1}(\theta - \theta') x_i^*(\theta) m(\theta', t). \quad (75)$$

We introduce a temporally smoothed version of the overlap order parameter,

$$\tilde{m}(\theta, t) = \int_0^t \frac{dt'}{\tau} e^{-(t-t')/\tau} m(\theta, t') + e^{-t/\tau} m(\theta, 0). \quad (76)$$

This satisfies

$$(1 + \tau \partial_t) \tilde{m}(\theta, t) = m(\theta, t), \quad (77)$$

which is the temporal low-pass filtering described in the main text ('Dynamical mean-field theory'). Applying $(1 + \tau \partial_t)^{-1}$ to both sides of the dynamics equation (Eq. 75) yields

$$x_i(t) = \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' [\Gamma^\phi]^{-1}(\theta - \theta') x_i^*(\theta) \tilde{m}(\theta', t). \quad (78)$$

Applying $\phi(x)$ to both sides of this equation, then taking the inner product with $\phi_i^*(\theta)$ and dividing by N , gives

$$m(\theta, t) = \frac{1}{N} \sum_{i=1}^N \phi(x_i^*(\theta)) \phi \left(\int_0^{2\pi} du \int_0^{2\pi} du' [\Gamma^\phi]^{-1}(u - u') x_i^*(u) \tilde{m}(u', t) \right). \quad (79)$$

The key step in deriving the mean-field equations is to replace the empirical average over neurons (i index) with an expectation over the generating Gaussian process. That is,

$$m(\theta, t) = \left\langle \underbrace{\phi(x_i^*(\theta))}_{\text{term 1}} \phi \left(\underbrace{\int_0^{2\pi} du \int_0^{2\pi} du' [\Gamma^\phi]^{-1}(u - u') x_i^*(u) \tilde{m}(u', t)}_{\text{term 2}} \right) \right\rangle_{x^* \sim \mathcal{GP}(0, \Gamma^x)}. \quad (80)$$

This replacement is exact for $N \rightarrow \infty$. Terms 1 and 2 are jointly Gaussian with zero mean. Their variances and covariance fully determine the right-hand side of the equation and are given by

$$\langle [\text{term 1}]^2 \rangle = \Gamma^x(0), \quad (81)$$

$$\langle [\text{term 2}]^2 \rangle = \int_0^{2\pi} du \int_0^{2\pi} du' \int_0^{2\pi} dv \int_0^{2\pi} dv' [\Gamma^\phi]^{-1}(u - u') [\Gamma^\phi]^{-1}(v - v') \Gamma^x(u - v) \tilde{m}(u', t) \tilde{m}(v', t), \quad (82)$$

$$\langle [\text{term 1}] [\text{term 2}] \rangle = \int_0^{2\pi} du \int_0^{2\pi} du' [\Gamma^\phi]^{-1}(u - u') \Gamma^x(\theta - u) \tilde{m}(u', t). \quad (83)$$

We define the gain kernel $\mathcal{U}(\Delta\theta)$ and the effective interactions $\mathcal{J}(\Delta\theta)$ in Fourier space (the latter being defined in main-text Eq. 9),

$$\begin{aligned} \hat{\mathcal{U}}_n &= \frac{\hat{\Gamma}_n^x}{(\hat{\Gamma}_n^\phi)^2}, \\ \hat{\mathcal{J}}_n &= \frac{\hat{\Gamma}_n^x}{\hat{\Gamma}_n^\phi}. \end{aligned} \quad (84)$$

After defining these kernels (and being careful about factors of $1/2\pi$), the correlations between terms 1 and 2 can be written

$$\langle [\text{term } 2]^2 \rangle = \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' \mathcal{U}(\theta - \theta') \tilde{m}(\theta, t) \tilde{m}(\theta', t) \equiv Q(t), \quad (85)$$

$$\langle [\text{term } 1][\text{term } 2] \rangle = \frac{1}{2\pi} \int_0^{2\pi} d\theta' \mathcal{J}(\theta - \theta') \tilde{m}(\theta', t). \quad (86)$$

Now, let $\mathcal{G}(C_{11}, C_{22}, C_{12})$ denote the expectation

$$\mathcal{G}(C_{11}, C_{22}, C_{12}) = \langle \phi(x_1) \phi(x_2) \rangle \left(\begin{matrix} x_1 \\ x_2 \end{matrix} \right) \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} \right). \quad (87)$$

This can be explicitly evaluated using the activation function $\phi(x) = 1 + \text{erf}(\beta\sqrt{\pi}x/2)$, which yields

$$\mathcal{G}(C_{11}, C_{22}, C_{12}) = 1 + \frac{2}{\pi} \tan^{-1} \left(\frac{\beta^2 C_{12}}{\sqrt{(\beta^2 C_{11} + \frac{2}{\pi})(\beta^2 C_{22} + \frac{2}{\pi}) - (\beta^2 C_{12})^2}} \right). \quad (88)$$

Then, Eq. 80 can be written as

$$m(\theta, t) = \mathcal{G} \left(\Gamma^x(0), Q(t), \frac{1}{2\pi} \int_0^{2\pi} d\theta' \mathcal{J}(\theta - \theta') \tilde{m}(\theta', t) \right). \quad (89)$$

Defining

$$G_Q(h) = \mathcal{G}(\Gamma^x(0), Q, h) \quad (90)$$

and using the relationship between $m(\theta, t)$ and $\tilde{m}(\theta, t)$, we obtain the mean-field dynamics at zero angular velocity (same as Eq. 10 in the main text, with $\omega(t) = 0$),

$$(1 + \tau \partial_t) \tilde{m}(\theta, t) = G_{Q(t)} \left(\frac{1}{2\pi} \int_0^{2\pi} d\theta' \mathcal{J}(\theta - \theta') \tilde{m}(\theta', t) \right).$$

6.2 Velocity-dependent dynamics

We now extend this derivation to include nonzero angular velocity $\omega(t)$. The network dynamics are modified to include an angular velocity-dependent term,

$$(1 + \tau \partial_t) x_i(t) = \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' [\Gamma^\phi]^{-1}(\theta - \theta') (1 + \tau \omega(t) \partial_\theta) x_i^*(\theta) m(\theta', t). \quad (91)$$

To solve Eq. 91, we again apply $(1 + \tau \partial_t)^{-1}$ to both sides, and now introduce a variable $\tilde{m}_\omega(\theta, t)$ defined as

$$\tilde{m}_\omega(\theta, t) = \int_0^t \frac{dt'}{\tau} e^{-(t-t')/\tau} \tau \omega(t') m(\theta, t') + e^{-t/\tau} \tau \omega(0) m(\theta, 0), \quad (92)$$

which satisfies

$$(1 + \tau \partial_t) \tilde{m}_\omega(\theta, t) = \tau \omega(t) m(\theta, t). \quad (93)$$

Then, we have

$$x_i(t) = \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' [\Gamma^\phi]^{-1}(\theta - \theta') (x_i^*(\theta) \tilde{m}(\theta', t) + \partial_\theta x_i^*(\theta) \tilde{m}_\omega(\theta', t)). \quad (94)$$

We now massage Eq. 92 by expanding $\omega(t')$ in a Taylor series around t (dropping the initial transient for simplicity),

$$\begin{aligned}\tilde{m}_\omega(\theta, t) &= \left[\int_0^t \frac{dt'}{\tau} e^{-(t-t')/\tau} \tau \left[\omega(t) + (\partial_t \omega(t))(t' - t) + \dots \right] m(\theta, t') \right] \\ &= \left[\tau \omega(t) \tilde{m}(\theta, t) - \tau^2 (\partial_t \omega(t)) \int_0^t \frac{ds}{\tau} \frac{s}{\tau} e^{-s/\tau} m(\theta, t - s) + \dots \right].\end{aligned}\quad (95)$$

The expansion in square brackets is in powers $(\tau \partial_t)^n \tau \omega(t)$ for $n = 0, 1, 2, \dots$, which we will assume can be truncated after $n = 0$. This assumption means that the angular acceleration $\partial_t \omega(t)$, times appropriate factors of τ , is small, while the angular velocity itself may be large. This gives

$$\tilde{m}_\omega(\theta, t) \approx \tau \omega(t) \tilde{m}(\theta, t). \quad (96)$$

Substituting Eq. 96 back into Eq. 94, we obtain

$$x_i(t) = \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' [\Gamma^\phi]^{-1}(\theta - \theta') (1 + \tau \omega(t) \partial_\theta) x_i^*(\theta) \tilde{m}(\theta', t). \quad (97)$$

Following the same averaging procedure as in the zero-velocity case, but now accounting for the velocity-dependent terms in Eq. 97, the correlations are modified in two ways:

1. The gain kernel $\mathcal{U}(\Delta\theta)$ that determines the activity-dependent gain $Q(t)$ (the variance of input currents) is modified to include an angular velocity-dependent term,

$$\mathcal{U}_{\omega(t)}(\Delta\theta) = (1 - \tau^2 \omega(t)^2 \partial_{\Delta\theta}^2) \mathcal{U}(\Delta\theta), \quad (98)$$

where $\mathcal{U}(\Delta\theta)$ is the zero-velocity gain kernel defined through its Fourier transform $\hat{\mathcal{U}}_n = \hat{\Gamma}_n^x / (\hat{\Gamma}_n^\phi)^2$.

2. The correlation between input currents and target tuning curves acquires an additional angular velocity-dependent term,

$$\frac{1}{2\pi} \int_0^{2\pi} d\theta' (1 - \tau \omega(t) \partial_\theta) \mathcal{J}(\theta - \theta') \tilde{m}(\theta', t). \quad (99)$$

These modifications lead to the full dynamical mean-field equation for the overlap order parameter (same as Eq. 10 in the main text),

$$\tau \partial_t \tilde{m}(\theta, t) = -\tilde{m}(\theta, t) + G_{Q(t)} \left(\frac{1}{2\pi} \int_0^{2\pi} d\theta' [1 - \tau \omega(t) \partial_\theta] \mathcal{J}(\theta - \theta') \tilde{m}(\theta', t) \right).$$

Here $Q(t)$ is given by Eq. 11 of the main text, with $\mathcal{U}_{\omega(t)}(\Delta\theta)$ (Eq. 98) in place of $\mathcal{U}(\Delta\theta)$.

6.3 Jacobian analysis of the mean-field dynamics

To analyze the stability of solutions to the mean-field equations, we need to compute the Jacobian $M(\theta, \theta')$ defined in the main text ('Dynamical mean-field theory'), which describes how perturbations to the overlap order parameter evolve in time. This requires expressions for derivatives of the effective activation function $G_Q(h)$ with respect to both h and Q . Recall that $G_Q(h)$ is defined through the function $\mathcal{G}(C_{11}, C_{22}, C_{12})$, which computes the expectation in Eq. 87.

We define auxiliary quantities

$$Y_1 = \left(\beta^2 C_{11} + \frac{2}{\pi} \right) \left(\beta^2 C_{22} + \frac{2}{\pi} \right) \text{ and } Y_2 = Y_1 - (\beta^2 C_{12})^2. \quad (100)$$

Then, the required derivatives are

$$\partial_{C_{12}} \mathcal{G}(C_{11}, C_{22}, C_{12}) = \langle \phi'(x_1) \phi'(x_2) \rangle_{(x_1, x_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{C})} = \frac{2\beta^2}{\pi} \frac{1}{\sqrt{Y_2}}, \quad (101)$$

$$\partial_{C_{22}} \mathcal{G}(C_{11}, C_{22}, C_{12}) = \frac{1}{2} \langle \phi(x_1) \phi''(x_2) \rangle_{(x_1, x_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{C})} = -\frac{\beta^4}{\pi} \frac{C_{12}}{Y_1 \sqrt{Y_2}} \left(\beta^2 C_{11} + \frac{2}{\pi} \right), \quad (102)$$

where $\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix}$.

For the bump solution $\tilde{m}(\theta) = \Gamma^\phi(\theta)$, the Jacobian, defined by $\tau \partial_t \delta \tilde{m}(\theta) = \int_0^{2\pi} d\theta' M(\theta, \theta') \delta \tilde{m}(\theta')$, takes the form

$$M(\theta, \theta') = -\delta(\theta - \theta') + \frac{1}{2\pi} \partial_{C_{12}} \mathcal{G}(\Gamma^x(0), \Gamma^x(0), \Gamma^x(\theta)) \mathcal{J}(\theta - \theta') + \frac{1}{2\pi} 2\partial_{C_{22}} \mathcal{G}(\Gamma^x(0), \Gamma^x(0), \Gamma^x(\theta)) \mathcal{J}(\theta'). \quad (103)$$

The three terms are exponential decay, direct coupling through the effective interactions, and modulation through the activity-dependent gain $Q(t)$.

For the homogeneous solution $\tilde{m}(\theta) = m(\theta) = \tilde{m}_0$, we first determine $\tilde{m}_0$ by numerically solving the self-consistent equations

$$\tilde{m}_0 = \mathcal{G}(\Gamma^x(0), Q_0, h_0), \quad (104)$$

$$Q_0 = \tilde{m}_0^2 \hat{\mathcal{U}}_0, \quad (105)$$

$$h_0 = \tilde{m}_0 \hat{\mathcal{J}}_0. \quad (106)$$

The Jacobian at this solution is then

$$M(\theta, \theta') = -\delta(\theta - \theta') + \frac{1}{2\pi} \partial_{C_{12}} \mathcal{G}(\Gamma^x(0), Q_0, h_0) \mathcal{J}(\theta - \theta') + \frac{1}{2\pi} 2\partial_{C_{22}} \mathcal{G}(\Gamma^x(0), Q_0, h_0) \hat{\mathcal{U}}_0 \tilde{m}_0. \quad (107)$$

To analyze the equations with Q fixed at $\Gamma^x(0)$, as in the main text ('Dynamical mean-field theory'), we simply omit the terms involving $\partial_{C_{22}} \mathcal{G}(C_{11}, C_{22}, C_{12})$ in both Jacobians (and, if we would like to find the new stable uniform state, we fix $Q_0 = \Gamma^x(0)$ in the self-consistent equations above rather than allowing it to depend on $\tilde{m}_0$). The presence of a positive eigenvalue of the Jacobian for the bump state when we fix $Q = \Gamma^x(0)$, shown in Fig. 7f, demonstrates that the activity-dependent gain modulation through $Q(t)$ is essential for bump stability.

6.4 Finite domain of effective activation function

The effective activation function $G_Q(h)$ (Eq. 90) is defined only for

$$|h| \leq \sqrt{Q \Gamma^x(0)}. \quad (108)$$

Beyond this range, the 2×2 covariance matrix of the underlying bivariate Gaussian is non-positive-definite.

This bound holds in the dynamical mean-field system, as we now show. Using matrix-vector notation for clarity, we have

$$h(\theta) = \mathbf{e}_\theta^T \Gamma^x \mathbf{y} \quad (109)$$

$$Q = \mathbf{y}^T \Gamma^x \mathbf{y} \quad (110)$$

where $\mathbf{y} = (\Gamma^\phi)^{-1} \tilde{\mathbf{m}}(t)$ and $\mathbf{e}_\theta$ selects angular component θ . Since Γ^x is positive definite, the Cauchy-Schwarz inequality implies that

$$(h(\theta))^2 = (\mathbf{e}_\theta^T \Gamma^x \mathbf{y})^2 \leq (\mathbf{e}_\theta^T \Gamma^x \mathbf{e}_\theta) (\mathbf{y}^T \Gamma^x \mathbf{y}) = \Gamma^x(0) Q, \quad (111)$$

which is the bound above.

6.5 Reciprocal symmetry of J at large N

At large N ,

$$\underbrace{\text{corr}(J_{ij}, J_{ji})}_{i \neq j} = \frac{\langle J_{ij} J_{ji} \rangle}{\langle J_{ij}^2 \rangle} = \frac{\alpha^2 [\mathcal{J}^2](0)}{\mathcal{J}(0)}, \quad (112)$$

where

$$[\mathcal{J}^2](\Delta\theta) = \frac{1}{2\pi} \int_0^{2\pi} du \mathcal{J}(\Delta\theta - u) \mathcal{J}(u). \quad (113)$$

Here $\mathcal{J}(\Delta\theta)$ are the circulant effective interactions in the dynamical mean-field theory, and α is given by Eq. 69.

6.6 Emergence of Mexican hat interactions from tuning-curve statistics

The Mexican hat profile emerges from the activation function's effect on correlation structure, specifically from a quartic term it generates in the Fourier transform of the effective interactions.

To illustrate this mechanism, consider correlation functions on the infinite line, with position s and frequency k in place of θ and n , since circularity is not essential. Define the input-current correlation function as $\Gamma^x(\Delta s) = \exp(-\sigma^2 \Delta s^2 / 2)$, the infinite-line analog of the main-text wrapped Gaussian with Fourier decay scale σ . The firing-rate correlation function is $\Gamma^\phi(\Delta s) = \mathcal{G}(1, 1, \Gamma^x(\Delta s))$, where $\mathcal{G}(C_{11}, C_{22}, C_{12})$ is defined in Eq. 87 and $\phi(x) = 1 + \text{erf}(\beta\sqrt{\pi}x/2)$ is the activation function of Setting 3 (Methods, ‘Model settings’).

Normalizing $\Gamma^x(\Delta s)$ and the connected firing-rate correlation $\Gamma^\phi(\Delta s) - \mathcal{G}(1, 1, 0)$ to probability distributions, $P^x(\Delta s) = \Gamma^x(\Delta s) / \int_{-\infty}^{\infty} d\Delta s \Gamma^x(\Delta s)$ and $P^\phi(\Delta s)$ defined the same way, gives cumulants κ_2^x , κ_4^x , κ_2^ϕ , and κ_4^ϕ . The subtracted constant $\mathcal{G}(1, 1, 0) = 1$ is the squared mean firing rate, and affects only the $k = 0$ mode, not the Mexican hat shape.

Because $\Gamma^x(\Delta s)$ is Gaussian in Δs , $\kappa_2^x = 1/\sigma^2$ and $\kappa_4^x = 0$. What happens to κ_2^ϕ and κ_4^ϕ is nontrivial.

The effective interactions in Fourier space are

$$\hat{\mathcal{J}}(k) \propto \frac{\hat{\Gamma}^x(k)}{\hat{\Gamma}^\phi(k)} \quad (114)$$

$$= \frac{\exp(-\frac{1}{2}(k/\sigma)^2)}{\int d\Delta s e^{ik\Delta s} P^\phi(\Delta s)} \quad (115)$$

$$= \exp\left(-\frac{1}{2}(k/\sigma)^2 - \log \int d\Delta s e^{ik\Delta s} P^\phi(\Delta s)\right) \quad (116)$$

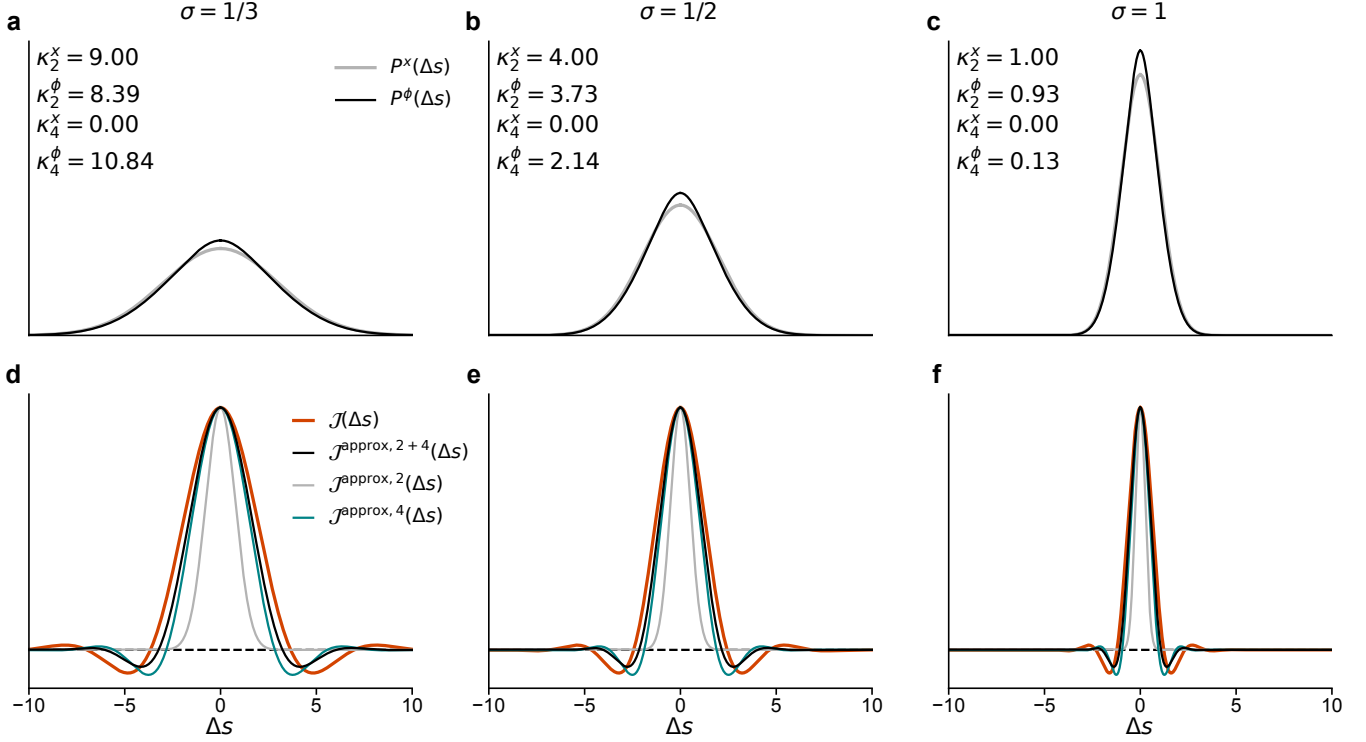
$$= \exp\left(-\frac{1}{2}(1/\sigma^2 - \kappa_2^\phi)k^2 - \frac{1}{24}\kappa_4^\phi k^4 + \mathcal{O}(k^6)\right), \quad (117)$$

where in the last line we recognized the cumulant generating function of $P^\phi(\Delta s)$.

The activation function “tightens” tuning, a well-known effect leading to $\kappa_2^\phi < \kappa_2^x$. Less considered is the fourth cumulant, which this activation function makes positive, $\kappa_4^\phi > 0$. This quartic term in Fourier space is essential for the Mexican hat profile. If it were zero, $\hat{\mathcal{J}}(k)$ would be purely Gaussian, producing a Gaussian (not Mexican hat) profile in position space. Indeed, as demonstrated in Supplementary Fig. 5, the observed quadratic and quartic terms are sufficient for generating a Mexican hat shape.

6.7 Continuous attractors without continuous symmetry

The correspondence with classical models developed in the main text rests on rotational symmetry of the tuning-curve statistics, but the general form of the mean-field framework does not require it. To illustrate this,



Supplementary Figure 5: Origin of the Mexican hat interaction profile (sharpness $\beta = 2$, error-function activation). Panels are paired by Fourier decay scale σ , namely **a** with **d**, **b** with **e**, and **c** with **f**. **a** to **c**, The input-current correlation function and the connected firing-rate correlation function, each normalized to a probability distribution, $P^x(\Delta s)$ and $P^\phi(\Delta s)$, annotated with their second and fourth cumulants κ_2 and κ_4 . **d** to **f**, Effective interactions $\mathcal{J}(\Delta s)$, from the ratio of Fourier coefficients, and three approximations, $\mathcal{J}^{\text{approx}, 2+4}(\Delta s)$ keeping the quadratic and quartic terms, $\mathcal{J}^{\text{approx}, 2}(\Delta s)$ the quadratic only, and $\mathcal{J}^{\text{approx}, 4}(\Delta s)$ the quartic only.

we consider the nonsymmetric correlation functions

$$\Gamma^x(\theta, \theta') = (1 + f \sin(2\theta))(1 + f \sin(2\theta')) \cos(\theta - \theta'), \quad (118)$$

$$\Gamma^\phi(\theta, \theta') = \mathcal{G}(\Gamma^x(\theta, \theta), \Gamma^x(\theta', \theta'), \Gamma^x(\theta, \theta')). \quad (119)$$

For $f = 0$ the correlation functions are circulant, whereas $f > 0$ breaks rotational symmetry. In $N_\theta \times N_\theta$ matrix notation, the interaction and gain kernels are

$$\mathcal{J} = \mathbf{\Gamma}^x(\mathbf{\Gamma}^\phi + \lambda \mathbf{I})^{-1}, \quad (120)$$

$$\mathcal{U} = (\mathbf{\Gamma}^\phi + \lambda \mathbf{I})^{-1} \mathbf{\Gamma}^x(\mathbf{\Gamma}^\phi + \lambda \mathbf{I})^{-1}, \quad (121)$$

which follow from the same derivation as in Sec. 6.1, without assuming translation invariance or going to Fourier space. The mean-field dynamics are

$$\tau \partial_t \tilde{m}(\theta, t) = -\tilde{m}(\theta, t) + \mathcal{G}(\Gamma^x(\theta, \theta), Q(t), h(\theta, t)), \quad (122)$$

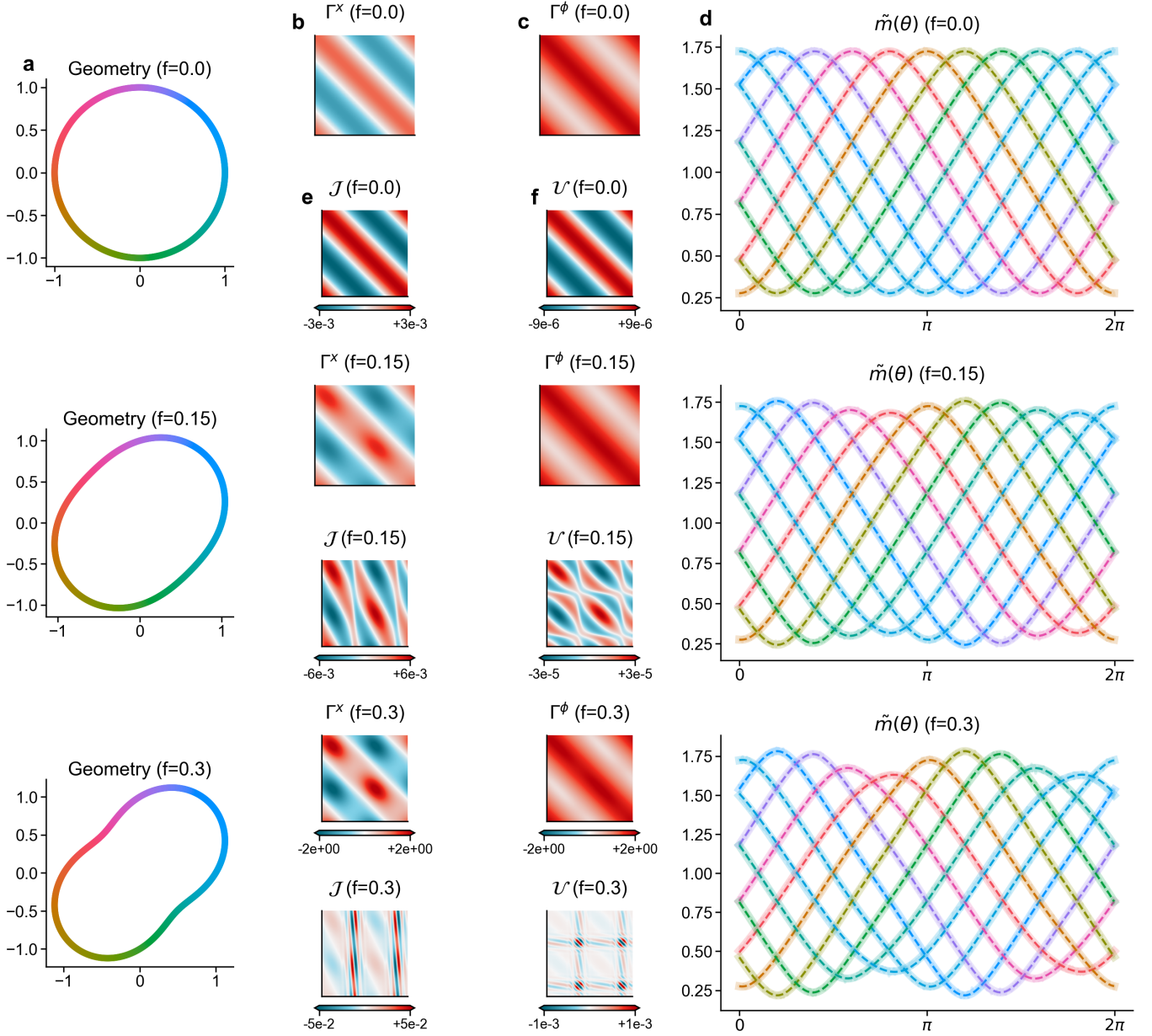
where

$$h(t) = \mathcal{J} \tilde{m}(t), \quad (123)$$

$$Q(t) = \tilde{m}(t)^T \mathcal{U} \tilde{m}(t). \quad (124)$$

Compared to the circulant ($f = 0$) case, the first argument $\Gamma^x(\theta, \theta)$ now depends on θ , and $\Gamma^x(\theta, \theta')$, $\Gamma^\phi(\theta, \theta')$, $\mathcal{J}(\theta, \theta')$, and $\mathcal{U}(\theta, \theta')$ are all noncirculant, so $\mathcal{J}$ is no longer a convolution.

For $\lambda = 0$, substituting $\tilde{m}(\theta) = \Gamma^\phi(\theta, \psi)$ gives $h(\theta) = \Gamma^x(\theta, \psi)$ and $Q = \Gamma^x(\psi, \psi)$, so the right-hand side reduces to $\mathcal{G}(\Gamma^x(\theta, \theta), \Gamma^x(\psi, \psi), \Gamma^x(\theta, \psi)) = \Gamma^\phi(\theta, \psi)$, and this profile is a fixed point for every ψ . Supplementary Fig. 6 shows the resulting attractors as the asymmetry f increases.



Supplementary Figure 6: Dynamical mean-field theory of asymmetric continuous attractors (sharpness $\beta = 2.5$, regularization $\lambda = 10^{-6}$, error-function activation). Color gives the angular coordinate θ in **a** and the initialization angle ψ in **d**. **a**, Geometric structure of the manifold, a circle at $f = 0$ and increasingly deformed for $f > 0$. **b**, Input-current correlation function $\Gamma^x(\theta, \theta')$. **c**, Firing-rate correlation function $\Gamma^\phi(\theta, \theta')$. **d**, Angular profiles of the overlap order parameter $\tilde{m}(\theta, t)$ for 10 values of ψ . Thick solid lines are unperturbed target states $\Gamma^\phi(\theta, \psi)$; thin dashed lines are the states reached after 30 s of evolution from those targets perturbed by independent Gaussian noise of standard deviation 0.01. **e**, **f**, Effective interactions $\mathcal{J}(\theta, \theta')$ (**e**) and gain kernel $\mathcal{U}(\theta, \theta')$ (**f**).

7 Fourier coefficients of correlation functions

This section computes the Fourier coefficients of the correlation functions used in the paper. These are wrapped functions, and thus their coefficients follow from the Poisson summation formula (Eq. 28).

Wrapped Gaussian. Consider first the wrapped Gaussian correlation function (same as Eq. 6 in the main text)

$$\Gamma^x(\Delta\theta) = \sum_{p=-\infty}^{\infty} \exp\left(-\frac{1}{2}\sigma^2(\Delta\theta + 2\pi p)^2\right).$$

We identify $f(x) = \exp(-\frac{1}{2}\sigma^2 x^2)$, which has Fourier transform

$$\hat{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx} e^{-\frac{1}{2}\sigma^2 x^2} dx \quad (125)$$

$$= \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{k^2}{2\sigma^2}\right). \quad (126)$$

Therefore, from Poisson summation, the Fourier coefficients of $\Gamma^x(\Delta\theta)$ are

$$\hat{\Gamma}_n^x = \hat{f}(n) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{n^2}{2\sigma^2}\right). \quad (127)$$

As discussed in Secs. 2 and 3.2, there is a doublet degeneracy of the Fourier coefficients ($\hat{\Gamma}_n^x = \hat{\Gamma}_{-n}^x$) for all n except $n = 0$, for which there is no degeneracy.

Grid-cell correlation function. To generate grid-cell input-current maps, we extend the Gaussian process to periodicity in two dimensions, with the periodic variables a two-dimensional vector of angles $\boldsymbol{\theta} = (\theta_1, \theta_2)$. The correlation function of the Gaussian process is

$$\Gamma^x(\Delta\boldsymbol{\theta}) = \sum_{\mathbf{p} \in \mathbb{Z}^2} A(\mathbf{p}) \exp\left(-\frac{\sigma^2}{2} \left\| \mathbf{H} \left(\Delta\boldsymbol{\theta} + \frac{2\pi\mathbf{p}}{\nu} \right) \right\|_2^2\right), \quad (128)$$

where $\mathbf{p} = (p_1, p_2)$ and the matrix

$$\mathbf{H} = \frac{1}{2\pi} \begin{pmatrix} 1 & 1/2 \\ 0 & \sqrt{3}/2 \end{pmatrix}, \quad |\det \mathbf{H}| = \frac{\sqrt{3}}{2(2\pi)^2}, \quad (129)$$

has columns defining the axes of a hexagonal lattice. The weighting function

$$A(\mathbf{p}) = \begin{cases} 1 & \mathbf{p} \in (\nu\mathbb{Z})^2 \\ \eta & \text{otherwise,} \end{cases} \quad (130)$$

with $\eta \in [0, 1]$, controls the periodicity of the patterns. Specializing to $\nu = 1$ and $A(\mathbf{p}) = 1$, the correlation function becomes

$$\Gamma^x(\Delta\boldsymbol{\theta}) = \sum_{\mathbf{p} \in \mathbb{Z}^2} \exp\left(-\frac{\sigma^2}{2} \left\| \mathbf{H} (\Delta\boldsymbol{\theta} + 2\pi\mathbf{p}) \right\|_2^2\right). \quad (131)$$

This function has a 12-fold symmetry $\Gamma^x(\Delta\boldsymbol{\theta}) = \Gamma^x(\mathbf{g} \Delta\boldsymbol{\theta})$, where $\mathbf{g}$ runs over the 12 matrices of a 2×2 representation of D_6 , the dihedral group of order 12. Using Poisson summation, the Fourier coefficients are

$$\hat{\Gamma}_{\mathbf{n}}^x = \frac{1}{\sqrt{\det(2\pi\sigma^2 \mathbf{H}^T \mathbf{H})}} \exp\left(-\frac{1}{2\sigma^2} \mathbf{n}^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{n}\right), \quad (132)$$

where $\mathbf{H}^T \mathbf{H} = \frac{1}{(2\pi)^2} \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$ and $(\mathbf{H}^T \mathbf{H})^{-1} = \frac{4(2\pi)^2}{3} \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}$. This can be further expressed as

$$\hat{\Gamma}_{\mathbf{n}}^x = \frac{(2\pi)^2}{\sqrt{3}\pi\sigma^2} \exp\left(-\frac{8\pi^2 L(\mathbf{n})}{3\sigma^2}\right), \quad (133)$$

where

$$L(\mathbf{n}) = n_1^2 + n_2^2 - n_1 n_2. \quad (134)$$

Thus, all integer pairs $\mathbf{n} \in \mathbb{Z}^2$ with the same value of $L(\mathbf{n})$ yield identical Fourier coefficients. The degeneracy of a given eigenvalue is therefore the number of integer solutions to $L(\mathbf{n}) = L$, which is a classical counting problem for the hexagonal lattice. Explicitly, for $L \geq 1$ one has

$$r(L) = |\{\mathbf{n} \in \mathbb{Z}^2 \mid L(n_1, n_2) = L\}| = 6(d_{1,3}(L) - d_{2,3}(L)), \quad (135)$$

where $d_{1,3}(L)$ and $d_{2,3}(L)$ denote the numbers of divisors of L congruent to 1 and 2, respectively, modulo 3. Every such degeneracy $r(L)$ is a multiple of 6. The constant mode $\mathbf{n} = \mathbf{0}$, with $L = 0$, is nondegenerate. For several of the first few representable values of L (including $L = 1$), the degeneracy is exactly 6.

In cartesian coordinates $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$, defined through the hexagonal-lattice matrix $\mathbf{H}$ above, and writing $\mathcal{J}(\mathbf{s})$ for the two-dimensional effective interactions with Fourier coefficients $\hat{\mathcal{J}}_{\mathbf{n}} = \hat{\Gamma}_{\mathbf{n}}^x / \hat{\Gamma}_{\mathbf{n}}^\phi$, the $N \rightarrow \infty$ mean-field description of the grid-cell network with velocity inputs takes a form similar to that of classical grid-cell models [12, 13],

$$\tau \partial_t \tilde{m}(\mathbf{s}, t) = -\tilde{m}(\mathbf{s}, t) + G_{Q(t)} \left(\frac{1}{(2\pi)^2 |\det \mathbf{H}|} \int_{\Omega} d\mathbf{s}' [1 - \tau \mathbf{v}^T(t) \boldsymbol{\nabla}_{\mathbf{s}}] \mathcal{J}(\mathbf{s} - \mathbf{s}') \tilde{m}(\mathbf{s}', t) \right), \quad (136)$$

where $\mathbf{v}(t) = (v_x(t), v_y(t))$ is the cartesian velocity and $\boldsymbol{\nabla}_{\mathbf{s}} = (\partial_{s_x}, \partial_{s_y})$ the cartesian gradient. The integration domain is $\Omega = \mathbf{H}[0, 2\pi)^2$. This equation exhibits two-dimensional translation symmetry. With $\nu = 1$ it is a single-bump model, in which $\tilde{m}(\mathbf{s}, t)$ is one localized region of activity moving in response to velocity input.

Supplementary Note 3: Model construction and stability

This Note describes how the finite-size networks we simulate are built, and how stable their attractors turn out to be.

8	Details of data-derived network model	31
8.1	Preprocessing for the data-derived network construction	31
8.2	Variants on the weight matrix	32
8.3	Stability and c	33
8.4	Impact of λ and c	33
8.5	Overlap order parameter provides the MAP estimate of head direction under certain assumptions	33
9	Structural and global stability	35
9.1	Structural stability of the manifold	35
9.2	Global stability of the manifold	35
10	Additional figures	38

8 Details of data-derived network model

8.1 Preprocessing for the data-derived network construction

Two additional steps precede the weight construction, beyond the cross-validated smoothing and per-neuron normalization of the Methods section ‘Data processing’. They apply only to the data-derived model; every other analysis of the tuning curves, including the statistics of the Methods section ‘Tuning-curve statistics’ and the generative-process fits of the main-text section ‘Statistical generative process’ (Setting 2; Methods, ‘Model settings’), uses the cross-validated, per-neuron normalized curves.

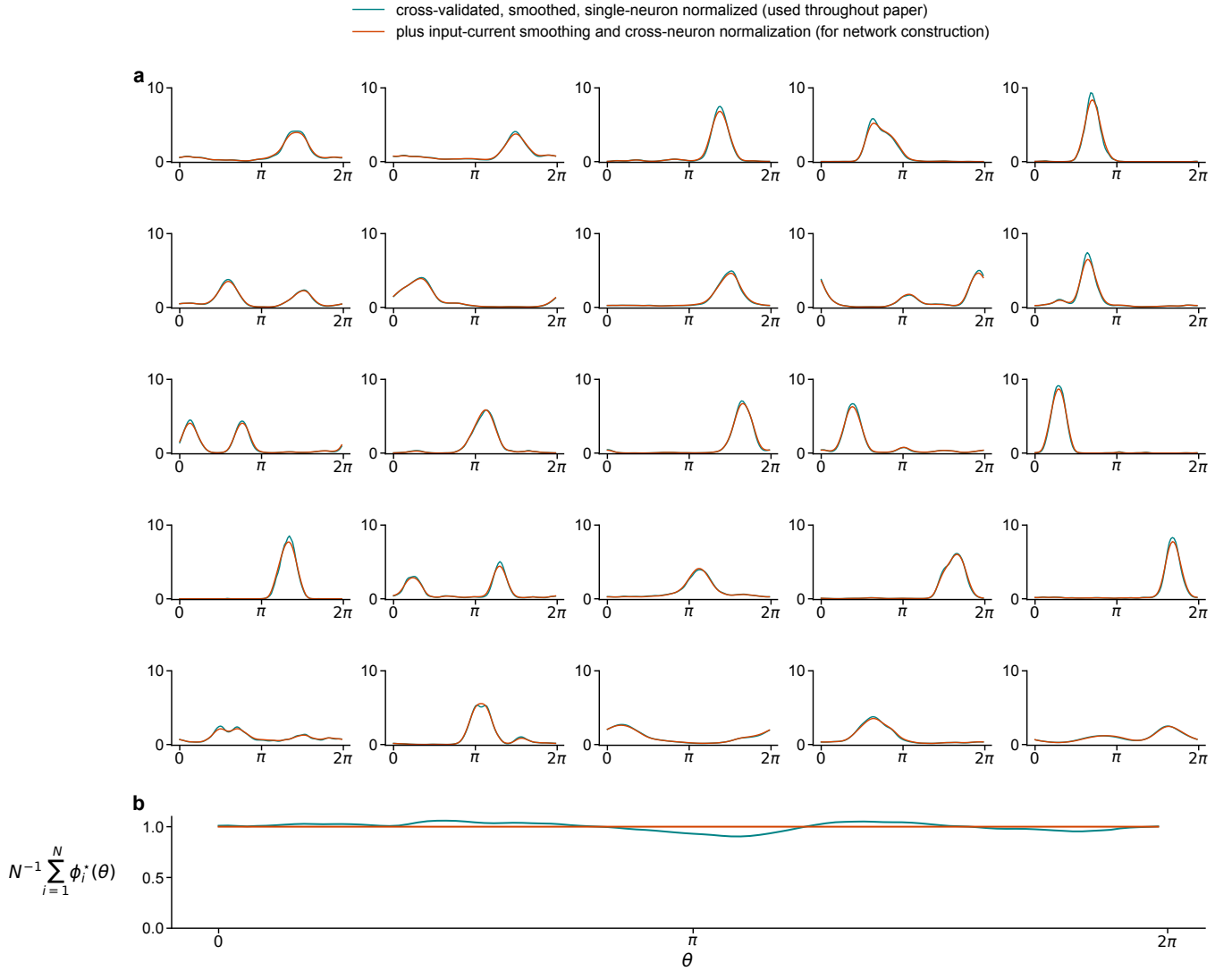
Smoothing of input currents. We obtain $x_i^*(\theta)$ by applying the inverse activation function $\phi^{-1}(\cdot)$ to the measured firing-rate tuning curves (main text, ‘Constructing a data-derived network model’). Where $\phi(x)$ is shallow, this inversion amplifies small fluctuations in the estimated rates into large, non-Gaussian excursions in the input currents. We therefore smoothed $x_i^*(\theta)$ with a circular Gaussian filter of width 2 bins (of 100). This reduces the largest excursions but does not make the estimated currents Gaussian. Consequently, quantitative large- N predictions that depend on the estimated currents of the individual recorded neurons, such as the eigenvalue spectrum of the data-derived $\mathbf{J}$, do not hold, although the generative process accurately describes the firing-rate statistics (main text, ‘Statistical generative process’). A principled alternative to this activation-function inversion would be to infer the input currents within the generative process using variational-inference methods. We leave this to future work (main text, ‘Unknown variables’).

Double normalization. We then apply the activation function to the smoothed input currents to obtain firing rates, and normalize those so that the population-averaged rate is independent of θ . While the integral over θ is exactly unity for each i by explicit normalization (Methods, ‘Data processing’), the population average over i for each θ exhibits small fluctuations around unity. We removed them by simultaneously normalizing both the neuronal (i) and angular (θ) averages to unity with the Sinkhorn–Knopp algorithm, which iteratively normalizes both axes until convergence. This results in a small correction to each tuning curve, displayed in Supplementary Fig. 7.

We make two remarks on this second step. First, in the limit $N \rightarrow \infty$, achieving the condition $N^{-1} \sum_{i=1}^N \phi_i^*(\theta) = \text{const.}$ for each θ does *not* require normalization along θ for each i ; at large N , this

condition results (up to $1/\sqrt{N}$ fluctuations) from the distributional circular symmetry of tuning curves, for which we have provided statistical evidence in the mouse data. Indeed, the corrections this procedure makes are very small, consistent with its removing a residual effect of order $1/\sqrt{N}$ (Supplementary Fig. 7).

Second, even in finite- N systems without exact enforcement of the θ -independent i -average condition (and thus without the reparameterization symmetry under the parameter c), one could presumably configure uniform inhibition and positive bias currents that stabilize the manifold. The normalization is required by our method of stabilization, which is chosen simply for convenience.



Supplementary Figure 7: Effect of double normalization on tuning curves ($N_{\text{data}} = 1,533$). Teal lines show cross-validated, smoothed tuning curves with single-neuron normalization (used throughout the paper). Orange lines show the same curves after both preprocessing steps used for the network construction, namely the 2-bin smoothing of the input currents and the cross-neuron normalization. **a**, 25 randomly selected tuning curves. **b**, Population-averaged firing rate $N^{-1} \sum_{i=1}^N \phi_i^*(\theta)$ against head direction θ .

8.2 Variants on the weight matrix

For the circulant and noisy circulant variants on the sorted data-derived weight matrix of the main-text section ‘Dynamics of the data-derived network model’ (Fig. 2c,d), we multiply each matrix by a factor of 1.275, chosen to produce bumps of similar height to those of the data-derived network.

8.3 Stability and c

With $c > 0$, the dynamics become

$$\tau \partial_t x_i(t) = -x_i(t) + \sum_{j=1}^N J_{ij} \phi_j(t) - c(\mu(t) - 1), \quad (137)$$

where $\mu(t) = N^{-1} \sum_{i=1}^N \phi_i(t)$ is the population average. Since $\mu(t) = 1$ when the network lies on the target manifold, this term subtracts input current when $\mu(t) > 1$ and adds input current when $\mu(t) < 1$, maintaining $\mu(t) = 1$. For $c \gg 1$, the mean mode exhibits exponential decay to unity with timescale $\sim \tau/c$. Equivalently, this is the uniform shift $J_{ij} \rightarrow J_{ij} - c/N$ together with $b = c$. The weight matrices displayed in Fig. 2 are shown before that shift, and the dynamics in that figure use $c = 1$.

8.4 Impact of λ and c

We analyzed how attractor stability depends on both c (mean-mode stabilization) and λ (regularization) in Supplementary Figs. 8 and 9. In the paper, we use $\lambda = 10^{-8}$ and $c = 1$; these figures show the full parameter dependence.

Consider first Supplementary Fig. 8. At $c = 0$ the activity either diverges or collapses to the origin, depending on the initial condition. Any $c > 0$ in the grid prevents the divergence, but small c leaves the collapse. Additionally, λ can be too large. In our grid, $\lambda = 10^{-2}$ destroys the manifold, causing collapse to the origin. Thus, c must be sufficiently large, and λ must be sufficiently small.

Supplementary Fig. 9 shows the mean squared flow rising with larger λ and also rising at small λ , where the optimization kernel becomes ill-conditioned. The value $\lambda = 10^{-8}$ used in the paper sits just above the minimum, beyond the ill-conditioned regime but with the flow error still negligible.

8.5 Overlap order parameter provides the MAP estimate of head direction under certain assumptions

In the main-text section ‘Dynamics of the data-derived network model’, the overlap order parameter is computed directly from recordings, using measured tuning curves and binned spike counts. Here we show that it is the maximum *a posteriori* estimate of θ under the assumptions below. Consider a random vector $\mathbf{r} = (r_1, \dots, r_N)$ of observed firing rates, distinct from the network state vector $\boldsymbol{\phi}(t)$ studied in the paper. Assume that each neuron’s firing rate is conditionally independent given head direction θ , with Gaussian noise around the tuning curve,

$$r_i | \theta \stackrel{\text{ind.}}{\sim} \mathcal{N}(\phi_i^*(\theta), \sigma_{\text{obs}}^2). \quad (138)$$

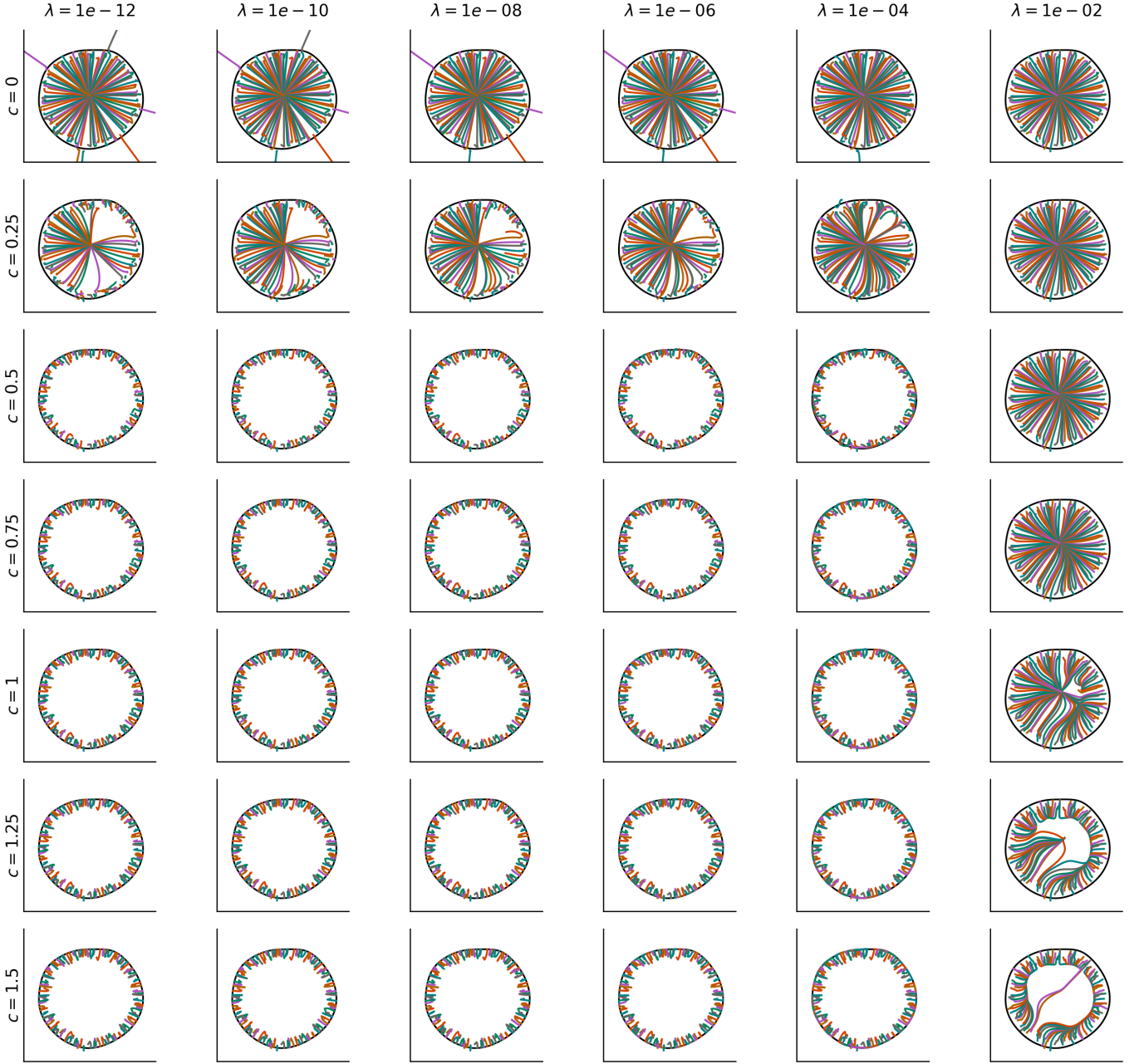
The likelihood is then

$$P(\mathbf{r}|\theta) \propto \exp \left(-\frac{1}{2\sigma_{\text{obs}}^2} \sum_{i=1}^N (r_i - \phi_i^*(\theta))^2 \right) \quad (139)$$

$$= \exp \left(-\frac{\|\mathbf{r}\|^2}{2\sigma_{\text{obs}}^2} \right) \exp \left(\frac{N}{\sigma_{\text{obs}}^2} \left[m(\theta) - \frac{1}{2N} \|\boldsymbol{\phi}^*(\theta)\|^2 \right] \right) \quad (140)$$

where $m(\theta) = N^{-1} \sum_{i=1}^N \phi_i^*(\theta) r_i$ is the overlap order parameter defined in this context. Under a uniform prior $P(\theta) = 1/2\pi$, neither the prior nor the first factor depends on θ , so both drop out, leaving

$$P(\theta|\mathbf{r}) \propto \exp \left(\frac{N}{\sigma_{\text{obs}}^2} \left[m(\theta) - \frac{1}{2N} \|\boldsymbol{\phi}^*(\theta)\|^2 \right] \right). \quad (141)$$

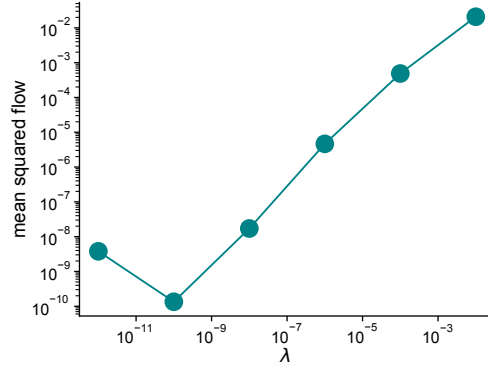


Supplementary Figure 8: Attractor stability across parameter values ($N = N_{\text{data}} = 1,533$, softplus activation). Each panel shows the plane of the top two principal components of the target manifold (black), with trajectories from 100 initial conditions obtained by input-current perturbations off of the manifold with independent Gaussian noise across neurons. Regularization values λ vary across columns; uniform inhibition values c vary across rows. Initial conditions appear inside the manifold for the reason given in Fig. 2e.

The maximum *a posteriori* estimate is therefore

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} \left[m(\theta) - \frac{1}{2N} \|\phi^*(\theta)\|^2 \right]. \quad (142)$$

When the squared firing-rate norm is independent of θ , this expression reduces to the peak of $m(\theta)$. Circularly symmetric tuning-curve statistics make the norm approximately independent of θ at large N , by the same argument applied to the population-averaged firing rate in Sec. 8.1.



Supplementary Figure 9: Flow error of the data-derived network construction as a function of the regularization parameter λ ($N = N_{\text{data}} = 1,533$, softplus activation). Flow error is the mean squared flow on the target manifold.

9 Structural and global stability

9.1 Structural stability of the manifold

Classical ring-attractor models degrade when structural (i.e., weight) noise is added. Our model already contains disorder, in the form of a disordered embedding. Here, we ask whether further weight noise degrades it as it degrades a classical circulant model.

We compared a disordered weight matrix from the generative process (Setting 3; Methods, ‘Model settings’) against a circulant counterpart with proportional eigenvalues at nonzero frequencies (Supplementary Fig. 10). The circulant matrix is the mean-field profile $\mathcal{J}(\Delta\theta)$ discretized over N neurons, with 1 subtracted from every weight to give the uniform inhibition required for unimodal bumps, then rescaled so that its diagonal element is $10/N$. The subtraction shifts only the $n = 0$ eigenvalue, so at $n \neq 0$ the circulant eigenvalues stay proportional to those of the disordered matrix.

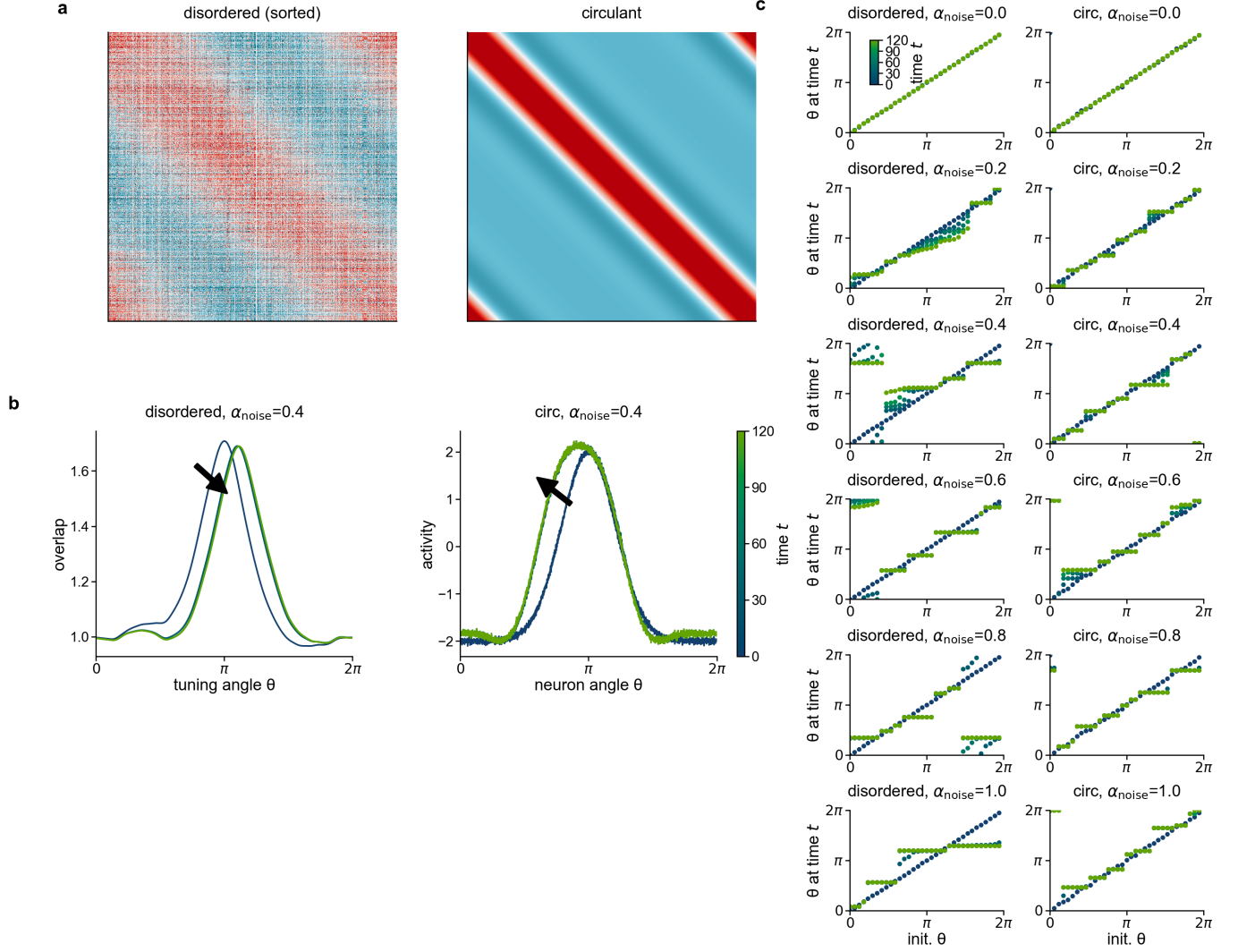
Both networks were initialized with slightly noisy patterns spaced uniformly around their attractor manifolds, then evolved with quenched independent Gaussian noise added to every weight, of standard deviation α_{noise} times that of the noiseless weights. We tracked the angular position of the activity over time, taken as the peak of the overlap order parameter for the disordered network and the peak across neurons for the circulant one.

Both networks fragment into discrete attractor states as the noise grows (Supplementary Fig. 10c), with the continuum absent by $\alpha_{\text{noise}} = 0.4$. At $\alpha_{\text{noise}} = 0.2$ the circulant network has already locked onto discrete states while the disordered one still drifts slowly, and at $\alpha_{\text{noise}} = 1$ this reverses, with the disordered network settling onto fewer states. Across the range, the two systems are comparably robust.

9.2 Global stability of the manifold

To assess the global stability of the target manifold in finite-size networks, we examined whether initial conditions far from the manifold converge to it or to spurious attractors elsewhere in state space. We used the model parameters from the main-text section ‘Weights in the large-network limit’ (Setting 3; Methods, ‘Model settings’) with $N = 2,500$.

We initialized the network at i.i.d. Gaussian states with zero mean and varying standard deviations per neuron, ranging from 1 to 1,000. A state with $\mathcal{O}(1)$ per-neuron standard deviation projects onto the leading two principal components of the target manifold with magnitude $\mathcal{O}(1)$, while the manifold radius in that plane is $\mathcal{O}(\sqrt{N})$, so initial conditions reach the scale of the ring only once the per-neuron standard deviation is



Supplementary Figure 10: Structural stability of disordered and circulant networks ($N = 2,500$, generative-process Fourier decay scale $\sigma = 1.42$, sharpness $\beta = 2.77$, error-function activation). **a**, Weight matrices of the disordered network with neurons sorted by center-of-mass angles (left) and of the corresponding circulant network (right). **b**, Angular profiles at five times up to 120 s for one example initial condition at relative weight-noise level $\alpha_{\text{noise}} = 0.4$, showing the overlap order parameter for the disordered network (left) and the activity over neurons for the circulant network (right). Arrows give the direction of drift. **c**, Initial angle versus the angle reached at the same five times, across α_{noise} from 0 to 1 (rows), for the disordered (left) and circulant (right) networks (columns). Points on the diagonal have not moved; departures from it measure drift.

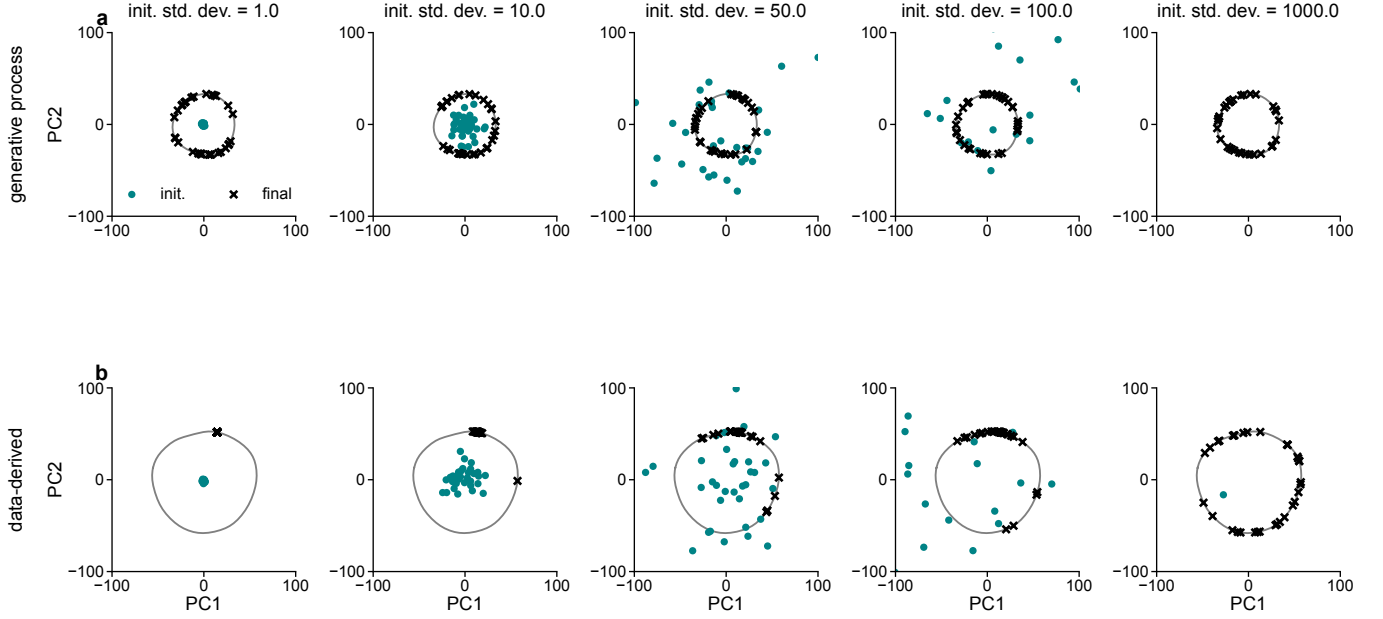
$\sim \sqrt{N} \approx 50$. Below that the initial conditions sit near the origin (a different effect from the inward offset in Fig. 2e, where the activation function rather than the dimensionality of space is responsible).

For each standard deviation we ran the dynamics for 30 s and projected the initial and final states onto the top two principal components of the target manifold (Supplementary Fig. 11).

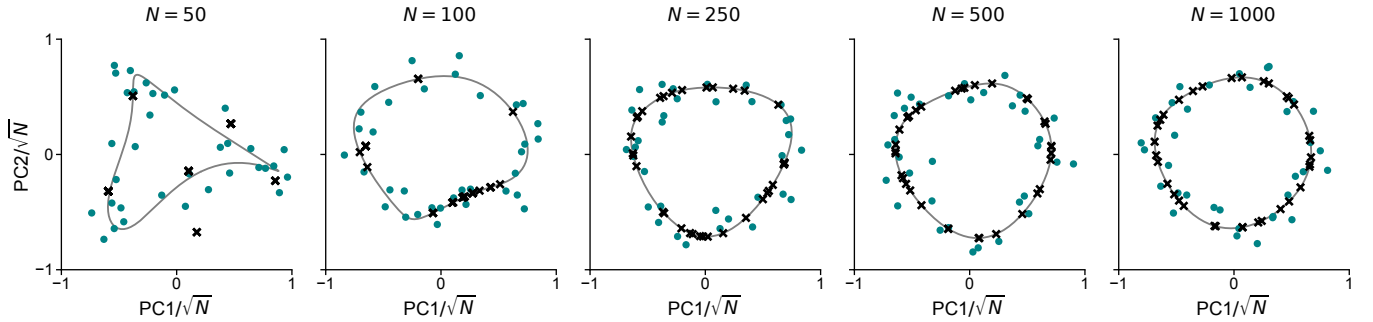
Across the whole range, initial conditions converged to points on the target manifold, with no spurious attractors apparent elsewhere in state space. In the data-derived network, initial conditions close to the origin converge to one or a small number of states on the ring, indicating a subtle asymmetry in the ring that matters when the spread of initial conditions is small.

Finite-size networks approach the mean-field behavior with increasing N (Supplementary Fig. 12). By $N = 1,000$ the target manifold resembles a geometric ring and initial conditions reliably converge to it. At

small N , trajectories sometimes settle at fixed points away from the manifold. No such off-manifold states remain at large N (Supplementary Fig. 11), as the dynamical mean-field theory predicts.



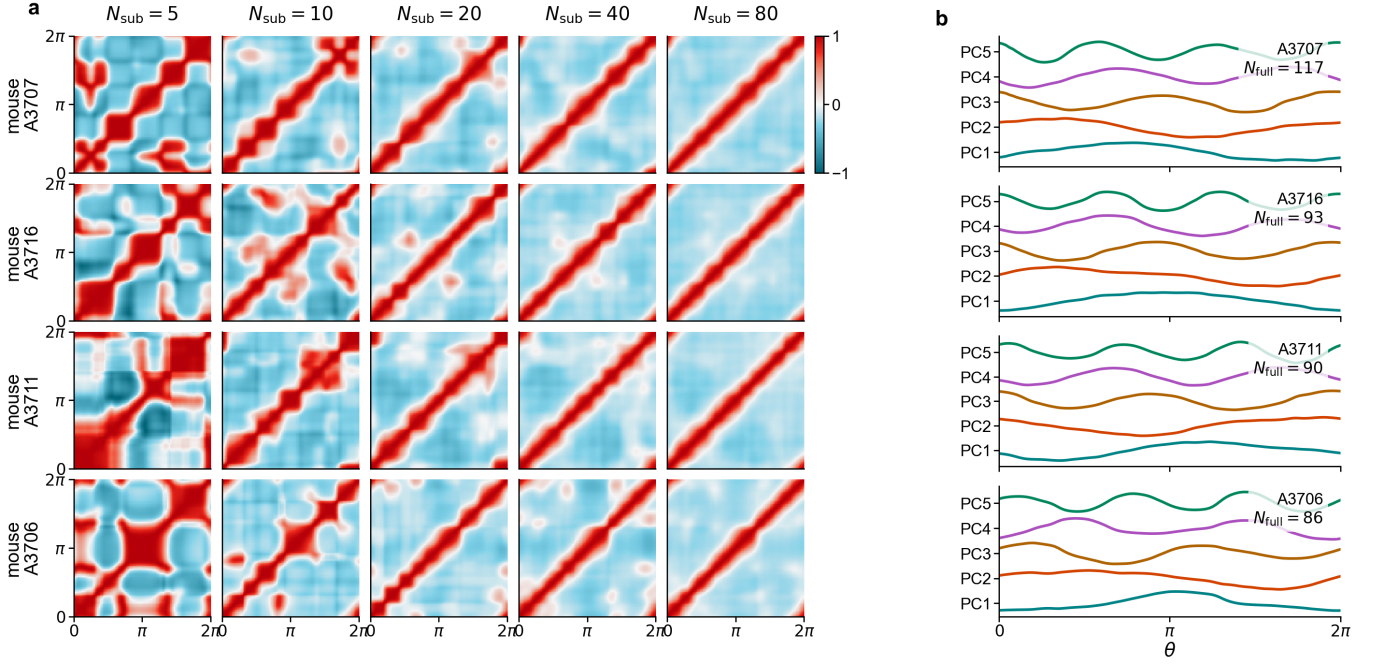
Supplementary Figure 11: Global stability of the target manifold in the generative-process and data-derived networks. Each panel shows initial conditions and final states after 30 s, projected onto the top two principal components of the target manifold (line). The per-neuron standard deviation of the initial conditions (around the origin) increases left-to-right. **a**, Generative-process network ($N = 2,500$, generative-process Fourier decay scale $\sigma = 1.42$, sharpness $\beta = 2.77$, error-function activation). **b**, Data-derived network ($N = N_{\text{data}} = 1,533$, uniform inhibition $c = 1$, softplus activation).



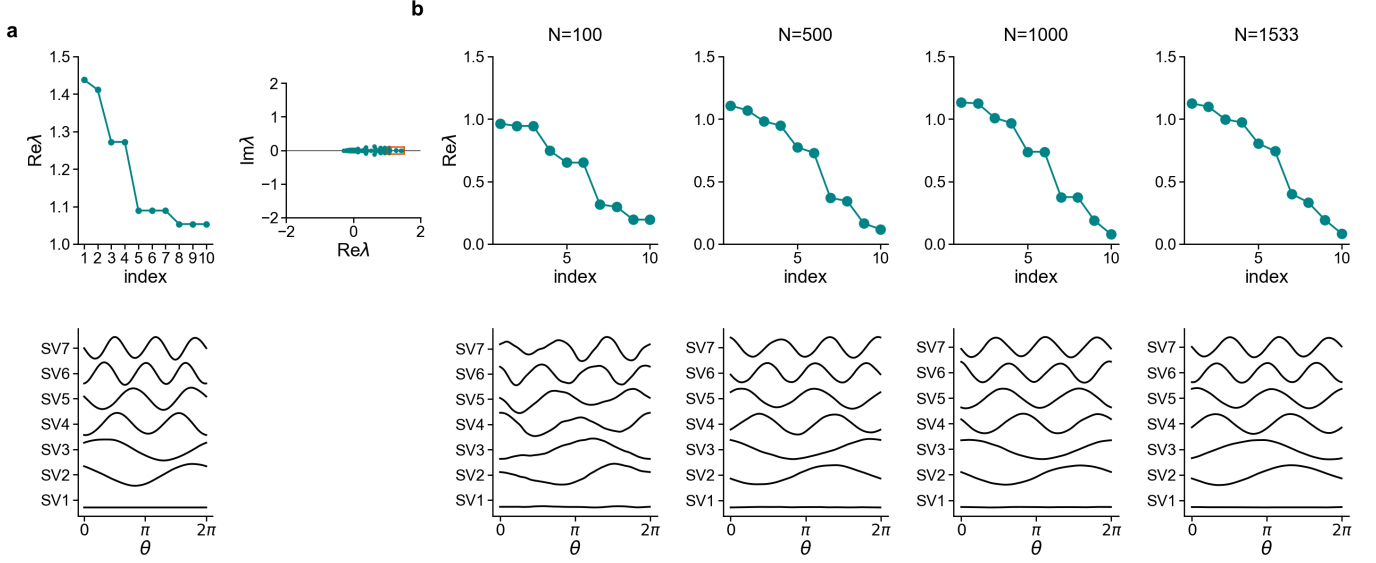
Supplementary Figure 12: Convergence of the dynamics to mean-field behavior with increasing N (generative-process Fourier decay scale $\sigma = 1.42$, sharpness $\beta = 2.77$, error-function activation). Markers, lines, colors, and simulation duration are as in Supplementary Fig. 11. The initial conditions are formed differently: here they are states on the target manifold perturbed by independent Gaussian noise of per-neuron standard deviation $0.1\sqrt{N}$, whereas in Supplementary Fig. 11 they are centered at the origin.

10 Additional figures

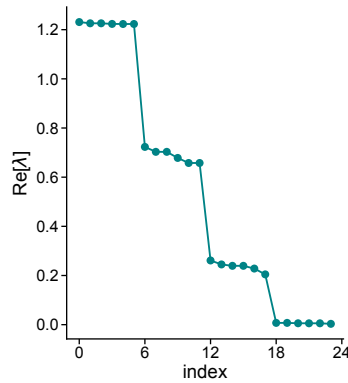
The figures in this section are referred to from the main text rather than from the Supplementary Information.



Supplementary Figure 13: Centered and variance-normalized two-point correlation functions and their eigenvectors in experimental data (four example mice). **a**, Centered and variance-normalized correlation functions over random subsets of N_{sub} neurons where $N_{\text{sub}} \in \{5, 10, 20, 40, 80\}$ (columns), in the same format as Fig. 3c, which shows the uncentered, unnormalized two-point correlation function. **b**, The five leading eigenvectors, PC1 to PC5, of the centered and variance-normalized correlation function of each mouse's full set of neurons, $N_{\text{full}} \in \{117, 93, 90, 86\}$.



Supplementary Figure 14: Spectral signatures in data-derived and finite-sample networks. Each block shows eigenvalues of the weight matrix (top) and left singular vectors of the target firing-rate matrix Φ used to construct it (bottom). Eigenvalues are sorted by descending real part and singular vectors by descending singular value of Φ . **a**, Data-derived optimal weight matrix J of the main-text section ‘Dynamics of the data-derived network model’ ($N = N_{\text{data}} = 1,533$, double-normalized tuning curves, softplus activation; Sec. 8.1). **Top left**: Real parts of the leading eigenvalues. **Top right**: All eigenvalues in the complex plane. The box marks the range plotted at top left. **Bottom**: Left singular vectors of the $N_\theta \times N_{\text{data}}$ matrix Φ . **b**, Networks constructed from generative-process samples of increasing size (generative-process Fourier decay scale $\sigma = 1.42$, sharpness $\beta = 2.77$, error-function activation).



Supplementary Figure 15: Real parts of the leading eigenvalues of the weight matrix J of the grid-cell model ($N = 2,500$, generative-process Fourier decay scale $\sigma = 4.04$, sharpness $\beta = 2.5$, regularization $\lambda = 10^{-8}$, number of periods $\nu = 1$, error-function activation). This network is smaller than the $N = 25,000$ network used for the simulations.

Supplementary Note 4: Dynamics of learned attractors

This Note is self-contained and uses its own notation, set out in Sec. 11.

11	Introduction	40
12	Problem formulation	41
13	Order parameters	42
14	Single-site dynamics	42
15	Differential equations for the order parameters	43
16	Linearization around fixed points	43
16.1	Relation to the response function	45
17	Gaussian feedback weights	46
17.1	Full dynamical mean-field theory for Gaussian weights	46
17.2	Interpretation as an effective recurrent network	46
17.3	Linearization around fixed points with Gaussian weights	47
18	Taxonomy of ring-attractor models	47
18.1	Comparison to the Darshan and Rivkind model	48
19	Path-integral derivation	49

11 Introduction

This Note develops the theory of reservoir-computing recurrent networks [14] that are trained with open-loop, teacher-forced inputs and then run autonomously in closed-loop mode. We focus on the closed-loop dynamics that result from learning open-loop fixed points. Rivkind and Barak [10] analyzed local stability around such fixed points using a response function whose poles correspond to outlier eigenvalues of the linearized dynamics. We go beyond this local, linear picture by deriving the nonlinear dynamics using dynamical mean-field theory.

The resulting order parameters are overlap functions $m^\mu(t)$, which depend on a single time, and correlation functions $C(t, t')$, which depend on two. When the feedback weights are i.i.d. Gaussian, the mean-field equations reduce to those of an effective recurrent network of P units, where P is the number of prescribed fixed points. Each unit carries the overlap $m^\mu(t)$ between the network state at time t and one learned pattern, and the units interact through a matrix determined by the teacher statistics. For $g > 0$, their effective activation function depends on the history of the two-time correlation $C(t, t')$. Linearizing these equations around a fixed point, we show that the order-parameter dynamics are governed by a linearized order-parameter matrix whose eigenvalues determine the poles of the response function computed by Rivkind and Barak [10]. Our aim is to establish the formalism and connect it to prior work; solving the equations and studying their solutions is left to future work.

We treat the general problem of a finite set of prescribed fixed points, indexed by μ , but the formulation specializes to a continuum of them, indexed here by the head direction θ , as Darshan and Rivkind [11] did for ring attractors. The model of the main text is that continuum case at $g = 0$, and Sec. 6.1 derives it directly, without path integrals. Sec. 18 uses the continuum case to organize ring-attractor models into a taxonomy.

Notation follows the main text wherever quantities are analogous, with the following exceptions. What we call $\Gamma^{\mu\nu}$ here is analogous to $\Gamma^\phi(\theta, \theta')$ in the main text, and $\Psi^{\mu\nu}$ is analogous to $\Gamma^x(\theta, \theta')$. Also, W_{ij} denotes the full weight matrix, corresponding to the learned matrix J_{ij} of the main text, while J_{ij}^{res} denotes its random component. Finally, in contrast to the main text, we set the single-neuron time constant to $\tau = 1$, equivalently measuring time in units of τ .

12 Problem formulation

We consider weight matrices of the form

$$W_{ij} = \frac{1}{N} \sum_{k=1}^K U_{ik} V_{jk} + J_{ij}^{\text{res}}, \quad (143)$$

where N is the number of neurons, K is the rank of the low-rank component, and P is the number of training examples. We take the scaling limit $N \rightarrow \infty$ with K and P finite, and we take $\phi(x)$ to be a saturating activation function such as the error function or hyperbolic tangent.

The weight matrix comprises two components, a learned low-rank structure and random reservoir weights. The low-rank component consists of fixed feedback weights U_{ik} and learned readout weights V_{ik} . Letting $\mathbf{u}_i$ denote the i th row of the feedback weights, we assume that averages of the form $\frac{1}{N} \sum_{i=1}^N f(\mathbf{u}_i)$ converge to limiting values, denoted $\langle f(\mathbf{u}) \rangle_{\mathbf{u}}$, as $N \rightarrow \infty$. The readout weights are learned by least-squares regression, as described below. The reservoir weights are sampled independently as $J_{ij}^{\text{res}} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, g^2/N)$, where g scales the reservoir connectivity strength.

During training, the system operates in open-loop mode, with feedback clamped to each of P teacher signals. Let $\bar{z}_k^\mu$ denote the teacher input for example $\mu = 1, \dots, P$. For input μ , the input currents $x_i(t)$ obey the open-loop dynamics

$$(1 + \partial_t)x_i(t) = \sum_{k=1}^K U_{ik} \bar{z}_k^\mu + \sum_{j=1}^N J_{ij}^{\text{res}} \phi(x_j(t)). \quad (144)$$

We assume the teacher inputs are large enough to suppress the chaotic activity that can occur at large g , so that the network settles to fixed points $\bar{x}_i^\mu$ satisfying

$$\bar{x}_i^\mu = \sum_{k=1}^K U_{ik} \bar{z}_k^\mu + \sum_{j=1}^N J_{ij}^{\text{res}} \phi(\bar{x}_j^\mu). \quad (145)$$

The teacher input correlation matrix is

$$\Psi^{\mu\nu} = \sum_{k=1}^K \bar{z}_k^\mu \bar{z}_k^\nu. \quad (146)$$

Once the readout weights have been learned, the system operates in closed-loop mode, with dynamics

$$(1 + \partial_t)x_i(t) = \sum_{k=1}^K U_{ik} z_k(t) + \sum_{j=1}^N J_{ij}^{\text{res}} \phi(x_j(t)), \quad (147)$$

$$z_k(t) = \frac{1}{N} \sum_{i=1}^N V_{ik} \phi(x_i(t)). \quad (148)$$

The readout weights V_{ik} are determined by least-squares regression, so that the closed-loop system regenerates the open-loop fixed-point activity. The loss function is

$$\mathcal{L}(\mathbf{V}) = \frac{1}{2} \sum_{\mu=1}^P \sum_{k=1}^K \left(\underbrace{\frac{1}{N} \sum_{i=1}^N V_{ik} \phi(\bar{x}_i^\mu)}_{\text{closed-loop readout}} - \underbrace{\bar{z}_k^\mu}_{\text{teacher input}} \right)^2 + \frac{\lambda}{2N} \sum_{i=1}^N \sum_{k=1}^K V_{ik}^2, \quad (149)$$

where $\lambda \geq 0$ is a regularization parameter. Setting the gradient with respect to V_{ik} to zero yields

$$V_{ik} = \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu \phi(\bar{x}_i^\nu), \quad (150)$$

$$\Gamma_\lambda = \Gamma + \lambda \mathbf{I}_P, \quad (151)$$

where $\Gamma^{\mu\nu}$ is the activity correlation matrix defined below (Eq. 152). In the limit $\lambda \rightarrow 0$ the training error vanishes, since there are finitely many training examples but infinitely many neurons, and the open-loop fixed points $\bar{x}_i^\mu$ become fixed points of the closed-loop system, though their stability is not guaranteed.

13 Order parameters

To analyze the closed-loop dynamics, we define three order parameters,

$$\Gamma^{\mu\nu} = \frac{1}{N} \sum_{i=1}^N \phi(\bar{x}_i^\mu) \phi(\bar{x}_i^\nu), \quad (152)$$

$$C(t, t') = \frac{1}{N} \sum_{i=1}^N \phi(x_i(t)) \phi(x_i(t')), \quad (153)$$

$$m^\mu(t) = \frac{1}{N} \sum_{i=1}^N \phi(\bar{x}_i^\mu) \phi(x_i(t)), \quad (154)$$

namely the activity correlation matrix between open-loop fixed points, the two-time correlation function of closed-loop activity, and the overlap function measuring alignment between open-loop fixed points and closed-loop states. In terms of $\Gamma^{\mu\nu}$ and $m^\mu(t)$, the closed-loop readout is

$$z_k(t) = \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu(t), \quad (155)$$

where $(\Gamma_\lambda^{-1})^{\mu\nu}$ is the inverse of the regularized correlation matrix (Eq. 151).

14 Single-site dynamics

Dynamical mean-field theory reduces the high-dimensional network dynamics to self-consistent equations for the order parameters. These equations follow from the saddle point of a path integral, derived in Sec. 19, and describe two effective single-neuron, or “single-site,” processes, one for the open-loop fixed points and one for the closed-loop dynamics,

$$\bar{x}^\mu = \sum_{k=1}^K u_k \bar{z}_k^\mu + \bar{\xi}^\mu, \quad (156)$$

$$(1 + \partial_t)x(t) = \sum_{k=1}^K u_k \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu(t) + \xi(t), \quad (157)$$

where $\mathbf{u} = (u_1, \dots, u_K)$ is a single row of the feedback weights U_{ik} . The effective noise terms $\bar{\xi}^\mu$ and $\xi(t)$ are zero-mean Gaussian variables with correlations

$$\langle \bar{\xi}^\mu \bar{\xi}^\nu \rangle_{\bar{\xi}} = g^2 \Gamma^{\mu\nu}, \quad (158)$$

$$\langle \xi(t) \xi(t') \rangle_{\xi} = g^2 C(t, t'), \quad (159)$$

$$\langle \bar{\xi}^\mu \xi(t) \rangle_{\bar{\xi}, \xi} = g^2 m^\mu(t), \quad (160)$$

where $\langle \cdots \rangle_{\bar{\xi}, \xi}$ denotes an average over $\bar{\xi} = (\bar{\xi}^1, \dots, \bar{\xi}^P)$ and $\xi(t)$. The order parameters obey the self-consistency relations

$$\Gamma^{\mu\nu} = \langle \phi(\bar{x}^\mu) \phi(\bar{x}^\nu) \rangle_{\mathbf{u}, \bar{\xi}}, \quad (161)$$

$$C(t, t') = \langle \phi(x(t)) \phi(x(t')) \rangle_{\mathbf{u}, \xi}, \quad (162)$$

$$m^\mu(t) = \langle \phi(\bar{x}^\mu) \phi(x(t)) \rangle_{\mathbf{u}, \bar{\xi}, \xi}. \quad (163)$$

15 Differential equations for the order parameters

To obtain differential equations for the time-dependent order parameters, we define their temporally low-pass filtered counterparts through

$$\left[\prod_{n'=1}^n (1 + \partial_{t_{n'}}) \right] \tilde{f}(t_1, \dots, t_n) = f(t_1, \dots, t_n), \quad (164)$$

where $n \in \{1, 2\}$. Starting the filters at a finite time adds exponentially decaying homogeneous terms determined by the filter and by the microscopic initial conditions. It follows that

$$\langle \tilde{\xi}(t) \tilde{\xi}(t') \rangle_{\bar{\xi}} = g^2 \tilde{C}(t, t'), \quad (165)$$

$$\langle \bar{\xi}^\mu \tilde{\xi}(t) \rangle_{\bar{\xi}, \tilde{\xi}} = g^2 \tilde{m}^\mu(t). \quad (166)$$

Integrating Eq. 157 then expresses the two single-site currents entirely in terms of filtered quantities, the first of which is unchanged (same as Eq. 156),

$$\bar{x}^\mu = \sum_{k=1}^K u_k \bar{z}_k^\mu + \bar{\xi}^\mu, \quad x(t) = \sum_{k=1}^K u_k \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu \tilde{m}^\nu(t) + \tilde{\xi}(t), \quad (167)$$

and the self-consistency relations become, with the relation for $\Gamma^{\mu\nu}$ unchanged (same as Eq. 161),

$$\Gamma^{\mu\nu} = \langle \phi(\bar{x}^\mu) \phi(\bar{x}^\nu) \rangle_{\mathbf{u}, \bar{\xi}}, \quad (168)$$

$$(1 + \partial_t)(1 + \partial_{t'}) \tilde{C}(t, t') = \langle \phi(x(t)) \phi(x(t')) \rangle_{\mathbf{u}, \tilde{\xi}}, \quad (169)$$

with $\bar{x}^\mu$ and $x(t)$ given by Eq. 167.

These form a closed set of differential equations for the order parameters, and the equation for $\tilde{C}(t, t')$ can be integrated numerically on a grid. Nonstationary solutions require initial conditions for the order parameters corresponding to a chosen initial state of the high-dimensional system, for example a state near one of the stored patterns. Alternatively, one may restrict attention to stationary solutions, in which case two-time functions depend only on $t - t'$ and single-time functions are constant. Below we treat these equations analytically only by linearizing around a fixed point, for which the initial conditions are immaterial.

16 Linearization around fixed points

We now analyze the linear dynamics of the order parameters around a fixed point, setting $\lambda \rightarrow 0$ so that the fixed points match the teacher signal exactly and $\Gamma_\lambda^{-1} = \Gamma^{-1}$. With the closed-loop state at fixed point $\mu = \bullet$, the order parameters take the values

$$\tilde{m}^\mu = m^\mu = \Gamma^{\mu\bullet}, \quad \tilde{C} = C = \Gamma^{\bullet\bullet}. \quad (170)$$

As a check, the closed-loop readout at this fixed point is

$$z_k = \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu = \sum_{\mu, \nu=1}^P [\Gamma^{-1}]^{\mu\nu} \bar{z}_k^\mu \Gamma^{\bullet\nu} = \sum_{\mu=1}^P \delta^{\bullet\mu} \bar{z}_k^\mu = \bar{z}_k^\bullet, \quad (171)$$

as expected.

Consider perturbations of the activity, of small amplitude ϵ ,

$$\phi(x_i(t)) = \phi(\bar{x}_i^\bullet) + \epsilon \delta\phi_i(t). \quad (172)$$

The corresponding perturbations to the order parameters are

$$\tilde{m}^\mu(t) = \Gamma^{\mu\bullet} + \epsilon \tilde{h}^\mu(t), \quad (173)$$

$$\tilde{C}(t, t') = \Gamma^{\bullet\bullet} + \epsilon (\tilde{c}(t) + \tilde{c}(t')) + \mathcal{O}(\epsilon^2), \quad (174)$$

where the first-order perturbation to $\tilde{C}(t, t')$ decomposes linearly into a term depending on t and a term depending on t' , both written as $\tilde{c}$. In fact $c(t) = h^\bullet(t)$.

Since the dynamics of the first-order perturbations are linear, we may drop the tildes, and obtain

$$(1 + \partial_t)c(t) = (2g^2\Gamma_{22}^{\bullet\bullet} + g^2\Gamma_{12}^{\bullet\bullet})c(t) + \sum_{\nu=1}^P \Xi^{\bullet\nu} h^\nu(t), \quad (175)$$

$$(1 + \partial_t)h^\mu(t) = 2g^2\Gamma_{22}^{\mu\bullet}c(t) + \sum_{\nu=1}^P (\Xi^{\mu\nu} + g^2\delta^{\mu\nu}\Gamma_{12}^{\mu\bullet})h^\nu(t), \quad (176)$$

where

$$\Gamma_{11}^{\mu\nu} = \frac{1}{2} \left\langle \phi'' \left(\sum_{k=1}^K u_k \bar{z}_k^\mu + \bar{\xi}^\mu \right) \phi \left(\sum_{k=1}^K u_k \bar{z}_k^\nu + \bar{\xi}^\nu \right) \right\rangle_{u, \bar{\xi}}, \quad (177)$$

$$\Gamma_{22}^{\mu\nu} = \frac{1}{2} \left\langle \phi \left(\sum_{k=1}^K u_k \bar{z}_k^\mu + \bar{\xi}^\mu \right) \phi'' \left(\sum_{k=1}^K u_k \bar{z}_k^\nu + \bar{\xi}^\nu \right) \right\rangle_{u, \bar{\xi}}, \quad (178)$$

$$\Gamma_{12}^{\mu\nu} = \left\langle \phi' \left(\sum_{k=1}^K u_k \bar{z}_k^\mu + \bar{\xi}^\mu \right) \phi' \left(\sum_{k=1}^K u_k \bar{z}_k^\nu + \bar{\xi}^\nu \right) \right\rangle_{u, \bar{\xi}}, \quad (179)$$

$$\Xi^{\mu\nu} = \sum_{k=1}^K \left\langle \phi \left(\sum_{\ell=1}^K u_\ell \bar{z}_\ell^\mu + \bar{\xi}^\mu \right) \phi' \left(\sum_{\ell=1}^K u_\ell \bar{z}_\ell^\nu + \bar{\xi}^\nu \right) u_k \right\rangle_{u, \bar{\xi}} \sum_{\rho=1}^P [\Gamma^{-1}]^{\nu\rho} \bar{z}_k^\rho. \quad (180)$$

Here $\Gamma_{11}^{\mu\nu}$ and $\Gamma_{22}^{\mu\nu}$ arise from perturbing the two variances entering the Gaussian averages, and $\Gamma_{12}^{\mu\nu}$ from perturbing the covariance; only Γ_{22} appears above, since $\Gamma_{11}^{\bullet\bullet} = \Gamma_{22}^{\bullet\bullet}$. The $(P+1) \times (P+1)$ linearized order-parameter matrix is therefore

$$\mathbf{T} = \begin{pmatrix} M_{cc} & M_{ch^\nu} \\ M_{h^\mu c} & M_{h^\mu h^\nu} \end{pmatrix} = \begin{pmatrix} 2g^2\Gamma_{22}^{\bullet\bullet} + g^2\Gamma_{12}^{\bullet\bullet} & \Xi^{\bullet\nu} \\ 2g^2\Gamma_{22}^{\mu\bullet} & \Xi^{\mu\nu} + g^2\delta^{\mu\nu}\Gamma_{12}^{\mu\bullet} \end{pmatrix}, \quad (181)$$

and stability requires every eigenvalue of $\mathbf{T}$ to have real part below one.

This matrix has a left eigenvector, that is, a right eigenvector of $\mathbf{T}^T$,

$$\begin{pmatrix} v_c \\ v_{h^\mu} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -\delta^{\mu\bullet} \end{pmatrix}, \quad (182)$$

with eigenvalue $g^2\Gamma_{12}^{\bullet\bullet}$, as direct calculation verifies. It follows that

$$c(t) - h^\bullet(t) = e^{(-1+g^2\Gamma_{12}^{\bullet\bullet})t} (c(0) - h^\bullet(0)). \quad (183)$$

The definitions of the order parameters require $c(t) = h^\bullet(t)$ for every physical perturbation, and the linearized equations preserve this identity, so treating the two as independent introduces a formal transverse mode with no direct physical interpretation. The mode is nonetheless informative, because the factor $g^2\Gamma_{12}^{\bullet\bullet}$ in its decay rate is the squared radius of the bulk of the random part of the Jacobian, $J_{ij}^{\text{res}}\phi'(\bar{x}_j^\bullet)$. The mode therefore reports on bulk-induced activity that is aligned with no particular pattern, and it loses stability exactly when the bulk edge does. Having extracted this, we drop $c(t)$ as redundant.

Setting $c(t) = h^\bullet(t)$ for all t gives the $P \times P$ linearized order-parameter matrix $\mathcal{M}^{\mu\nu} = M_{h^\mu h^\nu} + M_{h^\mu c}\delta^{\nu\bullet}$, or

$$\mathcal{M}^{\mu\nu} = \Xi^{\mu\nu} + g^2\delta^{\mu\nu}\Gamma_{12}^{\mu\bullet} + 2g^2\delta^{\nu\bullet}\Gamma_{22}^{\mu\bullet}. \quad (184)$$

In summary, the first-order perturbations to the overlaps evolve as

$$(1 + \partial_t)h^\mu(t) = \sum_{\nu=1}^P \mathcal{M}^{\mu\nu}h^\nu(t), \quad (185)$$

with $\mathcal{M}^{\mu\nu}$ given by Eq. 184.

16.1 Relation to the response function

We now relate the eigenvalues of $\mathcal{M}^{\mu\nu}$ to the poles of the closed-loop readout response function, which was the output of the calculation of Rivkind and Barak [10]. To match their analysis, we add a source term $w_k(t)$ to the closed-loop dynamics of the full N -dimensional network,

$$(1 + \partial_t)x_i(t) = \sum_{k=1}^K U_{ik} (z_k(t) + w_k(t)) + \sum_{j=1}^N J_{ij}^{\text{res}} \phi(x_j(t)), \quad (186)$$

and define the $K \times K$ response-function matrix, evaluated at a fixed point,

$$\chi_{k\ell}(t - t') = \left. \frac{dz_k(t)}{dw_\ell(t')} \right|_{\mathbf{w}=0}. \quad (187)$$

The prior work computed its Fourier transform $\hat{\chi}_{k\ell}(\omega)$, where here ω denotes frequency rather than angular velocity as in the main text.

With the source term, the first-order perturbations to the overlaps acquire an additional contribution,

$$(1 + \partial_t)h^\mu(t) = \sum_{\nu=1}^P \mathcal{M}^{\mu\nu}h^\nu(t) + \sum_{k=1}^K B_k^\mu w_k(t), \quad (188)$$

where

$$A_k^\mu = \sum_{\nu=1}^P [\Gamma^{-1}]^{\mu\nu} \bar{z}_k^\nu, \quad (189)$$

$$B_k^\mu = \left\langle \phi \left(\sum_{\ell=1}^K u_\ell \bar{z}_\ell^\mu + \bar{\xi}^\mu \right) \phi' \left(\sum_{\ell=1}^K u_\ell \bar{z}_\ell^\bullet + \bar{\xi}^\bullet \right) u_k \right\rangle_{\mathbf{u}, \bar{\xi}}. \quad (190)$$

The matrix $\mathbf{B}$ maps the source into the order-parameter dynamics, while $\mathbf{A}^T$ maps overlap perturbations to readout perturbations, and their product satisfies $\Xi = \mathbf{B}\mathbf{A}^T$. The same source contributes to $c(t)$, so it cancels from the equation for $c(t) - h^\bullet(t)$ and forcing preserves the constraint $c(t) = h^\bullet(t)$.

In Fourier space, we have

$$\hat{\chi}_{k\ell}(\omega) = \frac{1}{2\pi} \sum_{\mu,\nu=1}^P A_k^\mu [(1+i\omega)\mathbf{I}_P - \mathcal{M}]_{\mu\nu}^{-1} B_\ell^\nu. \quad (191)$$

Each eigenvalue ζ_a of $\mathcal{M}$ thus supplies a candidate pole at $\omega_a = i(1 - \zeta_a)$, which is an actual pole provided the corresponding eigenspace survives being sandwiched between $\mathbf{B}$ and $\mathbf{A}^T$. Generically it does, and every eigenvalue then appears as a pole even when $K = 1$ and $P > 1$.

17 Gaussian feedback weights

17.1 Full dynamical mean-field theory for Gaussian weights

We now assume that the feedback rows are $\mathbf{u}_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$, for which the mean-field equations simplify. Let $\mathcal{G}(\cdot, \cdot, \cdot)$ denote the bivariate Gaussian integral of the activation function (same as Eq. 87),

$$\mathcal{G}(C_{11}, C_{22}, C_{12}) = \langle \phi(x_1) \phi(x_2) \rangle \left(\begin{matrix} x_1 \\ x_2 \end{matrix} \right) \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} \right).$$

When applying this to kernels indexed by μ, ν or by t, t' , we abbreviate

$$\mathcal{G}(A^{\mu\nu}) = \mathcal{G}(A^{\mu\mu}, A^{\nu\nu}, A^{\mu\nu}), \quad \mathcal{G}(A(t, t')) = \mathcal{G}(A(t, t), A(t', t'), A(t, t')). \quad (192)$$

The self-consistent equations are then

$$\Gamma^{\mu\nu} = \mathcal{G}(\Psi^{\mu\nu} + g^2 \Gamma^{\mu\nu}), \quad (193)$$

$$(1 + \partial_t)(1 + \partial_{t'}) \tilde{C}(t, t') = \mathcal{G} \left(\sum_{\mu,\nu=1}^P \mathcal{U}^{\mu\nu} \tilde{m}^\mu(t) \tilde{m}^\nu(t') + g^2 \tilde{C}(t, t') \right), \quad (194)$$

$$(1 + \partial_t) \tilde{m}^\mu(t) = \mathcal{G} \left(\Psi^{\mu\mu} + g^2 \Gamma^{\mu\mu}, \sum_{\rho,\sigma=1}^P \mathcal{U}^{\rho\sigma} \tilde{m}^\rho(t) \tilde{m}^\sigma(t) + g^2 \tilde{C}(t, t), \sum_{\nu=1}^P \mathcal{J}^{\mu\nu} \tilde{m}^\nu(t) \right), \quad (195)$$

where we have defined the $P \times P$ matrices

$$\mathcal{J} = \Psi \Gamma_\lambda^{-1} + g^2 \mathbf{I}_P, \quad (196)$$

$$\mathcal{U} = \Gamma_\lambda^{-1} \Psi \Gamma_\lambda^{-1}. \quad (197)$$

These are analogous to the $N_\theta \times N_\theta$ circulant effective interactions and gain kernel that are central to the dynamical mean-field theory of the main text.

17.2 Interpretation as an effective recurrent network

The equations above can be rewritten to bring out their interpretation as an effective recurrent network of P units,

$$(1 + \partial_t) \tilde{m}^\mu(t) = G_{Q(t,t)}^\mu \left(\sum_{\nu=1}^P \mathcal{J}^{\mu\nu} \tilde{m}^\nu(t) \right), \quad (198)$$

with effective activation function (analogous to Eq. 90)

$$G_Q^\mu(h) = \mathcal{G}(\Psi^{\mu\mu} + g^2 \Gamma^{\mu\mu}, Q, h), \quad (199)$$

and with $Q(t, t')$ the two-time correlation function of the input currents (analogous to main-text Eq. 11),

$$Q(t, t') = \sum_{\mu, \nu=1}^P \mathcal{U}^{\mu\nu} \tilde{m}^\mu(t) \tilde{m}^\nu(t') + g^2 \tilde{C}(t, t'), \quad (200)$$

$$(1 + \partial_t)(1 + \partial_{t'}) \tilde{C}(t, t') = \mathcal{G}(Q(t, t')). \quad (201)$$

The units interact through the effective interactions $\mathcal{J}^{\mu\nu}$, which are built from the teacher correlation structure $\Psi^{\mu\nu}$, the regularized inverse firing-rate correlation structure $(\Gamma_\lambda^{-1})^{\mu\nu}$, and the random connectivity strength g^2 . The activation function $G_Q^\mu(h)$ is modulated by $Q(t, t)$, which for $g > 0$ depends on the full history of the system. For $g = 0$ it depends only on the instantaneous state, making the analogy to a conventional recurrent network more direct, although the activation function remains modulated by the global state through $Q(t, t)$.

17.3 Linearization around fixed points with Gaussian weights

The Gaussian assumption also simplifies the linearized order-parameter matrix. Integrating by parts,

$$\begin{aligned} \Xi^{\mu\nu} &= \sum_{k=1}^K \left\langle \phi \left(\sum_{\ell=1}^K u_\ell \bar{z}_\ell^\mu + \bar{\xi}^\mu \right) \phi' \left(\sum_{\ell=1}^K u_\ell \bar{z}_\ell^\bullet + \bar{\xi}^\bullet \right) u_k \right\rangle_{\mathbf{u}, \bar{\xi}} \sum_{\rho=1}^P [\Gamma^{-1}]^{\nu\rho} \bar{z}_k^\rho \\ &= \Gamma_{12}^{\mu\bullet} [\Psi \Gamma^{-1}]^{\mu\nu} + 2\Gamma_{22}^{\mu\bullet} [\Psi \Gamma^{-1}]^{\bullet\nu}, \end{aligned} \quad (202)$$

which gives, with $\mathcal{J} = \Psi \Gamma^{-1} + g^2 \mathbf{I}_P$ (Eq. 196 with $\Gamma_\lambda^{-1} = \Gamma^{-1}$, since $\lambda = 0$),

$$\mathcal{M}^{\mu\nu} = \Gamma_{12}^{\mu\bullet} \mathcal{J}^{\mu\nu} + 2\Gamma_{22}^{\mu\bullet} \mathcal{J}^{\bullet\nu}. \quad (203)$$

18 Taxonomy of ring-attractor models

Specialized to continuous circular manifolds of fixed points, the framework above provides a unifying taxonomy of continuous-attractor models, in which our model, classical models, the model of Darshan and Rivkind [11], and the models of Beiran et al. [2] and Mastrogiuseppe and Ostojic [1] all appear as special cases. The taxonomy is organized by three parameters. For Gaussian embeddings such as ours, rotational symmetry of the tuning curves produces rotational symmetry of the effective recurrent interactions.

We specialize the general framework to a continuous manifold by taking $P \rightarrow \infty$ after $N \rightarrow \infty$. To embed a geometric ring into the network, we define the teacher signal using scaled Fourier modes,

$$\bar{z}_1^\mu = \sqrt{\hat{\Psi}_0}, \quad (204)$$

$$\bar{z}_2^\mu = \sqrt{2\hat{\Psi}_1} \cos(\theta^\mu), \quad \bar{z}_3^\mu = \sqrt{2\hat{\Psi}_1} \sin(\theta^\mu), \quad (205)$$

$$\bar{z}_4^\mu = \sqrt{2\hat{\Psi}_2} \cos(2\theta^\mu), \quad \bar{z}_5^\mu = \sqrt{2\hat{\Psi}_2} \sin(2\theta^\mu), \quad (206)$$

where $\theta^\mu = 2\pi(\mu - 1)/P$ for $\mu = 1, \dots, P$ and $\hat{\Psi}_n$ is the n th Fourier coefficient of the teacher correlation function given below, in the convention of Sec. 2. The pattern continues through harmonic D , giving rank $K = 2D + 1$; if the constant component is absent, the first coordinate is omitted and $K = 2D$. The resulting teacher correlation matrix has the circulant form $\Psi^{\mu\nu} = \Psi(\theta^\mu - \theta^\nu)$, with

$$\Psi(\Delta\theta) = \hat{\Psi}_0 + \sum_{n=1}^D 2\hat{\Psi}_n \cos(n\Delta\theta). \quad (207)$$

- **Number of harmonics D :** the number of nonconstant Fourier modes in the teacher signal.
- **Random connectivity strength g :** controls chaotic fluctuations in the reservoir.
- **Embedding type:** Gaussian (random U_{ik}) or Fourier (structured $U_{ik} = u_k(\theta_i)$, with θ_i an angle assigned to neuron i).

Combining these gives eight possibilities.

1. **$D = 1, g = 0$, Fourier embedding:** the model of Ben-Yishai et al. [8], the classical ring attractor with cosine connectivity.
2. **$D > 1, g = 0$, Fourier embedding:** the classical ring attractor with higher-order Fourier modes in the connectivity.
3. **$D = 1, g > 0$, Fourier embedding:** the model of Darshan and Rivkind [11].
4. **$D > 1, g > 0$, Fourier embedding:** a generalization of Darshan and Rivkind [11], reducing to case 2 as $g \rightarrow 0$ (unpublished).
5. **$D = 1, g = 0$, Gaussian embedding:** the model of Beiran et al. [2], and of Mastrogiuseppe and Ostojic [1] with $g = 0$, up to non-Gaussianity as discussed below; a ring attractor with heterogeneous tuning-curve amplitudes.
6. **$D > 1, g = 0$, Gaussian embedding:** the model of this paper.
7. **$D = 1, g > 0$, Gaussian embedding:** a generalization of Darshan and Rivkind [11], reducing to case 5 as $g \rightarrow 0$ (unpublished).
8. **$D > 1, g > 0$, Gaussian embedding:** a generalization of Darshan and Rivkind [11], reducing to case 6 as $g \rightarrow 0$ (unpublished).

Case 5 requires two comments. First, Beiran et al. [2] and Mastrogiuseppe and Ostojic [1] constructed a rank-two ring attractor using jointly Gaussian loadings with appropriately configured correlations, whereas in case 5 the matrix $\mathbf{U}$ is marginally Gaussian but the joint distribution over $\mathbf{U}$ and $\mathbf{V}$ is non-Gaussian because of the activation function (Sec. 3). As explained there, this difference is minor for $D = 1$, since for both Gaussian and non-Gaussian loadings the symmetry group of the loading distribution is the circle, but it becomes pertinent for $D > 1$, where the two cases lead to different symmetry groups.

Second, Mastrogiuseppe and Ostojic [1] studied the same ring-attractor structure as Beiran et al. [2] with added i.i.d. random connectivity scaled by $g > 0$. This might appear equivalent to case 7, which likewise reduces to case 5 as $g \rightarrow 0$, but the two differ. In Mastrogiuseppe and Ostojic [1] the low-rank and i.i.d. components are independent, analogous to the circulant-plus-noise constructions of the main-text sections ‘Dynamics of the data-derived network model’ and ‘Weights in the large-network limit’ but with a Gaussian rather than Fourier embedding. In the framework underlying the taxonomy, by contrast, learning takes place in an open-loop setting in which the low-rank component is correlated with J_{ij}^{res} .

18.1 Comparison to the Darshan and Rivkind model

Our model achieves tuning heterogeneity through multiple Fourier modes ($D > 1$), without random background weights ($g = 0$), and with a Gaussian embedding, while the Darshan and Rivkind [11] model achieves it through one harmonic ($D = 1$), random background weights ($g > 0$), and a Fourier embedding. These differences have the following consequences.

- **Interpretability of the effective dynamics.** Gaussian feedback weights reduce the mean-field equations to the P -unit form of Eq. 198. A Fourier embedding instead retains an average over $\mathbf{u}$, which in the continuum limit becomes an integral over θ , and does not in general produce the same componentwise activation function. At $g = 0$ the Fourier cases recover classical ring-attractor dynamics directly. For either embedding, $g > 0$ introduces history dependence through $C(t, t')$, so the reduced dynamics are no longer those of a memoryless recurrent network.

- **Stability under increasing heterogeneity.** The Darshan and Rivkind [11] model is not minimum-norm and relies on operating near the chaotic regime to generate tuning heterogeneity. Once tuning curves deviate substantially from cosine-like profiles, chaotic fluctuations destabilize the manifold, preventing its use as a memory system (Supplementary Fig. 16). Strong teacher input is needed for the teacher to dominate the chaotic activity, which is in turn necessary but not sufficient for a stable closed-loop manifold. Unlike the circulant-plus-noise constructions of the main-text sections ‘Dynamics of the data-derived network model’ and ‘Weights in the large-network limit’, however, collapse to a few discrete fixed points is avoided in the non-chaotic regime. Given tuning curves produced by the Darshan and Rivkind [11] model, one could apply the minimum-norm prescription to obtain a version of our model with the same tuning-curve statistics, eliminating the chaotic instability at large heterogeneity. This raises the question of why a neural circuit would implement the Darshan and Rivkind [11] weight matrix at all. One possibility is that operating near chaotic instability is computationally useful in circuits that must combine memory of a continuous variable with dynamic computation.
- **Flexibility for matching experimental data.** The Darshan and Rivkind [11] model produces a one-dimensional family of tuning-curve statistics parameterized by g . Our minimum-norm prescription instead allows arbitrary correlation functions for the input currents. It provides enough freedom to fit experimental head-direction tuning curves quantitatively with only three parameters.

19 Path-integral derivation

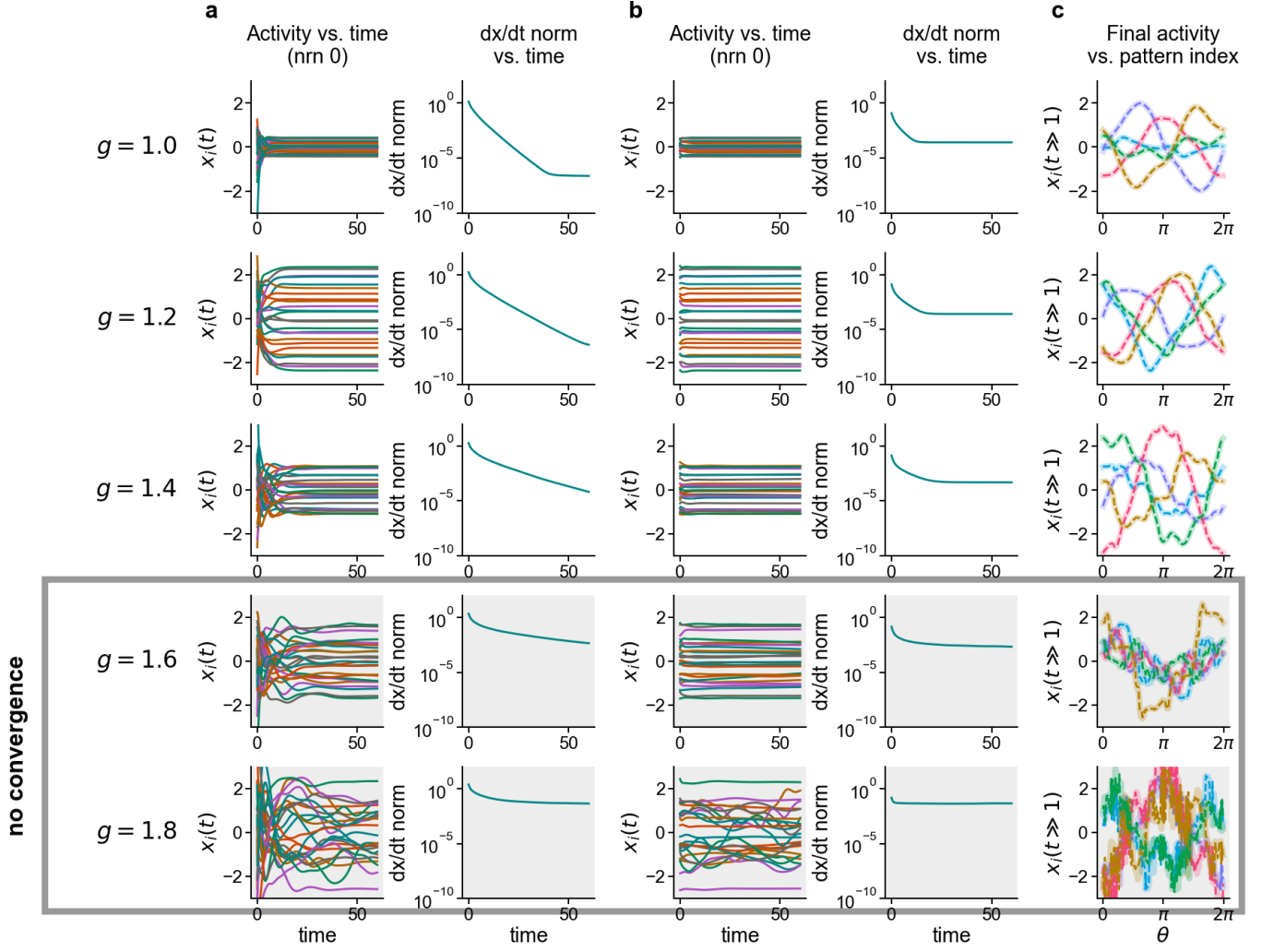
We now write down the path integral, within the Martin–Siggia–Rose–Janssen–De Dominicis framework, that describes both the open- and closed-loop activity [15]. This requires a set of conjugate fields, denoted with hats, which enforce the fixed-point (open-loop) and dynamical (closed-loop) network equations through integral representations of delta functions.

Using the least-squares solution for V_{ik} , the path integral depends on J_{ij}^{res} and U_{ik} and reads

$$\mathcal{Z}[\mathbf{J}^{\text{res}}, \mathbf{U}] = \int \mathcal{D}\bar{\mathbf{x}} \int \mathcal{D}\hat{\mathbf{x}} \int \mathcal{D}\mathbf{x} \int \mathcal{D}\hat{\mathbf{x}} \exp \left\{ \underbrace{i \sum_{\mu=1}^P \sum_{i=1}^N \hat{x}_i^\mu \left(\bar{x}_i^\mu - \sum_{k=1}^K U_{ik} \bar{z}_k^\mu - \sum_{j=1}^N J_{ij}^{\text{res}} \phi(\bar{x}_j^\mu) \right)}_{\text{delta functions enforcing open-loop fixed-point equations}} \right. \\ \left. + \underbrace{i \sum_{i=1}^N \int dt \hat{x}_i(t) \left((1 + \partial_t) x_i(t) - \sum_{k=1}^K U_{ik} \sum_{\mu,\nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu(t) - \sum_{j=1}^N J_{ij}^{\text{res}} \phi(x_j(t)) \right)}_{\text{delta functions enforcing closed-loop equations of motion}} \right\}, \quad (208)$$

where $\hat{x}_i^\mu$ and $\hat{x}_i(t)$ are the conjugate fields enforcing the open- and closed-loop dynamics, respectively.

We first average over the random reservoir weights, which multiply both the open- and closed-loop terms in the action. Averaging over Gaussian J_{ij}^{res} generates cross-terms between these contributions, yielding three types of term, open-open, closed-closed, and open-closed, corresponding to the three order parameters. In particular, the open-closed overlap $m^\mu(t)$ emerges automatically,



Supplementary Figure 16: Simulations of the Darshan and Rivkind model [11] ($N = 1,000$, $P = 250$). Rows are values of the random connectivity strength g . The box marks the two rows in which the dynamics do not converge. **a**, Open-loop operation with the teacher signal. The first column shows the input current of a single neuron, with a line for each of 25 teacher patterns; the second column shows the root mean square of $\partial_t x(t)$ over the N neurons and P patterns. **b**, Closed-loop (autonomous) operation, initialized near the open-loop final states, in the same two formats as in **a**. **c**, Tuning curves defined by the attractor manifold, plotting final activity against θ for five neurons, with open-loop steady states as thick faint lines and closed-loop final states as dashed lines, insofar as a stable manifold is formed.

$$\begin{aligned}
 & \left\langle \exp \left\{ -i \sum_{i,j=1}^N J_{ij}^{\text{res}} \left(\sum_{\mu=1}^P \hat{x}_i^{\mu} \phi(\bar{x}_j^{\mu}) + \int dt \hat{x}_i(t) \phi(x_j(t)) \right) \right\} \right\rangle_{\mathbf{J}^{\text{res}}} \\
 &= \exp \left\{ \sum_{i=1}^N \left(-\frac{g^2}{2} \sum_{\mu,\nu=1}^P \hat{x}_i^{\mu} \underbrace{\left[\frac{1}{N} \sum_{j=1}^N \phi(\bar{x}_j^{\mu}) \phi(\bar{x}_j^{\nu}) \right]}_{\Gamma^{\mu\nu}} \hat{x}_i^{\nu} - \frac{g^2}{2} \int dt \int dt' \hat{x}_i(t) \underbrace{\left[\frac{1}{N} \sum_{j=1}^N \phi(x_j(t)) \phi(x_j(t')) \right]}_{C(t,t')} \hat{x}_i(t') \right. \right. \\
 & \quad \left. \left. - g^2 \sum_{\mu=1}^P \int dt \hat{x}_i^{\mu} \underbrace{\left[\frac{1}{N} \sum_{j=1}^N \phi(\bar{x}_j^{\mu}) \phi(x_j(t)) \right]}_{m^{\mu}(t)} \hat{x}_i(t) \right) \right\}. \quad (209)
 \end{aligned}$$

To express the path integral as a sum over the neuron index i , we introduce the order parameters indicated in

Eq. 209 and enforce their definitions through further integral representations of delta functions. This requires additional conjugate fields $\hat{\Gamma}^{\mu\nu}$, $\hat{C}(t, t')$, and $\hat{m}^\mu(t)$, whose contours implicitly run along the imaginary axis,

$$\begin{aligned}
\mathcal{Z}[U] = & \int \mathcal{D}\bar{x} \int \mathcal{D}\hat{x} \int \mathcal{D}x \int \mathcal{D}\hat{x} \int \mathcal{D}\Gamma \int \mathcal{D}\hat{\Gamma} \int \mathcal{D}C \int \mathcal{D}\hat{C} \int \mathcal{D}m \int \mathcal{D}\hat{m} \\
& \exp \left\{ \underbrace{\frac{N}{2} \sum_{\mu, \nu=1}^P \hat{\Gamma}^{\mu\nu} \left(\Gamma^{\mu\nu} - \frac{1}{N} \sum_{i=1}^N \phi(\bar{x}_i^\mu) \phi(\bar{x}_i^\nu) \right)}_{\text{delta function enforcing } \Gamma^{\mu\nu} \text{ definition}} + \underbrace{\frac{N}{2} \int dt \int dt' \hat{C}(t, t') \left(C(t, t') - \frac{1}{N} \sum_{i=1}^N \phi(x_i(t)) \phi(x_i(t')) \right)}_{\text{delta function enforcing } C(t, t') \text{ definition}} \right. \\
& \quad + \underbrace{\frac{N}{2} \sum_{\mu=1}^P \int dt \hat{m}^\mu(t) \left(m^\mu(t) - \frac{1}{N} \sum_{i=1}^N \phi(\bar{x}_i^\mu) \phi(x_i(t)) \right)}_{\text{delta function enforcing } m^\mu(t) \text{ definition}} \\
& \quad + \sum_{i=1}^N \left(\underbrace{i \sum_{\mu=1}^P \hat{x}_i^\mu \left[\bar{x}_i^\mu - \sum_{k=1}^K U_{ik} \bar{z}_k^\mu \right]}_{\text{remaining open-loop terms}} + \underbrace{i \int dt \hat{x}_i(t) \left[(1 + \partial_t) x_i(t) - \sum_{k=1}^K U_{ik} \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu(t) \right]}_{\text{remaining closed-loop terms}} \right. \\
& \quad \left. \left. - \underbrace{\frac{g^2}{2} \sum_{\mu, \nu=1}^P \hat{x}_i^\mu \Gamma^{\mu\nu} \hat{x}_i^\nu - \frac{g^2}{2} \int dt \int dt' \hat{x}_i(t) C(t, t') \hat{x}_i(t') - g^2 \sum_{\mu=1}^P \int dt \hat{x}_i^\mu m^\mu(t) \hat{x}_i(t)}_{\text{terms from integrating out the disorder}} \right) \right\}. \quad (210)
\end{aligned}$$

This can be written more compactly as

$$\mathcal{Z}[U] = \int \mathcal{D}\Gamma \int \mathcal{D}\hat{\Gamma} \int \mathcal{D}C \int \mathcal{D}\hat{C} \int \mathcal{D}m \int \mathcal{D}\hat{m} \exp \left\{ -N\mathcal{S}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; U] \right\}, \quad (211)$$

with the order-one action

$$\begin{aligned}
\mathcal{S}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; U] = & -\frac{1}{2} \sum_{\mu, \nu=1}^P \hat{\Gamma}^{\mu\nu} \Gamma^{\mu\nu} - \frac{1}{2} \int dt \int dt' \hat{C}(t, t') C(t, t') - \frac{1}{2} \sum_{\mu=1}^P \int dt \hat{m}^\mu(t) m^\mu(t) \\
& - \frac{1}{N} \sum_{i=1}^N \log \mathcal{W}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; \mathbf{u}_i]. \quad (212)
\end{aligned}$$

The single-site path integral $\mathcal{W}[\dots; \mathbf{u}]$, which depends on a single row $\mathbf{u}$ of the feedback weight matrix, is

$$\begin{aligned}
\mathcal{W}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; \mathbf{u}] = & \int \mathcal{D}\bar{x} \int \mathcal{D}\hat{x} \int \mathcal{D}x \int \mathcal{D}\hat{x} \\
& \exp \left\{ i \sum_{\mu=1}^P \hat{x}^\mu \left(\bar{x}^\mu - \sum_{k=1}^K u_k \bar{z}_k^\mu \right) + i \int dt \hat{x}(t) \left((1 + \partial_t) x(t) - \sum_{k=1}^K u_k \sum_{\mu, \nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu(t) \right) \right. \\
& - \frac{1}{2} \sum_{\mu, \nu=1}^P \phi(\bar{x}^\mu) \hat{\Gamma}^{\mu\nu} \phi(\bar{x}^\nu) - \frac{1}{2} \int dt \int dt' \phi(x(t)) \hat{C}(t, t') \phi(x(t')) - \frac{1}{2} \sum_{\mu=1}^P \int dt \phi(\bar{x}^\mu) \hat{m}^\mu(t) \phi(x(t)) \\
& \left. - \frac{g^2}{2} \sum_{\mu, \nu=1}^P \hat{x}^\mu \Gamma^{\mu\nu} \hat{x}^\nu - \frac{g^2}{2} \int dt \int dt' \hat{x}(t) C(t, t') \hat{x}(t') - g^2 \sum_{\mu=1}^P \int dt \hat{x}^\mu m^\mu(t) \hat{x}(t) \right\}. \quad (213)
\end{aligned}$$

The final three terms can be reformulated using jointly Gaussian random fields. Introducing a Gaussian field $\xi(t)$ and a Gaussian vector $\bar{\xi}^\mu$ with zero means and the covariances of Eqs. 158–160, the single-site path integral becomes

$$\begin{aligned}
\mathcal{W}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; \mathbf{u}] &= \int \mathcal{D}\bar{x} \int \mathcal{D}\hat{x} \int \mathcal{D}x \int \mathcal{D}\hat{x} \\
&\left\langle \exp \left\{ i \sum_{\mu=1}^P \hat{x}^\mu \left(\bar{x}^\mu - \sum_{k=1}^K u_k \bar{z}_k^\mu - \bar{\xi}^\mu \right) + i \int dt \hat{x}(t) \left((1 + \partial_t)x(t) - \sum_{k=1}^K u_k \sum_{\mu,\nu=1}^P (\Gamma_\lambda^{-1})^{\mu\nu} \bar{z}_k^\mu m^\nu(t) - \xi(t) \right) \right. \right. \\
&\left. \left. - \frac{1}{2} \sum_{\mu,\nu=1}^P \phi(\bar{x}^\mu) \hat{\Gamma}^{\mu\nu} \phi(\bar{x}^\nu) - \frac{1}{2} \int dt \int dt' \phi(x(t)) \hat{C}(t, t') \phi(x(t')) - \frac{1}{2} \sum_{\mu=1}^P \int dt \phi(\bar{x}^\mu) \hat{m}^\mu(t) \phi(x(t)) \right\} \right\rangle_{\bar{\xi}, \xi}. \quad (214)
\end{aligned}$$

These transformations are exact at finite N . We now take $N \rightarrow \infty$, where the path integral is dominated by the saddle point at which the derivatives of the action vanish. In anticipation of this limit we make the replacement

$$\frac{1}{N} \sum_{i=1}^N \log \mathcal{W}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; \mathbf{u}_i] \rightarrow \left\langle \log \mathcal{W}[\Gamma, \hat{\Gamma}, C, \hat{C}, m, \hat{m}; \mathbf{u}] \right\rangle_{\mathbf{u}}, \quad (215)$$

which removes the remaining i dependence. The saddle-point derivatives are

$$\frac{\delta \mathcal{S}}{\delta \hat{\Gamma}^{\mu\nu}} = -\Gamma^{\mu\nu} + \left\langle \langle \phi(\bar{x}^\mu) \phi(\bar{x}^\nu) \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad (216)$$

$$\frac{\delta \mathcal{S}}{\delta \Gamma^{\mu\nu}} = -\hat{\Gamma}^{\mu\nu} + g^2 \left\langle \langle \hat{x}^\mu \hat{x}^\nu \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}} + i \left\langle \left\langle \int ds \hat{x}(s) \sum_{\sigma=1}^P m^\sigma(s) \frac{\partial q^\sigma(\mathbf{u})}{\partial \Gamma^{\mu\nu}} \right\rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad (217)$$

$$\frac{\delta \mathcal{S}}{\delta \hat{C}(t, t')} = -C(t, t') + \left\langle \langle \phi(x(t)) \phi(x(t')) \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad (218)$$

$$\frac{\delta \mathcal{S}}{\delta C(t, t')} = -\hat{C}(t, t') + g^2 \left\langle \langle \hat{x}(t) \hat{x}(t') \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad (219)$$

$$2 \frac{\delta \mathcal{S}}{\delta \hat{m}^\mu(t)} = -m^\mu(t) + \left\langle \langle \phi(\bar{x}^\mu) \phi(x(t)) \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad (220)$$

$$2 \frac{\delta \mathcal{S}}{\delta m^\mu(t)} = -\hat{m}^\mu(t) + 2i \left\langle \langle \hat{x}(t) q^\mu(\mathbf{u}) \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}} + 2g^2 \left\langle \langle \hat{x}^\mu \hat{x}(t) \rangle_{\mathcal{W}[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad (221)$$

where $q^\nu(\mathbf{u}) = \sum_{k=1}^K u_k \sum_{\rho=1}^P (\Gamma_\lambda^{-1})^{\rho\nu} \bar{z}_k^\rho$, and where a derivative with respect to a symmetric object acts on both orderings of its arguments, so on (t', t) as well as (t, t') and on (ν, μ) as well as (μ, ν) .

At zero conjugate fields, the causal single-site generating functional is normalized to a constant, since it then merely integrates over a delta function. Any average containing only response fields therefore vanishes, because each response field acts as a derivative with respect to a source conjugate to it, and the functional is constant for every value of that source. The saddle-point equations for the conjugate order parameters set each one equal to a sum of such response-only averages, so setting them all to zero is self-consistent. Marking the saddle-point values with a $*$ subscript,

$$\Gamma_*^{\mu\nu} = \left\langle \langle \phi(\bar{x}^\mu) \phi(\bar{x}^\nu) \rangle_{\mathcal{W}^*[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad \hat{\Gamma}_*^{\mu\nu} = 0, \quad (222)$$

$$C_*(t, t') = \left\langle \langle \phi(x(t)) \phi(x(t')) \rangle_{\mathcal{W}^*[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad \hat{C}_*(t, t') = 0, \quad (223)$$

$$m_*^\mu(t) = \left\langle \langle \phi(\bar{x}^\mu) \phi(x(t)) \rangle_{\mathcal{W}^*[\mathbf{u}]} \right\rangle_{\mathbf{u}}, \quad \hat{m}_*^\mu(t) = 0, \quad (224)$$

where $\langle \cdots \rangle_{\mathcal{W}^*[\mathbf{u}]}$ denotes an average within the process described by $\mathcal{W}[\Gamma_*, 0, C_*, 0, m_*, 0; \mathbf{u}]$. Since we consider only saddle-point values, we drop the $*$ subscript throughout.

References

- [1] Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.
- [2] Manuel Beiran, Alexis Dubreuil, Adrian Valente, Francesca Mastrogiuseppe, and Srdjan Ostojic. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Computation*, 33(6):1572–1615, 2021.
- [3] Alexis Dubreuil, Adrian Valente, Manuel Beiran, Francesca Mastrogiuseppe, and Srdjan Ostojic. The role of population structure in computations through neural dynamics. *Nature Neuroscience*, 25(6):783–794, 2022.
- [4] David G Clark and Manuel Beiran. Structure of activity in multiregion recurrent neural networks. *Proceedings of the National Academy of Sciences*, 122(10):e2404039122, 2025.
- [5] Louis Pezon, Valentin Schmutz, and Wulfram Gerstner. Linking neural manifolds to circuit structure in recurrent networks. *Neuron*, 114(9):1682–1694, 2026.
- [6] Daniel J Amit, Hanoach Gutfreund, and Haim Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32(2):1007, 1985.
- [7] Kai Fong Ernest Chong. A closer look at the approximation capabilities of neural networks. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=rkevSgrtPr>.
- [8] Rani Ben-Yishai, R Lev Bar-Or, and Haim Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848, 1995.
- [9] Barak Blumenfeld, Dmitri Bibitchkov, and Misha Tsodyks. Neural network model of the primary visual cortex: from functional architecture to lateral connectivity and back. *Journal of Computational Neuroscience*, 20(2):219–241, 2006.
- [10] Alexander Rivkind and Omri Barak. Local dynamics in trained recurrent neural networks. *Physical Review Letters*, 118(25):258101, 2017.
- [11] Ran Darshan and Alexander Rivkind. Learning to represent continuous variables in heterogeneous neural networks. *Cell Reports*, 39(1):110612, 2022.
- [12] Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Computational Biology*, 5(2):e1000291, 2009.
- [13] Alexis Guanella, Daniel Kiper, and Paul Verschure. A model of grid cells based on a twisted torus topology. *International Journal of Neural Systems*, 17(04):231–240, 2007.
- [14] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.
- [15] A Crisanti and H Sompolinsky. Path integral approach to random neural networks. *Physical Review E*, 98(6):062120, 2018.