

Symmetries and continuous attractors in disordered neural circuits

Received: 28 January 2025

Accepted: 13 August 2026

Published online: 14 October 2026



David G. Clark^{1,2,3,6}✉, L. F. Abbott^{1,2} & Haim Sompolinsky^{3,4,5}

A core challenge in neuroscience is reconciling idealized models with experimental data. Classical continuous-attractor models describe neural representations of continuous variables such as head direction. Symmetry in the weights generates a manifold of stable states, with tuning curves identical up to shifts. Rodent head-direction cells, by contrast, are markedly heterogeneous. Such heterogeneity, we show, is nevertheless compatible with these models. Here we build, directly from head-direction cell data, recurrent networks that behave as quasicontinuous attractors. A generative process then reproduces the heterogeneous tuning of these cells while making connectivity and dynamics tractable at large N . Although the weights appear disordered, they carry a circular symmetry whose eigenvalue degeneracies could signify continuous-attractor organization in connectome data. Dynamical mean-field theory recovers the ingredients of classical ring attractors, namely, Mexican hat interactions and a spontaneously broken symmetry producing bump states. Continuous-attractor mechanisms may thus operate in complex mammalian circuits.

Organisms from flies to humans maintain mental representations of continuous variables, including those relevant for navigation, such as self-orientation and location. Continuous-attractor models explain how networks of neurons, each individually lacking memory, sustain these representations through collective dynamics. In these models, the weights generate a manifold of stable states matching the group structure of the encoded variable, providing working memory and, with velocity inputs, integration of this variable.

Specifically, classical continuous-attractor models build such manifolds from three ingredients¹. First, neurons must transform input currents into firing rates nonlinearly. Second, the weights must support Turing pattern formation, typically through a Mexican hat profile with local excitation and longer-range inhibition (Fig. 1b). Third, these weights must possess a continuous symmetry corresponding to the encoded variable (circular for head direction, two-dimensional for grid-cell responses and so on; Fig. 1a). The essential mechanism is the spontaneous breaking of this symmetry, after which the resulting

stable states form a continuous manifold inheriting the symmetry of the weights and thus of the encoded variable. One direct consequence is that tuning curves are identical up to neuron-specific shifts (Fig. 1c). The canonical example is the ring-attractor model for circular variables such as head direction or stimulus orientation^{2–5}, realized in the central complex of *Drosophila melanogaster*^{6–9} and, more recently, in zebrafish¹⁰.

We ask how mammalian brains represent continuous variables. Rodents possess head-direction cells, analogous to the fly heading representation, in regions including the anterodorsal thalamic nucleus and the postsubiculum^{11,12}, and grid cells, which represent location through hexagonally arranged firing fields, in the medial entorhinal cortex^{13,14}. Whether these are instances of the classical mechanism remains open. A hallmark prediction of continuous-attractor models is that intrinsic circuit structure confines activity to a low-dimensional manifold even without sensory drive. In agreement with this prediction, manifold confinement persists during sleep in both head-direction

¹Zuckerman Institute, Columbia University, New York, NY, USA. ²Kavli Institute for Brain Science, Columbia University, New York, NY, USA. ³Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, MA, USA. ⁴Center for Brain Science, Harvard University, Cambridge, MA, USA. ⁵Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem, Jerusalem, Israel. ⁶Present address: Center for Computational Neuroscience, Flatiron Institute, New York, NY, USA. ✉e-mail: dclark@flatironinstitute.org

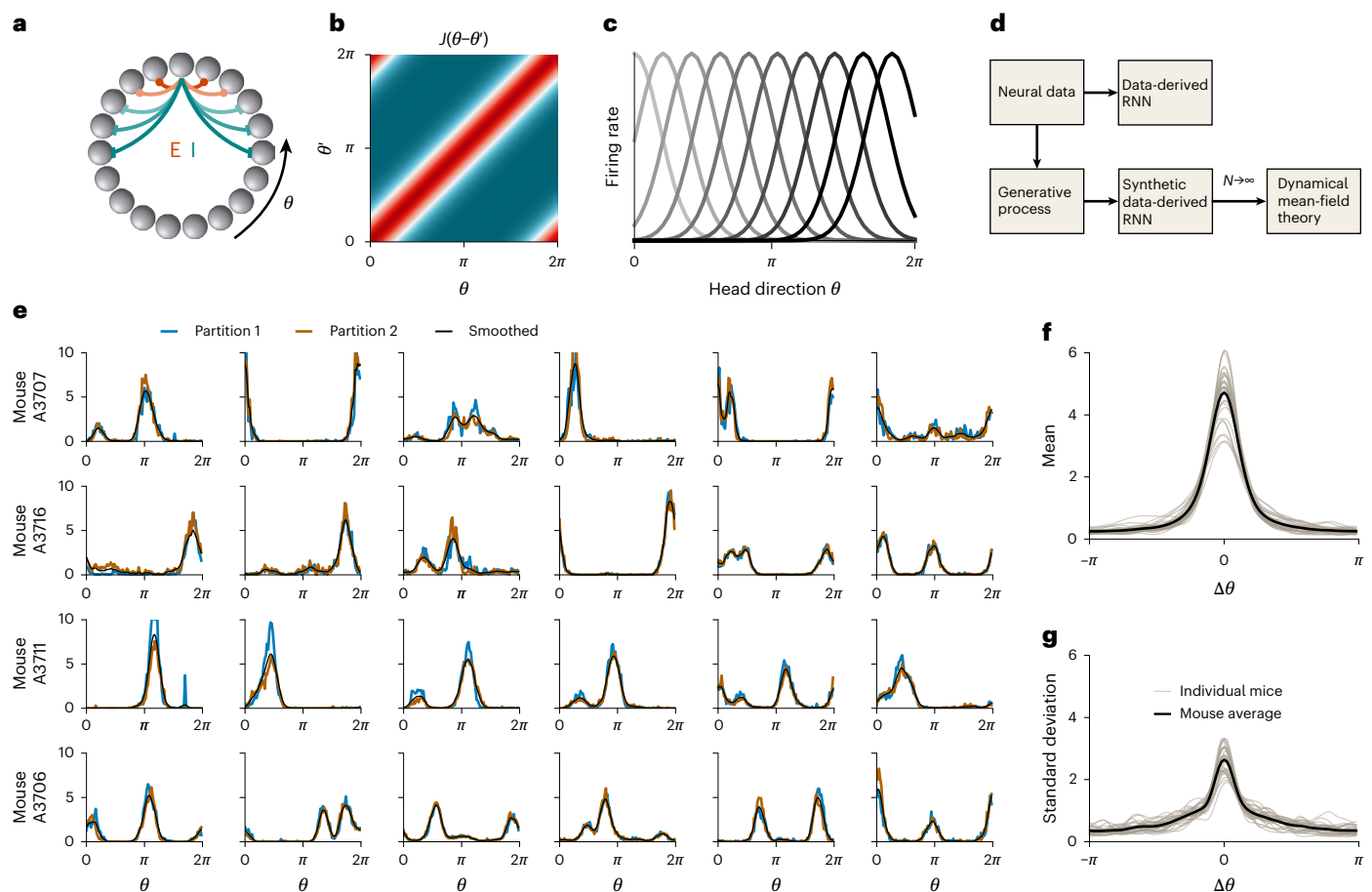


Fig. 1 | Classical ring-attractor models and postsubicular head-direction cell data. **a**, Schematic of a classical ring-attractor model with rotation-invariant weights. E, excitatory; I, inhibitory. **b**, Circular, Mexican hat weight profile characteristic of classical ring-attractor models. **c**, Tuning curves predicted by classical ring-attractor models, identical in shape up to neuron-specific shifts and tilting the θ axis. **d**, Schematic of the analyses. Neural data are used to build a data-derived recurrent neural network (RNN) model and to fit a generative process, which in turn allows larger network models and dynamical mean-field

analysis. **e**, Randomly selected single-neuron tuning curves (arbitrary units) from postsubicular head-direction cells^{16,39} for the four mice with the most recorded neurons. Partitions 1 and 2 are unsmoothed curves from two disjoint partitions of the recording; smoothed curves result from cross-validated smoothing (Methods). **f, g**, Mean (**f**) and standard deviation (**g**) of center-of-mass-aligned tuning curves, taken across neurons within each mouse; mouse average is across all $n = 31$ mice, $N_{\text{data}} = 1,533$ neurons in total.

cells, where the manifold has ring topology^{15,16}, and grid cells, where it has toroidal topology^{17–20}.

We now revisit the list of classical continuous-attractor ingredients. Mammalian circuits possess the requisite nonlinear neurons and strong recurrent excitation and inhibition, yet, at least in the regions we consider, their tuning to continuous variables is markedly different. Whereas classical weights make tuning curves identical up to shifts, these mammalian responses are diverse. This heterogeneity appears in the rodent head-direction system, which we examine in detail; in the medial entorhinal cortex, where the pristine periodicity of classical models is confined to a small subset of grid cells^{21–23}; and elsewhere in the brain²⁴.

These observations raise two interconnected questions:

- (1) What is the nature of heterogeneous tuning in biological neural circuits?
- (2) To what extent are classical models, based on dynamical breaking of continuous symmetry, relevant to such circuits?

Question (2) has practical implications given the growing availability of large-scale electron microscopy reconstructions of synaptic connectivity, or ‘connectomes’, in the mammalian cortex^{25,26}. In particular, while classical continuous-attractor models exhibit easily identifiable

connectivity structure, such as the circulant weight matrix of a ring attractor, it remains unknown which features, if any, are conserved between classical models and heterogeneous circuits and therefore what connectome analyses should target.

Recent work has begun to address these questions separately. Regarding question (1), Mainali et al.²⁷ modeled a population of hippocampal place cells as nonlinearly transformed Gaussian-process samples, capturing topological features of place fields across dimensionalities and environments; we take a similar approach here. Regarding question (2), several constructions of networks with heterogeneous continuous-attractor manifolds already exist^{28–32}, and constructability is in any case guaranteed on general grounds by the universal approximation capability of recurrent networks³³ (Supplementary Note 1). However, prior works were unconstrained by the statistical structure of real data and did not establish a relationship to classical models or identify which connectivity features are shared with them.

This paper addresses questions (1) and (2) within a unified framework, showing that heterogeneous mammalian circuits could rely on the same symmetry-based mechanism as classical continuous-attractor models (Fig. 1d). We focus on the rodent head-direction system. Examining recordings from postsubicular head-direction cells in freely moving mice¹⁶, we first show that their tuning heterogeneity rules out the

recorded neurons themselves forming a classical ring-attractor circuit. Their tuning could instead arise from feedforward input from a classical ring attractor elsewhere^{34,35}, from distributed dynamics across regions or from local recurrence. Whichever holds, understanding how recurrent connectivity alone can generate heterogeneous continuous-attractor manifolds is broadly relevant given the prevalence of such tuning across mammalian regions²⁴. We therefore build a recurrent network directly from the measured tuning curves, as has been done in other systems^{36–38}, and find that it behaves as a continuous attractor.

To go beyond the population sizes available experimentally, we develop a statistical generative process for sampling synthetic tuning curves, addressing question (1). A circularly symmetric Gaussian process motivated by the experimental tuning-curve distribution reproduces the statistical structure of biological head-direction cells quantitatively across multiple measures. We then compute and analyze optimal weight matrices at large N that generate such tuning. Although these matrices appear disordered, they encode an underlying circular symmetry, which manifests as doubly degenerate eigenvalues. The symmetry is realized differently than in classical models² and other constructions²⁸, which embed it along orderly Fourier eigenvectors, whereas ours embeds it along random, disorderly eigenvectors. For connectomics, this contrast implies that eigenvalue degeneracies, rather than eigenvector structure, could be a more robust signifier of continuous-attractor organization.

Finally, we develop a dynamical mean-field theory for these networks, yielding a central result that answers question (2). The effective large- N description recovers all three ingredients of classical continuous-attractor models, namely a nonlinear activation function, Mexican hat interactions and continuous symmetry. Spontaneous symmetry breaking then generates stable bump states through a Turing instability. Angular velocity inputs are faithfully integrated within these mean-field dynamics. We extend this framework to higher dimensions by modeling grid cells in the medial entorhinal cortex, which are thought to constitute a toroidal attractor manifold.

Results

Failure of classical ring-attractor models

We analyzed recordings of head-direction cells from the postsubiculum of freely moving mice¹⁶ (data are publicly available at ref. 39). Duszkievicz et al.¹⁶ emphasized the relative stereotypy of these cells by contrasting them with more heterogeneous fast-spiking neurons. Even these comparatively stereotyped cells, however, depart markedly from classical predictions, as we show below.

We analyzed $N_{\text{data}} = 1,533$ excitatory neurons, pooled across 31 mice, identified as head-direction cells by Duszkievicz et al.¹⁶ on the basis of significant head-direction information, with no constraint on tuning-curve shape such as unimodality. Tuning curves were computed as occupancy-normalized spike counts in 100 angular bins, denoised by cross-validated Gaussian smoothing and normalized to unit mean across angles to compare across neurons and animals (Methods and Fig. 1e).

Classical ring-attractor models (Fig. 1a,b) predict tuning curves identical up to neuron-specific angular shifts (Fig. 1c). Mouse head-direction cells instead show substantial heterogeneity, with variable numbers of irregularly spaced peaks (Fig. 1e). When tuning curves are aligned within each of the 31 mice by their circular centers of mass, the mean profiles are unimodal and symmetric (Fig. 1f), but the fluctuations about them are comparable in magnitude to the means (Fig. 1g), violating the classical prediction of zero fluctuations. Both mean and fluctuation profiles are consistent across mice, suggesting a universal statistical structure.

Constructing a data-derived network model

To test whether intrinsic recurrent connectivity can nonetheless generate the observed responses as a continuous-attractor manifold, we

constructed network models from the experimental tuning curves using an optimization principle (Methods).

We modeled a rate-based recurrent network of N neurons. Neuron i has input current $x_i(t)$ and firing rate $\phi_i(t) = \phi(x_i(t))$, with

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

where J_{ij} is the weight from neuron j to i and b is a bias current. We use bold symbols for vectors. The decay term $-x_i(t)$ makes an isolated neuron relax exponentially to rest with timescale $\tau = 50$ ms, so persistence over seconds, minutes or hours must arise from collective network dynamics. We chose $\phi(x)$ to be a ‘softplus’, interpolating between zero firing for strongly negative inputs and linear, unit-gain firing for strongly positive inputs (Methods and equation (18)).

The normalized tuning curves define a target manifold $\phi_i^*(\theta)$, $i = 1, \dots, N_{\text{data}}$, embedded in neuronal space and parameterized by θ , whose periodicity gives the manifold ring topology. Applying the inverse activation function yields the corresponding input-current manifold $x_i^*(\theta)$. These target patterns are a priori unrelated to the dynamical variables $\phi_i(t)$ and $x_i(t)$. We choose J to confine the dynamics to the manifold by asking the flow to vanish everywhere on it, $\partial_\theta x_i(t) = 0$ at every θ .

To prevent runaway activity under the nonsaturating activation function, we first normalized the target manifold across neurons so that the neuron-averaged firing rate is unity for all θ (Supplementary Note 3 and Supplementary Fig. 7). On the manifold, the right-hand side of equation (1) is then invariant under $J_{ij} \rightarrow J_{ij} - c/N$ combined with $b \rightarrow b + c$. We optimized the weights with $b = 0$ and afterward chose $c > 0$ large enough to supply the uniform inhibition that stabilizes the mean-activity mode (Supplementary Note 3).

We optimized the weights for the vanishing-flow condition by least squares, minimizing the loss function

$$\mathcal{L}(J) = \frac{1}{N} \frac{1}{2\pi} \int_0^{2\pi} d\theta \sum_{i=1}^N \left(\sum_{j=1}^N J_{ij} \phi_j^*(\theta) - x_i^*(\theta) \right)^2 + \lambda \sum_{ij} J_{ij}^2, \quad (2)$$

subject to $J_{ii} = 0$. The first term makes the flow vanish when activity lies on the target manifold (equation (1)), whereas the second penalizes large weights. The solution (Methods) is, for $i \neq j$,

$$J_{ij} = \frac{1}{N} \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' K_{i,\lambda}^{-1}(\theta, \theta') x_i^*(\theta) \phi_j^*(\theta'), \quad (3)$$

where $K_{i,\lambda}(\theta, \theta') = C^\phi(\theta, \theta') - N^{-1} \phi_i^*(\theta) \phi_i^*(\theta') + 2\pi\lambda\delta(\theta - \theta')$. The second term is related to the constraint $J_{ii} = 0$ and vanishes at large N . The first term,

$$C^\phi(\theta, \theta') = \frac{1}{N} \sum_{i=1}^N \phi_i^*(\theta) \phi_i^*(\theta'), \quad (4)$$

is the two-point correlation function of the target firing rates. Equation (3) is Hebbian in form, with $\phi_j^*(\theta')$ and $x_i^*(\theta)$ the presynaptic and postsynaptic activities, respectively. Because the activation function enters only the presynaptic factor, the weights are not transpose-symmetric, $J_{ij} \neq J_{ji}$, and the dynamics admit no energy function^{40,41}.

The problem appears overconstrained, because continuously varying θ imposes infinitely many constraints on the N weights in each row. The smoothness of the tuning curves, however, imposes an effective frequency cutoff at some $D \ll N$, which reduces the number of independent constraints to $2D + 1$ (‘Weights in the large-network limit’). The problem is therefore underconstrained, and infinitely many weight matrices achieve zero flow on the manifold. As $\lambda \rightarrow 0$, the optimization selects the unique minimum-norm (pseudoinverse) solution.

Biological head-direction systems integrate angular velocity. Classical models achieve this by adding a term proportional to the

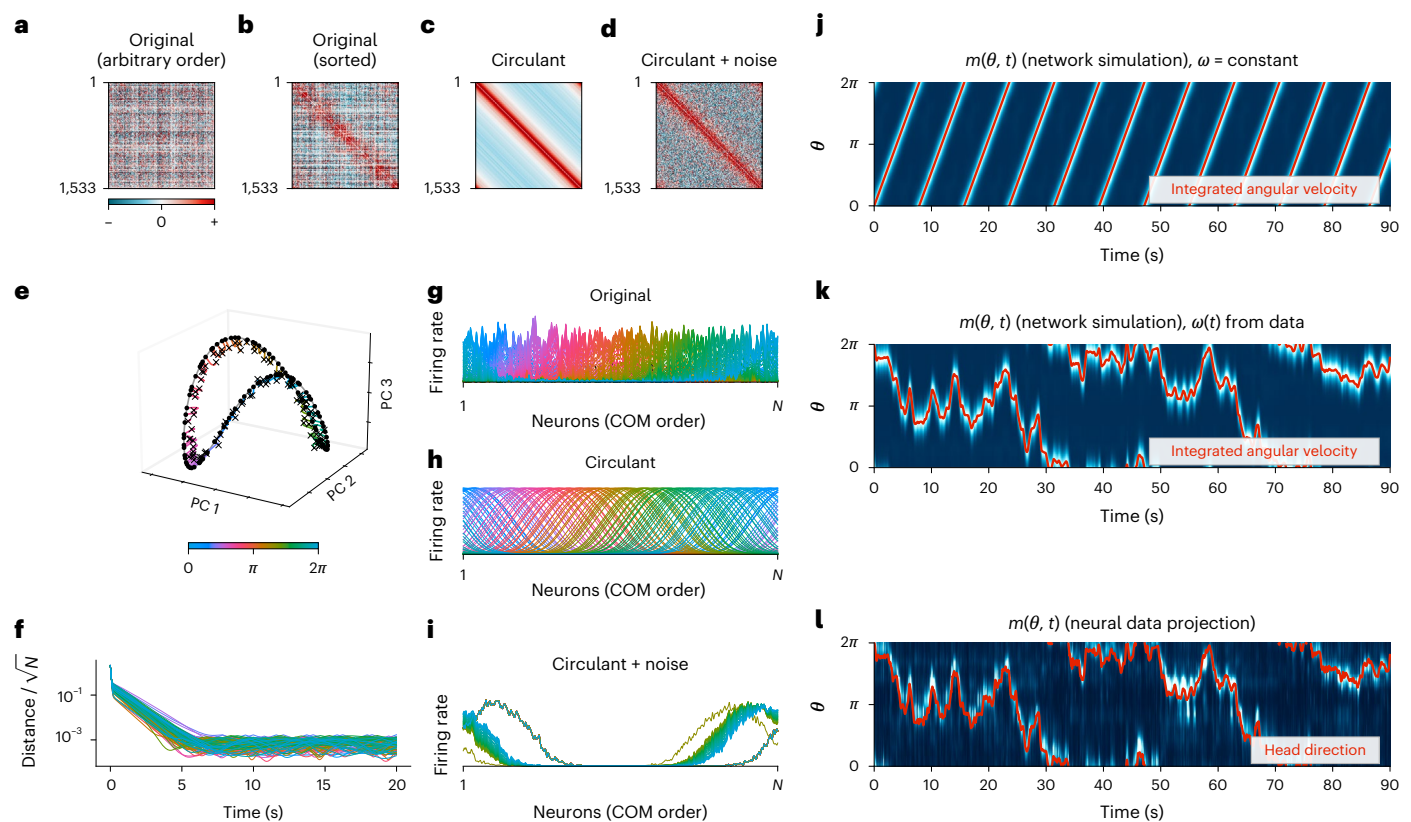


Fig. 2 | Dynamics of the data-derived network model ($N = N_{\text{data}} = 1,533$, uniform inhibition $c = 1$, softplus activation). **a**, Optimal data-derived weight matrix J , with neurons in arbitrary order. **b**, The same matrix, with neurons ordered by their tuning-curve circular centers of mass. **c**, Circulant version of **b**, obtained by averaging along diagonals. **d**, Matrix in **c** plus noise from randomly permuting the residuals between **b** and **c**. **e**, Network dynamics in the first three principal components (PC) of the target manifold. Crosses are initial conditions, which perturb the input currents off the target manifold; lines are 30-s trajectories, colored by initial-condition angle; and dots are final states. Initial conditions appear inside the manifold because the perturbation is applied to the input

currents while the projection is computed in firing-rate space, which shrinks the radial coordinate. **f**, Distance from each trajectory of **e** to the target manifold, divided by $\sqrt{N}$, over the first 20 s. **g–i**, Steady-state activity across ordered neurons at manifold coordinate θ , not to be confused with tuning curves across θ , for the weight matrices of **b** (**g**), **c** (**h**) and **d** (**i**), with neurons ordered as in **b**. COM, center of mass. **j–l**, Overlap order parameter $m(\theta, t)$ (equation (5)), with constant angular velocity input, $\omega = 0.8 \text{ rad s}^{-1}$ (**j**), angular velocity input from mouse head-direction recordings (**k**) and mouse head-direction cell recordings (**l**); $n = 117$ neurons in 100-ms spike-count bins from mouse A3707.

angular derivative of the rotation-invariant weight profile⁴. Although this nominally requires velocity-modulated weights, several biologically plausible mechanisms can mimic the effect^{4,6,8,42–44}. We generalize this construction by replacing $x_i^*(\theta)$ with $(1 + \tau\omega(t)\partial_\theta)x_i^*(\theta)$ in equation (3), which solves the same least-squares problem with a target flow of $\partial_t x_i(t) = \omega(t)\partial_\theta x_i^*(\theta)$, the tangent to the manifold, rather than a target flow of zero (Methods).

Dynamics of the data-derived network model

We now examine whether the data-derived network supports a quasicontinuous ring attractor. With neurons in arbitrary order, the optimized weight matrix appears unstructured (Fig. 2a). Ordering neurons by the circular centers of mass of their tuning curves reveals a noisy approximate Mexican hat structure (Fig. 2b); a similar ordering can be recovered from the singular vectors of the weights; Supplementary Note 2 and Supplementary Fig. 4).

We define a quasicontinuous-attractor manifold as an invariant manifold with fast, attracting normal flow and slow tangential flow, a separation of timescales. This dynamical criterion is more robust than counting fixed points, which are unreliable indicators of attractor quality⁴⁵. The optimized weights make $\tau\partial_t x_i(t)$ vanishingly small everywhere on the target manifold (Supplementary Fig. 9), but near-zero flow alone does not ensure that activity remains on the manifold or returns to it when perturbed. The uniform inhibition stabilizes the

$(N - 1)$ -dimensional hyperplane of unit mean activity, but does not stabilize the one-dimensional manifold within it.

Despite their apparent disorder, the optimal weights produce quasicontinuous-attractor dynamics. Initialized at regularly spaced manifold points perturbed by noise, trajectories return to the manifold within ~ 1 s (Fig. 2f) and then persist over 30 s (Fig. 2e), an order-of-magnitude separation between convergence and drift timescales. Furthermore, the target manifold is globally stable (Supplementary Fig. 11b).

Before turning to the nature of the disorder, we consider why stability emerges. Although not explicitly imposed, it follows from the minimum-norm prescription. The network Jacobian on the manifold is $-\delta_{ij} + J_{ij}\phi'(x_j^*(\theta))$, and because the minimum-norm weights are small, most of its eigenvalues are substantially negative, giving rapid relaxation. Non-minimum-norm constructions instead add structure in the directions the tuning curves leave unconstrained, which can impair stability (Supplementary Note 4). This stabilization therefore relies on overparameterization^{46,47}.

Are the optimal weights simply a circulant matrix plus weak noise? To test this, we constructed the circulant and circulant-plus-noise variants of the sorted weights (Supplementary Note 3) and compared the steady states reached after 30 s of evolution from perturbed manifold points. The optimal weights (Fig. 2b) maintain a continuum of stable, heterogeneous states (Fig. 2g). Averaging the sorted weights along

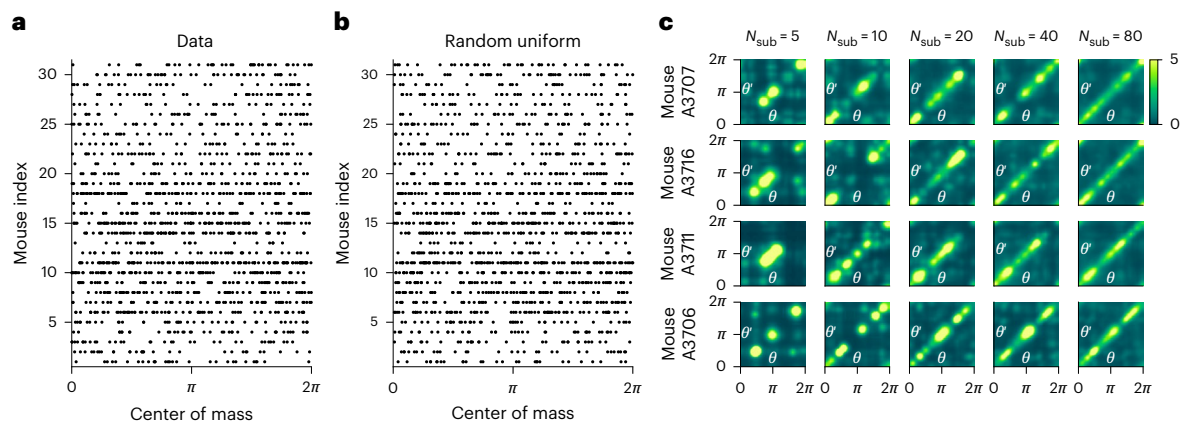


Fig. 3 | Tuning-curve center-of-mass angles and two-point correlation functions of head-direction cells ($N_{\text{data}} = 1,533$ neurons across $n = 31$ mice). **a**, Center-of-mass angles of individual tuning curves (dots). **b**, Uniformly distributed random angles, matched to the number of neurons in each session,

for visual comparison. **c**, Empirical two-point correlation functions $C^\phi(\theta, \theta')$ (equation (4)) for four example mice (rows), over random subsets of $N_{\text{sub}} = 5, 10, 20, 40$ and 80 neurons (columns).

diagonals gives a circulant matrix (Fig. 2c) whose bumps are identical up to translation, as in classical models (Fig. 2h). Adding back the scrambled residuals gives a circulant-plus-noise matrix (Fig. 2d), and this network collapses onto a few discrete states within the 30-s window (Fig. 2i), so its tangential drift occurs on the same timescale as its convergence, violating the separation-of-timescales criterion. The optimal weights are therefore not a noisy circulant matrix, and the nature of their apparent disorder is essential.

To characterize the alignment between network firing rates $\phi_i(t)$ and the target manifold $\phi_i^*(\theta)$, we define an overlap order parameter

$$m(\theta, t) = \frac{1}{N} \sum_{i=1}^N \phi_i^*(\theta) \phi_i(t). \quad (5)$$

If activity lies exactly on the manifold at coordinate ψ , then $m(\theta, t) = C^\phi(\theta, \psi)$ (equation (4)), a localized bump peaked at ψ . Under angular velocity input this bump travels, remaining centered on the integrated displacement. Classical ring-attractor models are described by an analogous order parameter with steady-state bumps. For a constant input $\omega = 0.8 \text{ rad s}^{-1}$, $m(\theta, t)$ tracks the input accurately (Fig. 2j), and the network also follows the more complex trajectories of the behavioral data (Fig. 2k).

The same order parameter can be computed from recordings, without network construction, by taking $\phi_i^*(\theta)$ to be the measured tuning curves within one mouse and $\phi_i(t)$ the binned spike counts. The resulting bump tracks the animal's head direction (Fig. 2l) and closely resembles that of the model (Fig. 2k), consistent with neural activity residing on a manifold aligned with the one defined by the tuning curves (prior work by Chaudhuri et al.¹⁵). The peak of $m(\theta, t)$ is, under certain assumptions, the maximum a posteriori estimate of head direction (Supplementary Note 3).

Statistical generative process

Single-mouse sessions contain at most $\mathcal{O}(10^2)$ neurons, and pooling across mice yields $N_{\text{data}} = \mathcal{O}(10^3)$, a small fraction of the postsubicular head-direction cell population in a single brain. We therefore develop a generative process describing the distribution of responses. Besides answering question (1), this lets us apply the optimization of equation (3) the previous section's optimization to synthetic populations of arbitrary size, enabling the large- N analysis underlying our answer to question (2).

Although tuning curves are heterogeneous, they could arise from a generative process that possesses circular symmetry. To this end, we examined the distribution of center-of-mass angles within each mouse

(Fig. 3a). The analysis must be done using a within-mouse approach because relative phases cannot be compared across animals. The distributions appear uniform (Fig. 3b), and the Kuiper test rejects uniformity in 2 of 31 sessions ($P < 0.05$) before multiple-comparison correction and in none after Benjamini–Hochberg correction, where the corrected P values across the 31 sessions range from 0.59 to 0.97.

As a stronger test, we computed the two-point correlation function of tuning curves, $C^\phi(\theta, \theta')$ (equation (4)), over increasingly large random subsets of neurons within a mouse. This function approaches a circulant form (Fig. 3c), as does the centered and variance-normalized correlation function, whose eigenvectors converge to Fourier modes (Supplementary Fig. 13), both being predictions of circular symmetry.

We developed a two-step generative process that incorporates this symmetry (Methods). First, an input-current tuning curve $x^*(\theta)$ is sampled from a circularly symmetric Gaussian process with $\langle x^*(\theta) \rangle = 0$, $\langle x^*(\theta) x^*(\theta') \rangle = \Gamma^x(\theta - \theta')$, where the correlation function depends only on the angular difference $\Delta\theta = \theta - \theta'$, as distributional circular symmetry requires. It is parameterized as a wrapped Gaussian,

$$\Gamma^x(\Delta\theta) = \sum_{p=-\infty}^{\infty} \exp\left(-\frac{\sigma^2}{2}(\Delta\theta + 2\pi p)^2\right), \quad (6)$$

where σ controls the typical number of peaks in the sampled tuning curves or, equivalently, the number of large Fourier coefficients $\hat{\Gamma}_n^x = \exp(-n^2/2\sigma^2)/\sqrt{2\pi\sigma^2}$ (Fig. 4a,b and Supplementary Note 2). Second, these input currents are transformed into firing rates by a normalized softplus function (Fig. 4c, Methods and equation (19)), $\phi^*(\theta) = \log(1 + e^{\beta(x^*(\theta) - b)})/Z[x^*]$, where β controls sharpness and b sets a soft threshold. The functional $Z[x^*]$ normalizes the θ -averaged firing rate to unity, which puts all tuning curves on a common scale and amplifies otherwise subtle tuning features (Methods). We denote by $\Gamma^\phi(\Delta\theta)$ the induced firing-rate correlation function, of which $C^\phi(\theta, \theta')$ (equation (4)) is the finite- N estimate.

Rather than fitting session-specific parameters, we sought a single-parameter set (σ, β, b) capturing the statistical structure across all recordings by minimizing the difference between the empirical two-point correlation function, pooled across mice, and that of the generative process. This yields $(\sigma, \beta, b) \approx (1.42, 2.77, 1.73)$; Fig. 4d–f) and tuning curves closely resembling the experimental ones (Fig. 4g).

We validated the process against several statistical features (Fig. 5 and Methods). The predicted center-of-mass-aligned mean and fluctuation profiles match the data (Fig. 5a,b and 'Failure of classical ring-attractor models'). The process also reproduces the distribution

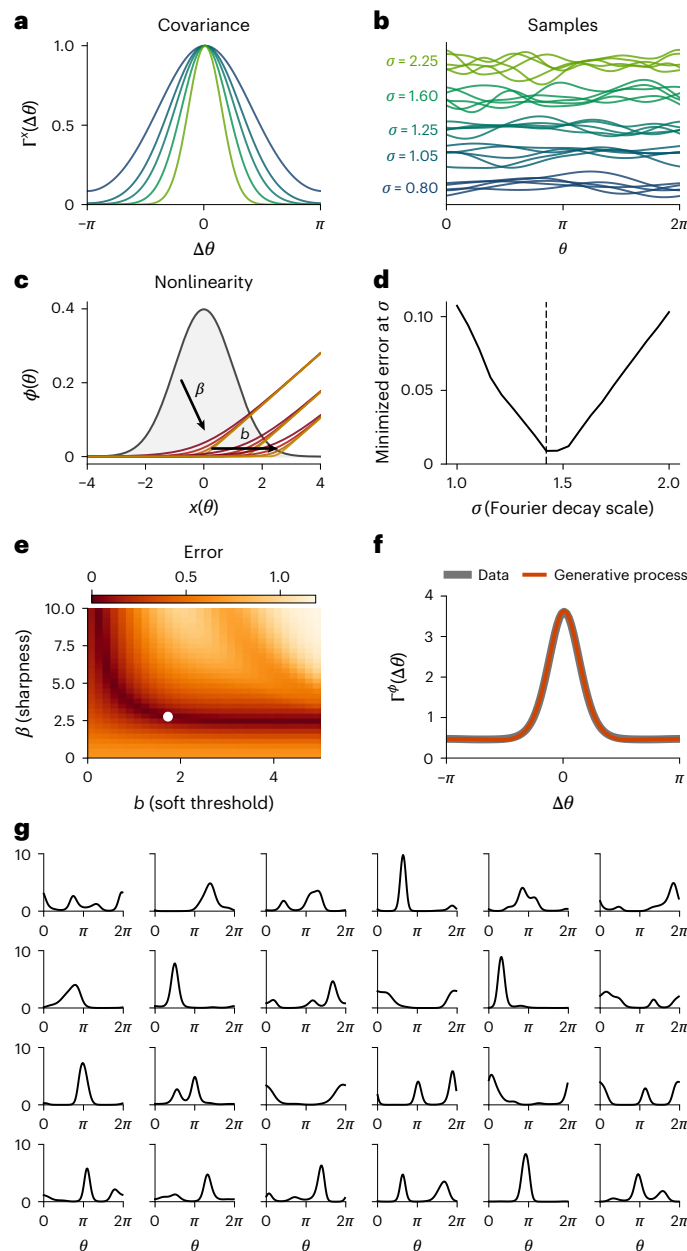


Fig. 4 | Statistical generative process for head-direction tuning curves.

a, Input-current correlation functions $\Gamma^x(\Delta\theta)$, for different values of the Fourier decay scale σ . **b**, Samples from the Gaussian process for the same values of σ as in **a**. **c**, Effect of parameters β (sharpness) and b (soft threshold) on the activation function. Gray curve shows the Gaussian distribution of input currents. **d**, Error as a function of σ , minimized over b and β , where the error is the Euclidean norm of the difference between the generative process and experimental two-point correlation functions over the first 20 distinct Fourier coefficients. The dashed line marks the minimum, $\sigma \approx 1.42$. **e**, Error as a function of b and β at $\sigma \approx 1.42$. The dot marks the minimum, $(b, \beta) \approx (1.73, 2.77)$. **f**, Two-point correlation function of the data, pooled over all $N_{\text{data}} = 1,533$ neurons from 31 mice and of the generative process at $(\sigma, \beta, b) \approx (1.42, 2.77, 1.73)$. **g**, Random samples from the generative process at optimal values.

of peak counts per tuning curve across detection thresholds z_{thresh} , with strongest agreement at higher thresholds (Fig. 5c), and the full distributions of tuning-curve values at the three highest peaks (Fig. 5d). It captures the distribution of flip-symmetry correlations ρ^{flip} (each curve's correlation with its center-of-mass reflection), particularly the strongly asymmetric tail (Fig. 5e, inset), and the distribution of head-direction information I_{HD} (ref. 48; Fig. 5f). The three-parameter

process therefore captures universal statistical structure in the rodent head-direction system.

Weights in the large-network limit

We now apply the generative process to study the optimal weight matrix J in the limit $N \rightarrow \infty$, focusing on its eigenvalues and eigenvectors (Fig. 6a). The features we identify at large N are approximately present in the finite-size data-derived weight matrix (Supplementary Fig. 14). We retain the circularly symmetric input-current correlation function $\Gamma^x(\Delta\theta)$ with Fourier decay scale σ , but adopt the saturating activation function $\phi(x) = 1 + \text{erf}(\beta\sqrt{\pi}x/2)$, which permits evaluation of certain integrals. Saturation also prevents runaway activity, so the uniform inhibition c used for the data-derived network is no longer needed, and we drop the soft threshold b and the normalization $Z[x^*]$ (Methods). This simplified process has parameters σ and β (Methods).

We organize the target input-current and firing-rate patterns $x_i^*(\theta)$ and $\phi_i^*(\theta)$ into matrices X and Φ of size $N_\theta \times N$, with θ discretized into N_θ bins. Tuning-curve smoothness allows truncation at a Fourier frequency D with $\sigma \ll D \ll N$, so both matrices have rank $2D + 1$ and admit singular value decompositions

$$X = U^x \text{diag}(\mathbf{S}^x)(V^x)^T \quad \text{and} \quad \Phi = U^\phi \text{diag}(\mathbf{S}^\phi)(V^\phi)^T, \quad (7)$$

in which the columns of matrices U^x, U^ϕ (size $N_\theta \times (2D + 1)$) are the orthonormal left singular vectors, the components of vectors $\mathbf{S}^x, \mathbf{S}^\phi$ (size $2D + 1$) are the singular values, and the columns of matrices V^x, V^ϕ (size $N \times (2D + 1)$) are the orthonormal right singular vectors. The structure of J follows from the fact that, as $N \rightarrow \infty$ with D fixed, these singular value decompositions converge to Fourier decompositions, a limit we derive in Supplementary Note 2. With the Fourier modes collected in the columns of the matrix F (size $N_\theta \times (2D + 1)$), the left singular vectors U^x and U^ϕ become F , while the singular values $\mathbf{S}^x$ and $\mathbf{S}^\phi$ become proportional to the square roots of the Fourier coefficients of the input-current and firing-rate correlation functions $\Gamma^x(\Delta\theta)$ and $\Gamma^\phi(\Delta\theta)$, which we collect in the vectors $\hat{\Gamma}^x$ and $\hat{\Gamma}^\phi$ (size $2D + 1$). At each non-zero frequency, the coefficients occur in equal pairs, a doublet degeneracy. The right singular vectors then specify how this Fourier representation, which defines a geometric ring, is embedded in the neuronal spaces of input currents and firing rates, with the i th row of V^x becoming proportional to the $2D + 1$ Fourier coefficients of $x_i^*(\theta)$.

The least-squares solution (equation (3)) maps the firing-rate manifold to the input-current manifold and can be written $J = X^T(\Phi\Phi^T + NN_\theta I_{N_\theta})^{-1}\Phi$, where we dropped the constraint $J_{ii} = 0$ in anticipation of the large- N limit. Replacing X and Φ with their singular value decompositions and taking $N \rightarrow \infty$, where the shared limit $U^x, U^\phi \rightarrow F$ renders the two sets of left singular vectors mutually orthonormal, gives the singular value decomposition of J (Supplementary Note 2),

$$J = V^x \text{diag}\left(\sqrt{\frac{\hat{\Gamma}^x}{\hat{\Gamma}^\phi}}\right)(V^\phi)^T. \quad (8)$$

The same limit makes the overlaps between the two sets of right singular vectors diagonal, $(V^\phi)^T V^x = \alpha \text{diag}(\sqrt{\hat{\Gamma}^x/\hat{\Gamma}^\phi})$, where α is the average slope of the activation function (Supplementary equation (69)). Consequently V^x and V^ϕ are the right and left eigenvectors of J , with eigenvalues

$$\lambda_n' = \alpha \hat{J}_n, \quad \text{where} \quad \hat{J}_n = \frac{\hat{\Gamma}_n^x}{\hat{\Gamma}_n^\phi}. \quad (9)$$

These eigenvalues are real and doubly degenerate, features inherited from $\hat{\Gamma}_n^x$ and $\hat{\Gamma}_n^\phi$, and they match numerical diagonalization at $N = 2,500$ (Fig. 6c,d).

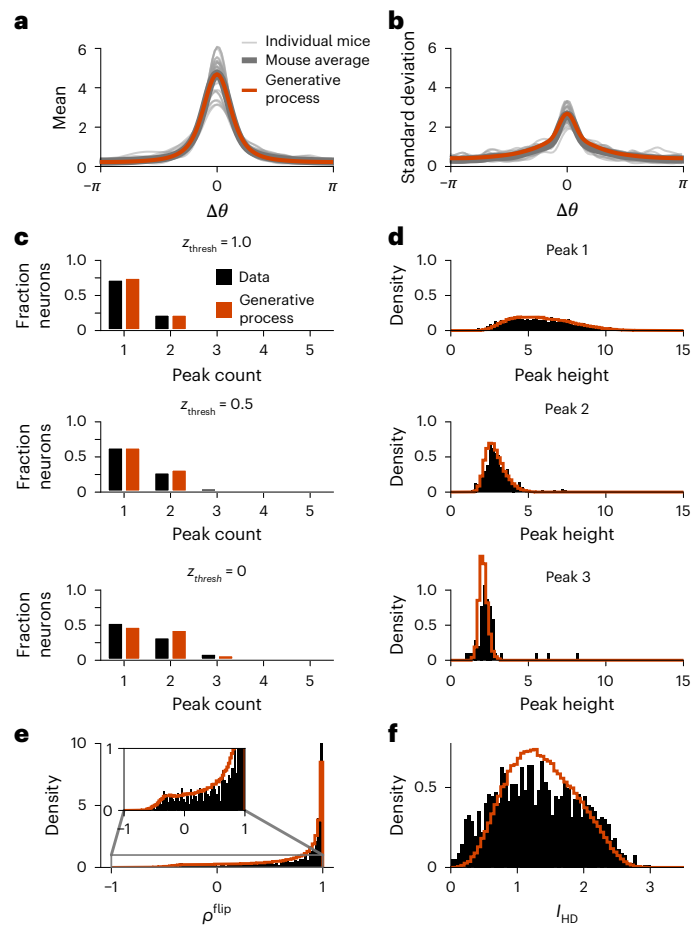


Fig. 5 | Tuning-curve statistics of the data and the generative process.

a, b, Mean (**a**) and standard deviation (**b**) of center-of-mass-aligned tuning curves predicted by the generative process, overlaid on the experimental curves of Fig. 1f, g. Individual mouse curves were averaged across one mouse's neurons, whereas mouse average curves were averaged over $n = 31$ mice. **c**, Distribution of peak counts per tuning curve at three peak-detection thresholds $z_{\text{thresh}} \in \{1, 0.5, 0\}$ (Methods). Here, and in **e** and **f**, $n = 1,533$ experimental neurons and $n = 100,000$ generative-process samples. **d**, Distributions of tuning-curve values at the three highest local maxima ($z_{\text{thresh}} = 1$), using the 1,533, 412 and 55 experimental curves and the 100,000, 24,271 and 1,043 generative-process samples that have at least 1, 2 and 3 peaks. **e**, Distribution of flip-symmetry correlations ρ^{flip} , each curve's correlation with its reflection about its own center of mass. Inset: an expanded view of the small- ρ^{flip} tail. **f**, Distribution of head-direction information I_{HD} (Methods).

To relate J to the weight matrix of a classical ring attractor, we interpret its eigenvalues and eigenvectors geometrically (Supplementary Note 2). Real, doubly degenerate eigenvalues are characteristic of a transpose-symmetric circulant matrix, yet J is neither, retaining a finite amount of transpose-asymmetry even as $N \rightarrow \infty$ (Supplementary Note 2). The difference from a circulant matrix must therefore lie in the eigenvectors, which are Fourier modes in the circulant case but disordered in J . These degenerate eigenvalues reflect the circular geometry of the attractor manifold⁴⁹, shared between our model and classical ones, while the eigenvectors reflect how that manifold is embedded in firing-rate space, through Fourier embeddings that give tuning curves identical up to shifts in classical models and disordered embeddings that give heterogeneous tuning curves in ours (Fig. 6h).

What is meant by 'disorder'? The elements of V^x are independently and identically Gaussian distributed across both rows and columns, because its rows hold the Fourier coefficients of independently sampled tuning curves and the Gaussian process leaves distinct Fourier

coefficients independent and Gaussian by definition. The rows of V^y are likewise independent, but its columns are not, because the activation function applied in θ -space induces subtle higher-order correlations among Fourier coefficients (Supplementary Note 1).

This eigenvector disorder undermines a tool for connectome analysis, spectral embedding, in which one computes the top eigenvectors of a related matrix, assigns each neuron a coordinate from its loadings on these eigenvectors and inspects the resulting point cloud for geometric structure. This procedure reveals rings in circulant matrices, whose eigenvectors are Fourier modes, but not in our model, where the point clouds from V^x and V^y are structureless (Fig. 6f and Supplementary Fig. 3). The higher-order structure in V^y does not help, because its scatter plots are no more ordered than those of V^x . Sorting neurons by an angle read from the leading two columns of V^x can recover a noisy circulant structure, but this procedure is not discriminative of ring-attractor organization, because generic low-rank matrices yield the same structure (Supplementary Note 2). Indeed, the classical intuition that each neuron should be assigned an angular coordinate is fundamentally wrong in our model, because each neuron carries not one angle but D angles (Supplementary Note 2).

An alternative model is a circulant matrix added to a background of independent random weights (Fig. 6b), a structure we found to be dynamically nonviable in the context of the data-derived network ('Dynamics of the data-derived network model' and Fig. 2d, i). Noise of order $1/\sqrt{N}$ adds a bulk to the eigenvalue spectrum⁵⁰ (Fig. 6e), but unless it is strong, the Fourier eigenvectors survive intact (Fig. 6g), as they do in models where the low-rank component is correlated, through learning, with the background, to obtain reliable continuous-attractor dynamics²⁸. In our model, by contrast, the eigenvectors are disordered by construction. Eigenvalue degeneracy, however, is present in every one of these cases. We therefore propose that eigenvalue degeneracy is a robust signature of ring-attractor structure in weight matrices, such as those from future connectome data.

Dynamical mean-field theory

The previous section identified connectivity features conserved between classical models and disordered models but did not relate their dynamics. We now describe the dynamics as $N \rightarrow \infty$ using dynamical mean-field theory, answering question (2) (Supplementary Note 2). These effective dynamics exhibit the ingredients of classical ring-attractor models, namely a nonlinear activation function (though modulated by global activity, as we explain), Mexican hat interactions and continuous symmetry. The dynamical mechanism, spontaneous breaking of continuous symmetry leading to bump states, is therefore conserved. We retain the generative process with the error-function activation function (Methods). Convergence to the mean-field behavior is shown in Supplementary Fig. 12.

The theory describes the high-dimensional dynamics in terms of a low-dimensional order parameter $\tilde{m}(\theta, t)$, a temporally low-pass-filtered version of $m(\theta, t)$ (equation (5)). For heading movements with small angular acceleration ($\tau^2 |\partial_t \omega(t)| \ll 1$, satisfied under the behavioral conditions), this order parameter obeys

$$\tau \partial_t \tilde{m}(\theta, t) = -\tilde{m}(\theta, t) + G_Q(t) \left(\frac{1}{2\pi} \int_0^{2\pi} d\theta' [1 - \tau \omega(t) \partial_\theta] \mathcal{J}(\theta - \theta') \tilde{m}(\theta', t) \right), \quad (10)$$

with interactions $\mathcal{J}(\Delta\theta)$ and effective activation function $G_Q(h)$. This equation is identical in form to that of the classical ring-attractor model of Ben-Yishai et al.²

The interactions $\mathcal{J}(\Delta\theta)$ are circulant and transpose-symmetric ($\mathcal{J}(\Delta\theta) = \mathcal{J}(-\Delta\theta)$), with Fourier coefficients $\hat{\mathcal{J}}_n = \hat{\Gamma}_n^x / \hat{\Gamma}_n^y$ proportional to the eigenvalues of J . $\mathcal{J}(\Delta\theta)$ has the Mexican hat profile characteristic of pattern-forming systems, with shape set by the Fourier decay scale

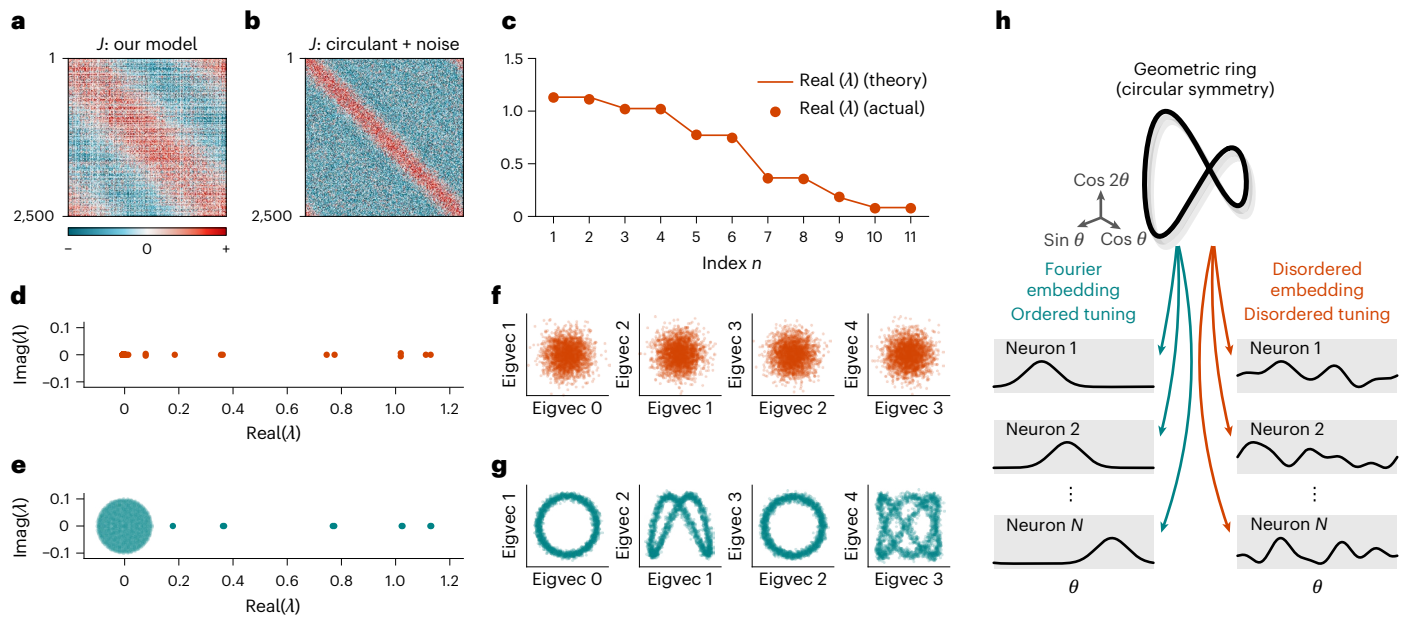


Fig. 6 | Eigenvalues and eigenvectors of the optimal weight matrix ($N=2,500$, generative-process Fourier decay scale $\sigma=1.42$, sharpness $\beta=2.77$, error-function activation). **a, Optimal weight matrix J for tuning curves drawn from the generative process; neurons were ordered by their tuning-curve circular centers of mass. **b**, Circulant matrix with the same eigenvalues as J at large N (equation (9)) and elements of order $1/N$, plus independent Gaussian noise of standard deviation $0.1/\sqrt{N}$. **c**, Sorted real parts of the leading eigenvalues of the matrix of **a** (markers), with the large- N prediction of equation (9) (line). The ninth**

is the constant mode, which does not occur in a doublet. **d,e**, Eigenvalues in the complex plane for the matrix of **a** (**d**), whose leading real parts are those in **c**, and for the circulant-plus-noise matrix of **b** (**e**). **f,g**, Scatter plots of the components of consecutive eigenvector (Eigvec) pairs of J , equivalently the left singular vectors of J , for the matrix of **a** (**f**) and for the circulant-plus-noise matrix of **b** (**g**). **h**, Schematic of the Fourier and disordered embeddings of the geometric ring in firing-rate space.

σ and sharpness β (Fig. 7a). This profile emerges from the statistical structure of the tuning curves (Supplementary Note 2 and Supplementary Fig. 5).

The effective activation function $G_Q(h)$ (Supplementary equation (90) and Fig. 7b) is modulated by the activity-dependent gain Q , a scalar that measures the overall excitation across θ . At time t , this scalar is

$$Q(t) = \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' \mathcal{U}_{\omega(t)}(\theta - \theta') \tilde{m}(\theta, t) \tilde{m}(\theta', t), \quad (11)$$

where $\mathcal{U}_{\omega(t)}(\Delta\theta)$ is an angular velocity-dependent kernel (Supplementary Note 2). Like the single-neuron activation function $\phi(x)$, $G_Q(h)$ has an input-dependent local gain, but it also carries a global gain: increasing Q lowers the slope of $G_Q(h)$ at every h (Fig. 7b). At fixed Q , this slope rises with $|h|$, rather than falling as the slope of $\phi(x)$ does under saturation. This activity-dependent gain control compensates for the absence of local saturation and is required for stability, as shown below.

For $\omega(t) = 0$ the equations admit static solutions $\tilde{m}(\theta) = m(\theta)$, of which the stable ones are bump states $m(\theta) = \Gamma^\psi(\theta - \psi)$, where ψ , the angular location of the peak, is arbitrary, reflecting the circular symmetry of the equations. For $\omega(t) \neq 0$, there are traveling-bump solutions,

$$m(\theta, t) = \Gamma^\psi \left(\theta - \psi - \int^t dt' \omega(t') \right), \quad (12)$$

and $\tilde{m}(\theta, t)$ is a temporally smoothed version of this profile. These solutions track head direction by integrating angular velocity (Fig. 7c), matching both the network simulations (Fig. 2j,k) and the recordings (Fig. 2l). For both static and traveling bumps, $Q = \Gamma^x(0)$, independent of time.

Stability is determined by the Jacobian $M(\theta, \theta') = \partial_{\tilde{m}(\theta', t)} [\tau \partial_t \tilde{m}(\theta, t)]$ (Supplementary Note 2). An eigenvalue with positive real part signals instability. At a bump state, all eigenvalues of $M(\theta, \theta')$ have negative real parts except the symmetry-induced zero (Fig. 7f, teal curve), so

perturbations relax onto the manifold with timescale $t_* = -\tau/\lambda_2^M$, set by the eigenvalue λ_2^M of second-largest real part, which is itself real; for $\beta \gtrsim 1$, this timescale is a few τ , decreasing with β and relatively insensitive to σ (Fig. 7d). At the uniform state $\tilde{m}(\theta) = \tilde{m}_0$, the Jacobian instead has a real, positive eigenvalue λ_1^M for all σ and β (a Turing instability) so perturbations grow with timescale $t_* = \tau/\lambda_1^M$, which, like the previous timescale, is a few τ for $\beta \gtrsim 1$, decreasing with β and insensitive to σ (Fig. 7e). As noted above, activity-dependent gain control is essential to this pattern formation. Fixing Q at $\Gamma^x(0)$ for all time, a manipulation with no analog in the high-dimensional network, destabilizes the bump, producing a positive eigenvalue (Fig. 7f, orange) and driving the system to a uniform state.

Grid cells

Our approach generalizes to higher-dimensional symmetries. Here, we provide a proof-of-concept application to grid cells in the medial entorhinal cortex, thought to constitute a two-dimensional toroidal attractor¹⁹.

Classical grid-cell models⁴² use a two-dimensional sheet of neurons with translation-invariant weights, which forms a hexagonally periodic pattern of activity bumps. The pattern shifts as the animal moves through the environment. Single-bump variants on a twisted torus also exist⁵¹. These models produce pristine, hexagonally periodic patterns. By contrast, medial entorhinal grid cells⁵² show a range of responses (Fig. 8a). Grid cells are typically identified by a large ‘grid score’ quantifying hexagonal periodicity, yet experimental grid scores form a continuous distribution with no clear boundary between grid and non-grid responses (Fig. 8b).

To model grid cells, we extended the Gaussian-process approach to a two-dimensional torus, parameterized by a pair of angles, with a correlation function periodic on a hexagonal lattice when viewed in cartesian coordinates (Supplementary Note 2). The pattern repeats with period $2\pi/\nu$, where the integer ν is the number of periods spanning the 2π range of each angle, and a parameter $\eta \in [0, 1]$ sets how strongly

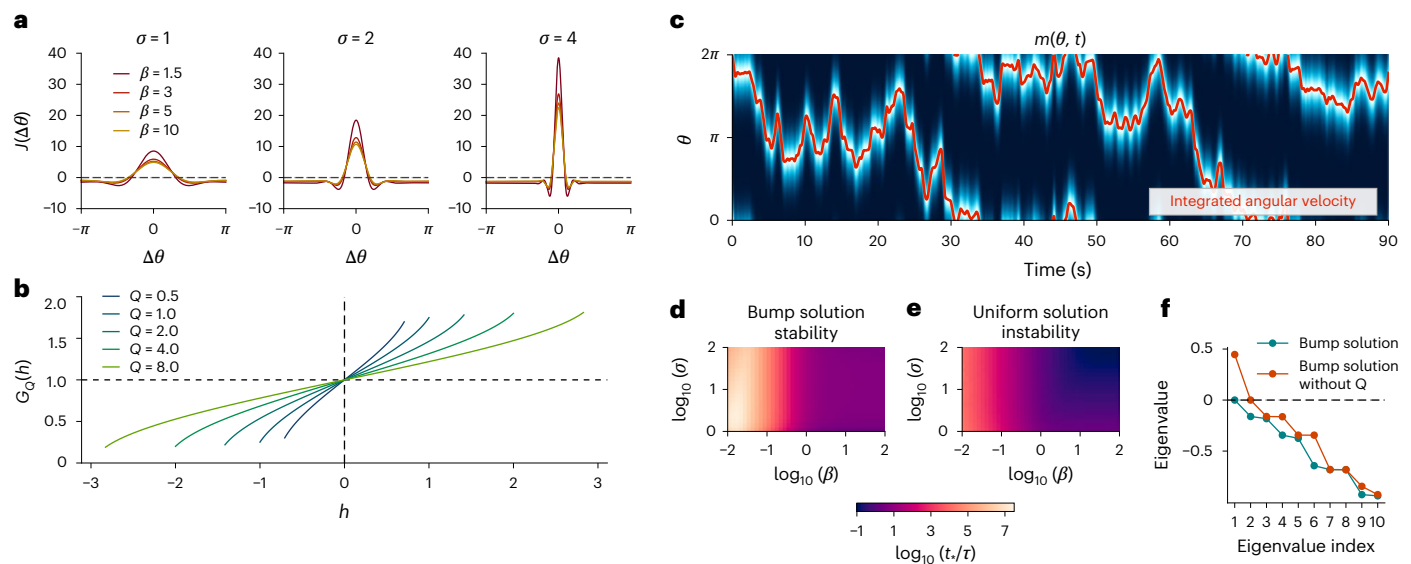


Fig. 7 | Dynamical mean-field description of the network at large N (error-function activation). a, Effective interactions $J(\Delta\theta)$, which are circulant with a Mexican hat profile, across Fourier decay scale σ (panels) and sharpness β (curves). Higher β reduces the amplitude of both excitation and inhibition, whereas larger σ narrows the profile and increases its amplitude. **b**, Effective activation function $G_Q(h)$, across values of the global gain Q (curves) with $\sigma = 1.42$ and $\beta = 2.77$. The function has a finite domain in h (Supplementary Note 2). **c**, Solution of the mean-field equations under angular velocity input taken from

mouse behavioral data, as in Fig. 2k, with $\sigma = 1.42$ and $\beta = 2.77$. **d, e**, Timescales t_* , plotted as $\log_{10}(t_*/\tau)$. **d**, Convergence to the attractor manifold, $t_* = -\tau/\lambda_2^M$, with λ_2^M the eigenvalue of the second-largest real part of the Jacobian $M(\theta, \theta')$ at the bump state. **e**, Divergence from the uniform state, $t_* = \tau/\lambda_1^M$, with λ_1^M the leading eigenvalue of $M(\theta, \theta')$ there. **f**, Real parts of the leading eigenvalues of $M(\theta, \theta')$ at the bump state, for the full mean-field equations and for modified equations with Q fixed at $\Gamma^*(0)$, with $\sigma = 1.42$ and $\beta = 2.77$.

periodic the samples are (Fig. 8d,e). At $\eta = 1$ they are exactly periodic with period $2\pi/v$; for $\eta < 1$, they are only approximately so but remain exactly periodic over 2π , preserving the manifold's toroidal symmetry. We set $v = 3$ for the sampled maps and grid-score analyses. The sampled maps produce the continuous spread of grid scores, as seen in the data (Fig. 8b,c). The mean grid score rises monotonically with η .

The two-dimensional autocorrelograms of the sampled input-current maps are typically more hexagonal than the maps themselves (Fig. 8d,e), as has been recognized in real grid-cell recordings since their discovery¹³. This is extremely natural within our approach: the grid pattern resides in the correlation function of the generative process, not in any individual neuron, and a pristine grid cell is simply one whose firing-rate map happens to resemble that correlation function. Such resemblance is not essential, because cells across the full range of grid scores contribute to the attractor we construct. The prevailing grid-score criterion may therefore be overly restrictive.

We constructed attractor networks from samples drawn with $v = 1$, a case analogous to the head-direction model, using least-squares optimization. We used $N = 25,000$, an order of magnitude more neurons than in the head-direction model, because the number of neurons needed for stable dynamics scales exponentially with manifold dimension. The network converges to the toroidal manifold from nearby initial conditions (Fig. 8f). The weight matrix at large N exhibits eigenvalue degeneracies in multiples of six, with the first few being sixfold, reflecting the hexagonal symmetry of the lattice (Supplementary Note 2 and Supplementary Fig. 15), analogous to the doublet degeneracy of head-direction networks. The same degeneracies appear for $v > 1$.

When written in cartesian coordinates, the full $N \rightarrow \infty$ mean-field description with velocity inputs takes the form of classical grid-cell models^{42,51}, with two-dimensional translation symmetry (Supplementary equation (136) and Supplementary Note 2). With $v = 1$, it is a single-bump twisted-torus model, in which the mean-field activity is a single localized region moving in response to velocity input.

Discussion

Mammalian circuits show heterogeneous tuning at odds with continuous-attractor models, which rely on continuously symmetric weights and generate tuning curves identical up to shifts. We showed that this heterogeneity is nonetheless consistent with an underlying circular symmetry, that the corresponding tuning curves form stable attractor manifolds whose minimum-norm weights show the same eigenvalue degeneracies as classical circulant weights, but with disordered eigenvectors, and that the resulting networks admit a dynamical mean-field theory displaying the essential features of classical ring-attractor models. Classical continuous-attractor dynamics could therefore be realized in the mammalian brain, exactly so at large N .

The steps we followed for head-direction and grid cells suggest a general framework for modeling how ordered computation emerges within large, disordered circuits and for connecting such models to data. This framework applies wherever Gaussian-process tuning models are apt, as they have proven to be for place cells²⁷ and, when defined over time and task parameters, for movement-related trajectories in motor cortex³³.

Is symmetry necessary?

Our model's correspondence with classical models rests on rotational symmetry of tuning-curve statistics, raising the question of whether symmetry is necessary for continuous-attractor dynamics. Strictly speaking, it is not, because universal approximation³³ implies that, for sufficiently large N , recurrent networks can support, to arbitrary accuracy, a continuous-attractor manifold of arbitrary, potentially nonsymmetric shape embedded in a fixed-dimensional subspace (Supplementary Note 1).

What is the nature of the effective dynamics in such nonsymmetric systems? Our framework extends to asymmetric cases, where a continuum of stable states arises without continuous symmetry (Supplementary Note 2 and Supplementary Fig. 6). There, both the interaction and gain kernels become noncirculant, and the input-current

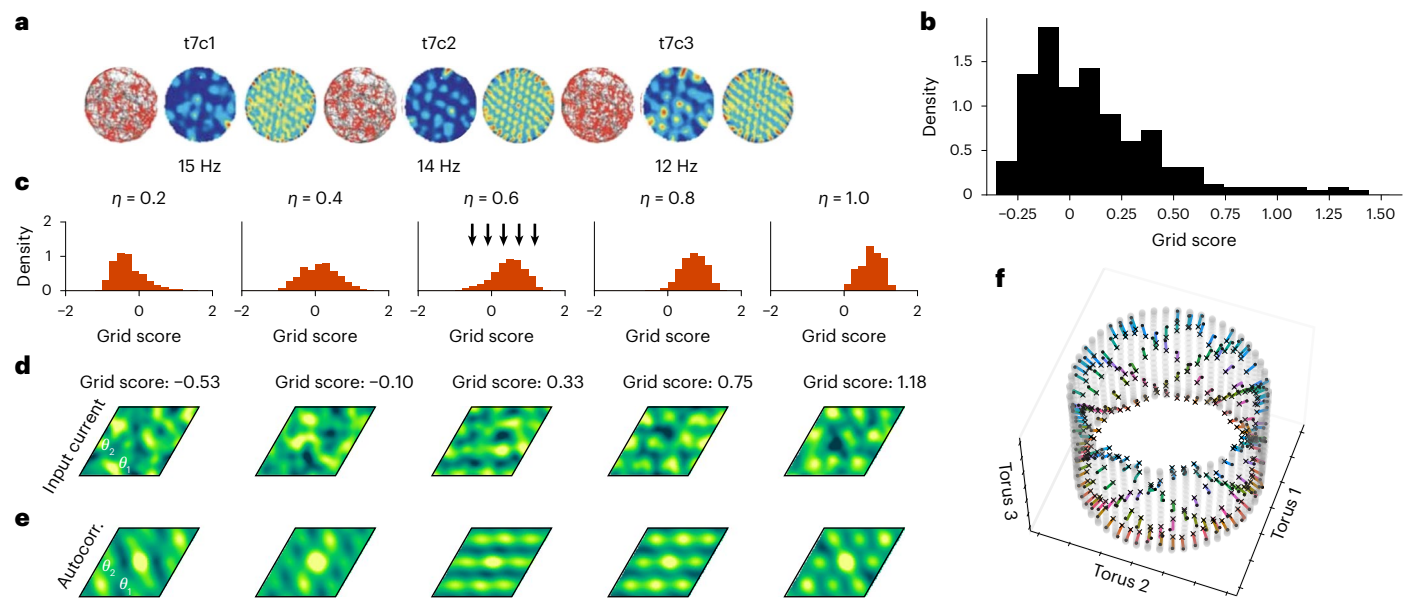


Fig. 8 | Grid-cell model derived from heterogeneous tuning curves.

a, Experimental grid-cell responses, adapted with permission from Hafting et al.¹³. For each cell, the rat's two-dimensional trajectory with spike locations overlaid (left), the firing-rate map (middle) and the spatial autocorrelogram (right), with the peak firing rate below, are shown. **b**, Distribution of grid scores from experimental recordings ($n = 620$ medial entorhinal units), digitized from Nayeibi et al.²³, which analyzed data from Mallory et al.³². **c**, Grid-score distributions of input currents sampled from the Gaussian process, across the periodicity strength η (panels; generative-process Fourier decay scale $\sigma = 12.12$,

$v = 3$ periods per 2π , $n = 1,000$ samples per panel). **d**, Input-current maps at five grid scores evenly spaced across the $\eta = 0.6$ distribution of **c**, indicated by arrows, over the torus. **e**, Spatial autocorrelograms (autocorr.) of the input-current maps in **d**. **f**, Network dynamics ($N = 25,000$, $\sigma = 4.04$, activation sharpness $\beta = 2.5$, $v = 1$, error-function activation) projected onto three orthonormal directions within the space of the six near-degenerate leading principal components, chosen so that one toroidal phase closes a loop in the plotted plane. Markers and lines are the same as in Fig. 2e, with color indicating initial condition.

variance depends on position along the manifold. Together these changes compensate for the asymmetric correlation functions, and the attractor survives.

Why, then, might biological systems exhibit symmetry? Head-direction circuits appear rigid and thus may reflect lifetime-averaged environmental statistics, which are angularly isotropic. This isotropy is broken within the particular environments an animal explores over minutes to hours, raising the experimentally testable question of whether head-direction systems show correspondingly short-term asymmetries. Symmetry may also be easier to develop, because existing models of ring-attractor self-organization produce symmetric weights³⁴.

Unknown variables

Our approach assumes that the encoded variable is known, using measured head direction to compute tuning curves. However, most variables encoded in the brain likely have no simple counterpart in the external world and are unknown to experimenters. Relatedly, our data-derived network construction uses a crude method of estimating input currents, which are not measured, namely, applying the inverse activation function to measured tuning curves ('Constructing a data-derived network model'). Due to this crudeness, the estimates it yields are not Gaussian, unlike the input currents the generative process postulates, so quantitative large- N predictions for that particular network, such as its eigenvalue spectrum, do not hold. Unsupervised methods address both of these 'unmeasured variable' limitations by inferring latent variables and tuning curves jointly from activity within a generative model, and several place a Gaussian-process prior on input currents^{55,56}, as our generative process does. Combining them with our approach would remove the need to know the encoded variable and would permit fitting the activation and correlation functions by maximum likelihood.

From world structure to circuit structure

What can one infer about neural circuits from the structure of the world? Near the sensory periphery, such inferences have repeatedly succeeded. Helmholtz predicted frequency decomposition in the cochlea from the superposition of frequencies in sound⁵⁷; Hering inferred color opponency, realized in retinal ganglion cells, through perceptual observations⁵⁸; Olshausen and Field⁵⁹ showed that natural image statistics yield V1-like receptive fields under a sparse, overcomplete code; and the many-receptor, multidimensional organization found by Buck and Axel⁶⁰ resembles compressed sensing^{61,62} (the latter two examples being postdictions that could in principle have been predictions).

Deeper in the brain, such inferences are harder, but our approach suggests a route for continuous variables. Begin with a variable in the world and assume a neural circuit represents it by generating a persistent activity manifold, which must share the variable's topology. The manifold's geometry follows if environmental symmetries over an animal's lifetime lead the distribution of tuning curves to obey the variable's symmetry, a far weaker requirement than identical tuning-curve shapes. At large N , such distributional symmetry becomes a geometric symmetry of the manifold. The weights must encode the manifold to generate it, and, as we have shown, its geometric symmetry appears in them as eigenvalue degeneracies. On general grounds, then, symmetries of the world should be reflected in the spectral structure of connectivity, even far from the periphery.

Applying this observation to connectome data comes with caveats, however. The eigenvalue degeneracy is suggestive of, but not sufficient for, continuous-attractor organization (Supplementary Note 1) and can arise for unrelated reasons. For example, a Hopfield network storing a fixed number of patterns acquires a degeneracy equal to that number, though lacking the multilevel structure associated with a reducible representation of a continuous symmetry group (see below). It is also only approximate on finite data. In our data-derived

weight matrix, the eigenvalues are approximately real, and the doublets are visible but not exact (Supplementary Fig. 14). Finally, extending eigenvalue-based analyses to circuits with distinct cell types, such as separate excitatory and inhibitory populations, is an important direction for methods development.

Low-rank networks, symmetry and degeneracy

The optimal weights (equation (3)) are approximately low rank, having far fewer large eigenvalues than neurons. Low-rank recurrent networks admit low-dimensional descriptions of their dynamics^{31,32,63,64}, a property we exploited. This setting is a natural one in which to formalize a broader theme of this paper, namely how continuous symmetries, dynamics and eigenvalue degeneracies relate. In Supplementary Note 1, we show that symmetries in the distribution of the weights generate corresponding symmetries in the dynamics (see also Pezon et al.⁶⁵, who establish, among other results, a closely related one for networks whose weights are built from a feature-space kernel). Our model instantiates this principle, because circular symmetry in the connectivity distribution produces circular symmetry in the overlap dynamics. Such symmetries in turn appear in the weight-matrix spectrum as eigenvalue degeneracies, with multiplicities given by the dimensions of the irreducible representations of the symmetry group, so that the ring's circle group yields doublets. The converse does not hold: eigenvalue degeneracies are highly suggestive of geometric structure, but unlike the distributional symmetries that produce them, they depend only on the second-order statistics of the loadings defining the low-rank weights and thus do not by themselves fix the symmetry group. In particular, we can construct ensembles that match our model's second-order statistics and hence its spectrum, while differing beyond second order, yielding different symmetry groups and attractor structures (Supplementary Figs. 1 and 2).

Our construction could be generalized beyond the circle and torus to other compact Lie groups, including non-abelian ones, which Di Bernardo et al.⁶⁶ study in the nondisordered setting. Heterogeneously tuned attractors built on these symmetries are a natural next step.

How should one model the connectivity that underlies a task and the dynamics it supports? One route has been to train a recurrent network via backpropagation and subsequently approximate the connectivity with an analytically tractable low-rank ensemble^{63,67}. In this paper, we impose, rather than learn, the target tuning curves, which yields minimum-norm weights in closed form and allows us to link tuning statistics, weights and mean-field dynamics analytically. A more ambitious goal, on which work has begun^{68,69}, is an end-to-end theory connecting a task to the connectivity that learning induces and in turn to the resulting dynamics. These routes offer different insights and are complementary.

Order and disorder

We have shown that continuous attractors admit a broad class of dynamically equivalent implementations, within which highly ordered classical models form a measure-zero special case ('Weights in the large-network limit' and Supplementary Note 2). Heterogeneous tuning may therefore be prevalent in mammalian brains simply because it is generic. What then favors ordered architectures? Robustness appears not to, because our disordered networks degrade under weight perturbations much like classical circulant models (Supplementary Note 3 and Supplementary Fig. 10). Limited neuronal resources are a stronger candidate, because the disordered models we have proposed require large N . This would explain why flies, with far fewer neurons, implement an anatomical ring attractor closely resembling classical models⁷⁰.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-026-02452-4>.

References

1. Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* **23**, 744–766 (2022).
2. Ben-Yishai, R., Bar-Or, R. L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
3. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
4. Ocko, S. A., Hardcastle, K., Giocomo, L. M. & Ganguli, S. Emergent elasticity in the neural code for space. *Proc. Natl Acad. Sci. USA* **115**, E11798–E11806 (2018).
5. Barak, O. & Romani, S. Mapping low-dimensional dynamics to high-dimensional neural activity: a derivation of the ring model from the neural engineering framework. *Neural Comput.* **33**, 827–852 (2021).
6. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).
7. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).
8. Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).
9. Hulse, B. K. et al. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).
10. Petrucco, L. et al. Neural dynamics and architecture of the heading direction circuit in zebrafish. *Nat. Neurosci.* **26**, 765–773 (2023).
11. Ranck, J. B. Head direction cells in the deep cell layer of dorsolateral pre-subiculum in freely moving rats. in *Electrical Activity of the Archicortex* (eds Buzsáki, G. & Vanderwolf, C. H.) 217–220 (Akadémiai Kiadó, 1985).
12. Taube, J. S., Muller, R. U. & Ranck, J. B. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).
13. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
14. Moser, E. I., Kropff, E. & Moser, M.-B. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* **31**, 69–89 (2008).
15. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, 1512–1520 (2019).
16. Duzkiewicz, A. J. et al. Local origin of excitatory–inhibitory tuning equivalence in a cortical network. *Nat. Neurosci.* **27**, 782–792 (2024).
17. Yoon, K. et al. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.* **16**, 1077–1084 (2013).
18. Gardner, R. J., Lu, L., Wernle, T., Moser, M.-B. & Moser, E. I. Correlation structure of grid cells is preserved during sleep. *Nat. Neurosci.* **22**, 598–608 (2019).
19. Gardner, R. J. et al. Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128 (2022).
20. Trettel, S. G., Trimper, J. B., Hwaun, E., Fiete, I. R. & Colgin, L. L. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nat. Neurosci.* **22**, 609–617 (2019).

21. Hardcastle, K., Maheswaranathan, N., Ganguli, S. & Giocomo, L. M. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron* **94**, 375–387 (2017).
22. Gu, Y. et al. A map-like micro-organization of grid cells in the medial entorhinal cortex. *Cell* **175**, 736–750 (2018).
23. Nayeibi, A. et al. Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks. In *Proc. 35th Conference on Neural Information Processing Systems (NeurIPS 2021)* (eds Ranzato, M. et al.) 12167–12179 (Curran Associates, 2021).
24. Posani, L., Wang, S., Muscinelli, S. P., Paninski, L. & Fusi, S. Rarely categorical, highly separable representations along the cortical hierarchy. *Nature* <https://doi.org/10.1038/s41586-026-10668-4> (2026).
25. Abbott, L. F. et al. The mind of a mouse. *Cell* **182**, 1372–1376 (2020).
26. Kuan, A. T. et al. Synaptic wiring motifs in posterior parietal cortex support decision-making. *Nature* **627**, 367–373 (2024).
27. Mainali, N., da Silveira, R. A. & Burak, Y. Universal statistics of hippocampal place fields across species and dimensionalities. *Neuron* **113**, 1110–1120 (2025).
28. Darshan, R. & Rivkind, A. Learning to represent continuous variables in heterogeneous neural networks. *Cell Rep.* **39**, 110612 (2022).
29. Rivkind, A. & Barak, O. Local dynamics in trained recurrent neural networks. *Phys. Rev. Lett.* **118**, 258101 (2017).
30. Pollock, E. & Jazayeri, M. Engineering recurrent neural networks from task-relevant manifolds and dynamics. *PLoS Comput. Biol.* **16**, e1008128 (2020).
31. Mastrogiuseppe, F. & Ostojic, S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**, 609–623 (2018).
32. Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F. & Ostojic, S. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Comput.* **33**, 1572–1615 (2021).
33. Doya, K. *Universality of Fully Connected Recurrent Neural Networks Technical Report* (Department of Biology, Univ. of California, 1993).
34. Viejo, G., Skromne Carrasco, S. & Peyrache, A. The head-direction signal is generated by multiple attractor-like networks. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.09.10.675474> (2025).
35. Taube, J. S. The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.* **30**, 181–207 (2007).
36. Andalman, A. S. et al. Neuronal dynamics regulating brain and behavioral state transitions. *Cell* **177**, 970–985 (2019).
37. Fisher, D., Olasagasti, I., Tank, D. W., Aksay, E. R. & Goldman, M. S. A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. *Neuron* **79**, 987–1000 (2013).
38. Rajan, K., Harvey, C. D. & Tank, D. W. Recurrent network models of sequence generation and memory. *Neuron* **90**, 128–142 (2016).
39. Duszkiwicz, A., Skromne Carrasco, S. & Peyrache, A. Large-scale recordings of head-direction cells in mouse postsubiculum. *DANDI Archive* <https://doi.org/10.48324/dandi.000939/O.240528.1542> (2024).
40. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
41. Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl Acad. Sci. USA* **81**, 3088–3092 (1984).
42. Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).
43. Turner-Evans, D. et al. Angular velocity integration in a fly heading circuit. *eLife* **6**, e23496 (2017).
44. Sorscher, B., Mel, G. C., Ocko, S. A., Giocomo, L. M. & Ganguli, S. A unified theory for the computational and mechanistic origins of grid cells. *Neuron* **111**, 121–137 (2023).
45. Ságodi, Á., Martín-Sánchez, G., Sokół, P. & Park, I. M. Back to the continuous attractor. In *Proc. 38th International Conference on Neural Information Processing Systems (eds Globerson, A. et al.)* 66856–66906 (Curran Associates, 2024).
46. Radhakrishnan, A., Belkin, M. & Uhler, C. Overparameterized neural networks implement associative memory. *Proc. Natl Acad. Sci. USA* **117**, 27162–27170 (2020).
47. Belkin, M., Hsu, D., Ma, S. & Mandal, S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proc. Natl Acad. Sci. USA* **116**, 15849–15854 (2019).
48. Skaggs, W., McNaughton, B., Gothard, K. & Markus, E. J. An information-theoretic approach to deciphering the hippocampal code. In *Proc. 6th International Conference on Neural Information Processing Systems (eds Hanson, S. J. et al.)* 1030–1037 (Morgan Kaufmann Publishers, 1992).
49. Kriegeskorte, N. & Wei, X.-X. Neural tuning and representational geometry. *Nat. Rev. Neurosci.* **22**, 703–718 (2021).
50. Baik, J., Ben Arous, G. & Pécché, S. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33**, 1643–1697 (2005).
51. Guanella, A., Kiper, D. & Verschure, P. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.* **17**, 231–240 (2007).
52. Mallory, C. S. et al. Mouse entorhinal cortex encodes a diverse repertoire of self-motion signals. *Nat. Commun.* **12**, 671 (2021).
53. Gao, P. et al. A theory of multineuronal dimensionality, dynamics and measurement. Preprint at *bioRxiv* <https://doi.org/10.1101/214262> (2017).
54. Vafidis, P., Oswald, D., D’Albis, T. & Kempter, R. Learning accurate path integration in ring attractor models of the head direction system. *eLife* **11**, e69841 (2022).
55. Wu, A., Pashkovski, S., Datta, S. R. & Pillow, J. W. Learning a latent manifold of odor representations from neural responses in piriform cortex. In *Proc. 32nd International Conference on Neural Information Processing Systems (eds Bengio, S. et al.)* 5383–5393 (Curran Associates, 2018).
56. Jensen, K. T., Kao, T.-C., Tripodi, M. & Hennequin, G. Manifold GPLVMs for discovering non-Euclidean latent structure in neural data. In *Proc. 34th International Conference on Neural Information Processing Systems (eds Larochelle, H. et al.)* 22580–22592 (Curran Associates, 2020).
57. Helmholtz, H. L. *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Cambridge Univ. Press, 2009).
58. Hering, E. *Outlines of a Theory of the Light Sense* (Harvard Univ. Press, 1964).
59. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
60. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
61. Grabska-Barwińska, A. et al. A probabilistic approach to demixing odors. *Nat. Neurosci.* **20**, 98–106 (2017).
62. Zavatone-Veth, J. A. et al. Neural circuits for fast Poisson compressed sensing in the olfactory bulb. In *Proc. 37th International Conference on Neural Information Processing Systems (eds Oh, A. et al.)* 64793–64828 (Curran Associates, 2023).

63. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F. & Ostojic, S. The role of population structure in computations through neural dynamics. *Nat. Neurosci.* **25**, 783–794 (2022).
 64. Clark, D. G. & Beiran, M. Structure of activity in multiregion recurrent neural networks. *Proc. Natl Acad. Sci. USA* **122**, e2404039122 (2025).
 65. Pezon, L., Schmutz, V. & Gerstner, W. Linking neural manifolds to circuit structure in recurrent networks. *Neuron* **114**, 1682–1694 (2026).
 66. Di Bernardo, A., Valente, A., Mastrogiuseppe, F. & Ostojic, S. Shaping manifolds in equivariant recurrent neural networks. Preprint at <https://doi.org/10.48550/arXiv.2511.04802> (2025).
 67. Valente, A., Pillow, J. W. & Ostojic, S. Extracting computational mechanisms from neural data using low-rank RNNs. In *Proc. 36th International Conference on Neural Information Processing Systems* (eds Koyejo, S. et al.) 24072–24086 (Curran Associates, 2022).
 68. Bordelon, B., Cotler, J., Pehlevan, C. & Zavatone-Veth, J. A. Dynamics of learning to integrate in linear recurrent neural networks. Preprint at <https://doi.org/10.48550/arXiv.2503.18754> (2025).
 69. Clark, D. G., Bordelon, B., Zavatone-Veth, J. A. & Pehlevan, C. Structure, disorder, and dynamics in task-trained recurrent neural circuits. Preprint at *bioRxiv* <https://doi.org/10.64898/2026.03.02.708943> (2026).
 70. Noorman, M., Hulse, B. K., Jayaraman, V., Romani, S. & Hermundstad, A. M. Maintaining and updating accurate internal representations of continuous variables with a handful of neurons. *Nat. Neurosci.* **27**, 2207–2217 (2024).
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.
- © The Author(s), under exclusive licence to Springer Nature America, Inc. 2026

Methods

Data processing

Tuning-curve estimation. We analyzed neural recordings from mice, collected by Duszkievicz et al.¹⁶ (see ref. 39 for the publicly available dataset), comprising 1,533 excitatory neurons that the authors identified as head-direction cells.

Neural and behavioral data were stored in Neurodata Without Borders format. For each mouse, we extracted the animal's head direction and spike times from the epoch of square-environment exploration. Epochs during which the virus-expressing mice underwent optogenetic manipulation were excluded. Head-direction angles were binned into 100 equally sized bins spanning the full 2π rad. For each neuron, we computed raw tuning curves by dividing the spike count in each angular bin by the time spent in that bin. Time bins in which head direction was not tracked were excluded from the occupancy, spike-count and likelihood computations.

To reduce noise in the resulting histogram-based tuning-curve estimates while preserving non-noise features, we implemented a cross-validated smoothing procedure. For each neuron, we performed the following:

- (1) Divided the recording session into eight temporal segments of equal duration, labeled alternately as partition 1 or 2 (that is, 1-2-1-2-1-2-1-2).
- (2) Computed raw tuning curves independently for each partition.
- (3) Applied Gaussian smoothing kernels of varying widths to the tuning curve from each partition, with widths logarithmically spaced from 0.1 to 100 bins (corresponding to $2\pi/1,000$ to 2π rad).
- (4) Evaluated the Poisson log-likelihood of the spiking data from one partition given the smoothed tuning curve from the other partition. Let $r_{i,\sigma_{\text{smooth}}}^{P1}(\theta)$ denote the tuning curve for neuron i obtained from partition 1 and smoothed with width σ_{smooth} , $s_i^{P2}(t)$ the spike count in a bin of width $\Delta t_{\text{bin}} = 10$ ms from partition 2 and $\theta(t)$ the animal's head direction. We evaluated

$$\begin{aligned} \mathcal{L}_{\text{Pois},i}^{P2|P1}(\sigma_{\text{smooth}}) \\ = \sum_{t \in P2} \left[s_i^{P2}(t) \log \left(r_{i,\sigma_{\text{smooth}}}^{P1}(\theta(t)) \Delta t_{\text{bin}} \right) - r_{i,\sigma_{\text{smooth}}}^{P1}(\theta(t)) \Delta t_{\text{bin}} - \log(s_i^{P2}(t)!) \right], \end{aligned} \quad (13)$$

with $\mathcal{L}_{\text{Pois},i}^{P1|P2}(\sigma_{\text{smooth}})$ defined symmetrically. Before smoothing, zero-rate angular bins were set to 10^{-3} times the mean rate so that the log-likelihood remained finite.

- (5) For each neuron and partition, selected the smoothing width that maximized the likelihood of the opposite partition.
- (6) Computed final tuning curves by averaging the optimally smoothed curves from both partitions.

Finally, we normalized all tuning curves to have a mean firing rate of unity (arbitrary units) across all head directions.

The weight construction ('Computing the weight matrix') uses a finer angular grid than these 100 bins, so after the smoothing and double normalization of Supplementary Note 3, we resampled the curves onto that grid by zero padding the Fourier coefficients of the input currents and reapplying the activation function to yield firing rates.

Tuning-curve statistics. We use four summary statistics of these tuning curves, all evaluated on the angular grid $\theta_a = 2\pi(a-1)/N_\theta$ for $a=1, \dots, N_\theta$.

The circular center of mass of neuron i is

$$\theta_i^{\text{COM}} = \arg \sum_{a=1}^{N_\theta} \phi_i^*(\theta_a) e^{i\theta_a}. \quad (14)$$

We use it to align tuning curves before averaging them across neurons ('Failure of classical ring-attractor models' and Fig. 1f,g) and to order neurons so that structure in the weight matrix is visible (Figs. 2b and 6a).

The flip-symmetry correlation of neuron i is

$$\rho_i^{\text{flip}} = \text{corr}_a(\phi_i^*(\theta_a), \phi_i^*(2\theta_i^{\text{COM}} - \theta_a)), \quad (15)$$

the Pearson correlation across angles between the tuning curve and its reflection about its circular center of mass.

The head-direction information of neuron i is

$$I_{\text{HD},i} = \frac{1}{N_\theta} \sum_{a=1}^{N_\theta} \phi_i^*(\theta_a) \log_2 \phi_i^*(\theta_a), \quad (16)$$

the measure of Skaggs et al.⁴⁸ under uniform occupancy. Because the curves have unit mean over θ by construction, the values $\phi_i^*(\theta_a)/N_\theta$ form a probability distribution over θ , and I_{HD} is $\log_2 N_\theta$ minus its entropy (Fig. 5f).

Finally, a local maximum of $\phi_i^*(\theta)$ counts as a peak when

$$\frac{\phi_i^*(\theta_a) - \langle \phi_i^* \rangle_\theta}{\text{SD}_\theta[\phi_i^*]} > z_{\text{thresh}}, \quad (17)$$

so lowering z_{thresh} admits progressively shallower peaks (Fig. 5c).

Model settings

The activation function, network dynamics and weight construction differ slightly across contexts. The three settings are collected here.

- Setting 1: experimental data-derived network model. For constructing a network model from experimental tuning curves ('Dynamics of the data-derived network model'), we use the softplus function

$$\phi(x) = \frac{1}{\beta} \log(1 + \exp(\beta x)), \quad (18)$$

where β controls the sharpness of the kink. We set $\beta = 2$ in all data-derived analyses. The bias does not appear in the activation function here. It enters instead as a current b in the network dynamics (same as equation (1)),

$$\tau \partial_t x_i(t) = -x_i(t) + \sum_{j=1}^N J_{ij} \phi_j(t) + b.$$

The weights are constructed via least-squares optimization ('Computing the weight matrix') to satisfy the vanishing-flow condition on the target manifold. The firing-rate target manifold is defined by the tuning curves themselves, while the input-current target manifold results from applying $\phi^{-1}(\cdot)$ to the firing-rate manifold. Thus, using an invertible activation function (that is, $\beta < \infty$) is important. After fitting the weights with $b = 0$, we introduce a parameter $c > 0$ to stabilize the mean activity mode; in particular, we shift $J_{ij} \rightarrow J_{ij} - N^{-1}c$ and set $b = c$ (Supplementary Note 3).

Setting 2: experimental data-fitted statistical generative process. For the generative process ('Statistical generative process'), we use a normalized softplus function

$$\phi^*(\theta) = \frac{1}{Z[x^*]} \log(1 + e^{\beta(x^*(\theta) - b)}), \quad (19)$$

where β controls the sharpness, b is the bias, and

$$Z[x^*] = \frac{1}{2\pi} \int_0^{2\pi} d\theta \log(1 + \exp(\beta(x^*(\theta) - b))) \quad (20)$$

is a normalization functional ensuring unit mean firing rate over θ . The normalization places all tuning curves on a common scale, both across mice and between data and generative process. It also

rescales each curve by its own mean, which matters most for the curves whose input current is mostly well below the bias, because for those the unnormalized rate is small everywhere and the tuning features are correspondingly faint; dividing by that small mean turns them into pronounced peaks. We use fitted parameters $(\sigma, \beta, b) \approx (1.42, 2.77, 1.73)$, obtained as described in ‘Statistical generative process’, where σ is the Fourier decay scale of the input-current correlation function $\Gamma^x(\Delta\theta)$ and does not appear in the activation function itself. This activation function is used for statistical modeling of tuning curves; no weight matrix is constructed here.

- Setting 3: analytically tractable statistical generative process and network model. For large- N theoretical analysis of the weights and dynamics (‘Weights in the large-network limit’ and ‘Dynamical mean-field theory’), we use the error-function activation

$$\phi(x) = 1 + \operatorname{erf}\left(\frac{\beta\sqrt{\pi}x}{2}\right), \quad (21)$$

where β controls the sharpness. In this setting we typically use $(\sigma, \beta) = (1.42, 2.77)$, carried over from the Setting 2 fit, where σ is the Fourier decay scale of the input-current correlation function $\Gamma^x(\Delta\theta)$. Three analyses depart from these values:

- The grid-cell model (‘Grid cells’), on the two-torus, with $(\sigma, \beta) = (4.04, 2.5)$ at $v=1$ period per 2π and $(12.12, 2.5)$ at $v=3$;
- The asymmetric continuous attractors of Supplementary Note 2, with $\beta=2.5$ and an input-current correlation function set by the manifold geometry, so no σ enters; and
- The Mexican hat construction of Supplementary Note 2, with $\beta=2$ and $\sigma \in \{1/3, 1/2, 1\}$.

We also vary σ and β to examine parameter dependencies (Fig. 7a,d,e and Supplementary Fig. 5). This activation function contains no bias term. The corresponding network dynamics are (same as equation (1) with $b=0$)

$$\tau \partial_t x_i(t) = -x_i(t) + \sum_{j=1}^N J_{ij} \phi_j(t).$$

The weights are constructed via the same least-squares optimization procedure as in the data-derived case, but applied to tuning curves sampled from the generative process.

In Settings 1 and 3, J may additionally undergo modulation by angular velocity $\omega(t)$ (‘Angular velocity integration’). In Settings 1 and 2, we use two mathematically equivalent forms of bias, an external term b in the network dynamics (Setting 1) and a threshold within the activation function (Setting 2). The change of variables $x_i \rightarrow x_i - b$ converts a threshold b into an external bias $-b$.

Nonsaturating and saturating activation functions. For the data-derived network model of Setting 1 and the generative process of Setting 2, we use unbounded activation functions. This is neurobiologically realistic in the sense that cortical neurons typically operate below saturation in their dynamic range⁷¹. Additionally, the experimental tuning curves show no obvious sign of saturation at high activation levels, so inferring a saturation point would not be well posed. For the theoretical analysis of Setting 3, we switch to bounded activation functions for reasons summarized in the introduction to ‘Weights in the large-network limit’. A bounded activation lets us evaluate certain mean-field integrals in closed form, in particular the function $\mathcal{G}(C_{11}, C_{22}, C_{12})$ of Supplementary Note 2 and Supplementary equation (87). Saturation also prevents runaway activity, so the uniform inhibition c used for the data-derived network is no longer needed. Finally, dropping the bias b leaves firing rates nonsparsely over θ , so the

normalization functional $Z[x^*]$ is not needed to amplify faint tuning features. The specific functional form within each class is not critical.

Computing the weight matrix

We use the same method to construct the weight matrix J in Settings 1 and 3.

Problem setup. We minimize the loss function of equation (2) over J subject to $J_{ii} = 0$. As in the analysis of ‘Weights in the large-network limit’, we define Φ as an $N_\theta \times N$ matrix whose columns contain the tuning curves $\phi_i^*(\theta)$ and X as an $N_\theta \times N$ matrix containing the corresponding input currents $x_i^*(\theta)$. Specifically, $X_{ai} = x_i^*(\theta_a)$ and $\Phi_{aj} = \phi_j^*(\theta_a)$, where we sample θ uniformly on $[0, 2\pi]$ at $\theta_a = 2\pi(a-1)/N_\theta$ for $a=1, \dots, N_\theta$, with $N_\theta = 10,000$ in Setting 1 (resampled from the 100-bin grid of ‘Tuning-curve estimation’) and $N_\theta = 500$ in Setting 3. We convert between continuous and discrete representations using the dictionary in Supplementary Information.

Loss function and unconstrained solution. The loss function (same as equation (2)) is

$$\mathcal{L} = \frac{1}{N} \frac{1}{2\pi} \int_0^{2\pi} d\theta \sum_{i=1}^N \left(x_i^*(\theta) - \sum_{j=1}^N J_{ij} \phi_j^*(\theta) \right)^2 + \lambda \sum_{i,j=1}^N J_{ij}^2.$$

We first solve the problem without the diagonal constraint. Differentiating with respect to J_{ij} gives

$$\frac{N}{2} \frac{d\mathcal{L}}{dJ_{ij}} = -\frac{1}{2\pi} \int_0^{2\pi} d\theta x_i^*(\theta) \phi_j^*(\theta) + \sum_{k=1}^N J_{ik} \left(\frac{1}{2\pi} \int_0^{2\pi} d\theta \phi_k^*(\theta) \phi_j^*(\theta) \right) + \lambda \delta_{ij} = 0. \quad (22)$$

Converting to the discretized setting gives

$$\frac{N}{2} \frac{d\mathcal{L}}{dJ} = -\frac{1}{N_\theta} X^T \Phi + J \left(\frac{1}{N_\theta} \Phi^T \Phi + N M_N \right) = 0. \quad (23)$$

Solving this system yields

$$J^{\text{unconstrained}} = \frac{1}{N_\theta} X^T \Phi \left(\frac{1}{N_\theta} \Phi^T \Phi + N M_N \right)^{-1}. \quad (24)$$

Using the push-through identity

$$A(A^T A + \lambda I_N)^{-1} = (A A^T + \lambda I_M)^{-1} A \quad (25)$$

for an $M \times N$ matrix A , this can be rewritten as

$$J^{\text{unconstrained}} = \frac{1}{N} X^T K_\lambda^{-1} \Phi, \quad (26)$$

where the optimization kernel is

$$K_\lambda = \frac{1}{N} \Phi \Phi^T + N_\theta M_{N_\theta}. \quad (27)$$

In the continuous limit, this becomes

$$J_{ij}^{\text{unconstrained}} = \frac{1}{N} \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' K_\lambda^{-1}(\theta, \theta') x_i^*(\theta) \phi_j^*(\theta'), \quad (28)$$

where the optimization kernel is

$$K_\lambda(\theta, \theta') = C^\phi(\theta, \theta') + 2\pi\lambda\delta(\theta - \theta'), \quad (29)$$

and $C^\phi(\theta, \theta') = \frac{1}{N} \sum_{i=1}^N \phi_i^*(\theta) \phi_i^*(\theta')$ is the empirical correlation function of the tuning curves (equation (4)).

Constrained solution. We now enforce $J_{ii} = 0$. The problem decouples across rows of J , because each row predicts $x_i^*(\theta)$ from the regressors $\phi_j^*(\theta)$ for $j = 1, \dots, N$. Setting $J_{ii} = 0$ amounts to dropping regressor $\phi_i^*(\theta)$ from row i and letting the remaining regressors take up the slack. Modifying equations (26) and (27) accordingly gives

$$J_{ij} = \frac{1}{N} x_i^T K_{i,\lambda}^{-1} \phi_j \quad (i \neq j), \quad \text{where} \quad K_{i,\lambda} = \frac{1}{N} \Phi \Phi^T - \frac{1}{N} \Phi_i \Phi_i^T + N_\theta M_{N_\theta}, \quad (30)$$

where $\mathbf{x}_i$ and Φ_i are the N_θ -dimensional vectors containing $x_i^*(\theta)$ and $\phi_i^*(\theta)$ (that is, the i th columns of X and Φ). Taking the continuous limit yields the weight formula (same as equation (3))

$$J_{ij} = \frac{1}{N} \int_0^{2\pi} d\theta \int_0^{2\pi} d\theta' K_{i,\lambda}^{-1}(\theta, \theta') x_i^*(\theta) \phi_j^*(\theta'),$$

where the optimization kernel is

$$K_{i,\lambda}(\theta, \theta') = C^\phi(\theta, \theta') - \frac{1}{N} \phi_i^*(\theta) \phi_i^*(\theta') + 2\pi\lambda\delta(\theta - \theta'). \quad (31)$$

This kernel shows that the constraint of zero on-diagonals provides a small, $\mathcal{O}(1/N)$ correction to the off-diagonals relative to the unconstrained solution. Thus, at large N , it may be omitted.

Unfortunately, the above formula requires inverting a separate matrix for each neuron i , which is computationally expensive. This can be circumvented using an equivalent but different formula. One can derive it using the Sherman–Morrison formula for a rank-one update to a matrix inverse; here, we instead use Lagrange multipliers to obtain the same result. The modified loss function becomes

$$\mathcal{L} = \frac{1}{N} \frac{1}{2\pi} \int_0^{2\pi} d\theta \sum_{i=1}^N \left(x_i^*(\theta) - \sum_{j=1}^N J_{ij} \phi_j^*(\theta) \right)^2 + \lambda \sum_{i,j=1}^N J_{ij}^2 + \frac{2}{N} \sum_{i=1}^N \gamma_i J_{ii}, \quad (32)$$

where γ_i are Lagrange multipliers for $i = 1, \dots, N$. Differentiating with respect to J_{ij} gives

$$\begin{aligned} \frac{N}{2} \frac{d\mathcal{L}}{dJ_{ij}} &= -\frac{1}{2\pi} \int_0^{2\pi} d\theta x_i^*(\theta) \phi_j^*(\theta) \\ &+ \sum_{k=1}^N J_{ik} \left(\frac{1}{2\pi} \int_0^{2\pi} d\theta \phi_k^*(\theta) \phi_j^*(\theta) + N\lambda\delta_{kj} \right) + \delta_{ij} \gamma_i = 0. \end{aligned} \quad (33)$$

In discretized form,

$$\frac{N}{2} \frac{d\mathcal{L}}{dJ} = -\frac{1}{N_\theta} X^T \Phi + J \Omega + \text{diag}(\gamma) = 0, \quad (34)$$

where we have defined the $N \times N$ matrix Ω ,

$$\Omega = \frac{1}{N_\theta} \Phi^T \Phi + N\lambda I_N. \quad (35)$$

In the continuous limit,

$$\Omega_{ij} = \frac{1}{2\pi} \int_0^{2\pi} d\theta \phi_i^*(\theta) \phi_j^*(\theta) + N\lambda\delta_{ij}. \quad (36)$$

Solving for J yields

$$J = J^{\text{unconstrained}} - \text{diag}(\gamma) \Omega^{-1}. \quad (37)$$

The Lagrange multipliers are determined by the constraint $J_{ii} = 0$, yielding

$$\gamma_i = \frac{J_{ii}^{\text{unconstrained}}}{[\Omega^{-1}]_{ii}}. \quad (38)$$

Therefore, the final constrained solution is

$$J_{ij} = J_{ij}^{\text{unconstrained}} - \frac{J_{ii}^{\text{unconstrained}} [\Omega^{-1}]_{ij}}{[\Omega^{-1}]_{ii}}. \quad (39)$$

This is the formula we use in practice.

Angular velocity integration. To make the network integrate angular velocity ω , we replace the target current $x_i^*(\theta)$ with $(1 + \tau\omega\partial_\theta)x_i^*(\theta)$ throughout the derivation above. The target flow on the manifold then points along the tangent vector with magnitude proportional to ω , $\tau\partial_\theta x_i = \tau\omega\partial_\theta x_i^*(\theta)$, rather than vanishing as in ‘Constructing a data-derived network model’. Because the least-squares solution is linear in the target, with or without the constraint $J_{ii} = 0$, the weight matrix separates into a static part and a velocity-coupling part,

$$J(\omega) = J + \tau\omega J', \quad (40)$$

where J is the solution with target X , and J' is the solution obtained by the identical procedure with X replaced by X' , the matrix of angular derivatives $X'_{ai} = \partial_\theta x_i^*(\theta_a)$. Explicitly,

$$\begin{aligned} J^{\text{unconstrained}} &= \frac{1}{N} X'^T K_\lambda^{-1} \Phi, \\ J' &= J^{\text{unconstrained}} - \text{diag}(\gamma') \Omega^{-1}, \\ \gamma'_i &= \frac{[J^{\text{unconstrained}}]_{ii}}{[\Omega^{-1}]_{ii}}, \end{aligned} \quad (41)$$

so that $J'_{ii} = 0$. In simulations, $\omega = \omega(t)$, and the term $\tau\omega(t)J'$ is added to the dynamics.

Simulation details

All simulations used network time constant $\tau = 50$ ms and Euler integration with timestep Δt , except the simulations of Supplementary Fig. 16 and the theory of Supplementary Note 4, which measure time in units of τ .

We simulated recurrent neural networks (Settings 1 and 3) by integrating the rate-based dynamics

$$x_i(t + \Delta t) = \left(1 - \frac{\Delta t}{\tau}\right) x_i(t) + \frac{\Delta t}{\tau} \left(\sum_{j=1}^N J_{ij} \phi_j(t) + b\right). \quad (42)$$

We used $\Delta t = 1$ ms for the data-derived network simulations of Fig. 2 and $\Delta t = 5$ ms for all other network simulations. Network construction used regularization parameter $\lambda = 10^{-8}$ unless otherwise noted.

We simulated the evolution of the overlap order parameter $\bar{m}(\theta, t)$ by Euler integration of a form of the mean-field equations that is exact at large N for any angular velocity and reduces to equation (10) of the main text at slowly varying angular velocity (Supplementary Note 2). Angular velocity enters through a second field $\bar{m}_\omega(\theta, t)$ (‘Angular velocity integration’), so the scheme is

$$\bar{m}(\theta, t + \Delta t) = \left(1 - \frac{\Delta t}{\tau}\right) \bar{m}(\theta, t) + \frac{\Delta t}{\tau} G_{Q(t)}(h(\theta, t)), \quad (43)$$

$$\bar{m}_\omega(\theta, t + \Delta t) = \left(1 - \frac{\Delta t}{\tau}\right) \bar{m}_\omega(\theta, t) + \Delta t \omega(t) G_{Q(t)}(h(\theta, t)), \quad (44)$$

where

$$h(\theta, t) = \frac{1}{2\pi} \int_0^{2\pi} d\theta' [\mathcal{J}(\theta - \theta') \bar{m}(\theta', t) - \partial_\theta \mathcal{J}(\theta - \theta') \bar{m}_\omega(\theta', t)]. \quad (45)$$

At zero angular velocity, $\bar{m}_\omega(\theta, t) = 0$, and the first equation is the scheme on its own. For main text analyses, we used $\Delta t = 1$ ms. For supplementary analyses, we used $\Delta t = 5$ ms.

Statistics and reproducibility

This is a theoretical and computational study. No new experiments were performed for this work. The only experimental data analyzed are previously published recordings from the postsubiculum of freely moving mice, collected by Duszkievicz et al.¹⁶ and publicly available as DANDI Dandiset 000939 (ref. 39).

Sample size. No statistical method was used to predetermine sample size. The size of the dataset was fixed by the original study, namely 31 recording sessions, 1 per mouse, comprising 2,691 units, of which $N_{\text{data}} = 1,533$ were classified as head-direction cells.

Data exclusions. No recording sessions were excluded. Our analyses use the subset of units that Duszkievicz et al.¹⁶ classified as head-direction cells, on the basis of significant head-direction information relative to a time-reversed control. That criterion was defined by the original study's authors, and the resulting labels are distributed with the dataset. All 1,533 units carrying the head-direction label were analyzed. For each session, only the epoch recorded during exploration of the square environment was analyzed. Nine mice additionally underwent optogenetic manipulation, which took place in a separately recorded epoch that no analysis touches. Those mice are otherwise included in full, contributing 315 of the 1,533 head-direction cells from their square-environment epochs.

Randomization and blinding. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. Neither consideration applies here because we performed no new experiments and analyzed a public dataset with no experimental groups, conditions or interventions to allocate. For the randomization and blinding of the original recordings, see the Methods of Duszkievicz et al.¹⁶.

Statistical tests. Uniformity of the within-session distribution of tuning-curve circular centers of mass was tested with the Kuiper test, a two-sided goodness-of-fit test, against the uniform distribution on $[0, 2\pi)$, applied independently to each of the 31 sessions. The 31 resulting *P* values were corrected for multiple comparisons with the Benjamini–Hochberg false discovery rate procedure at the 5% level. Two of the 31 sessions deviated significantly ($P < 0.05$) before correction and none after correction, with corrected *P* values ranging from 0.59 to 0.97.

Reproducibility. Simulations and analyses are deterministic given a fixed random seed, and every notebook that draws random numbers sets its seed explicitly. The released code reproduces every figure in the paper from the released cached results. Those cached results are themselves regenerable by downloading the raw recordings from the DANDI Archive and running the same code, so no step of the pipeline is a black box.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw electrophysiological and behavioral recordings analyzed in this study were collected by Duszkievicz et al.¹⁶ and are publicly available from the DANDI Archive as Dandiset 000939 (ref. 39). The cached intermediate results used by the analysis notebooks are publicly available via Zenodo at <https://doi.org/10.5281/zenodo.21827768> (ref. 72). Every cached file is regenerable from the raw data and the released code.

Code availability

All analysis code is publicly available as an installable Python package, ring, together with 11 Jupyter notebooks that generate every figure in

the paper, apart from two panels of the grid-cell figure that reproduce experimental data from previously published studies. The code is released under the MIT License and is available on GitHub at <https://github.com/davidclark1/symmetries-continuous-attractors>, archived on Zenodo at <https://doi.org/10.5281/zenodo.21829584> (ref. 73). Analyses were run under Python 3.12 with NumPy 1.26.4, SciPy 1.13.1, Matplotlib 3.8.4, PyTorch 2.4.0, pynwb 2.6, Astropy 6.1 and statsmodels 0.14, and the full environment specification is distributed with the code.

References

- Rubin, D. B., Van Hooser, S. D. & Miller, K. D. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**, 402–417 (2015).
- Clark, D. G., Abbott, L. F. & Sompolsinsky, H. Cached intermediate results for ‘Symmetries and continuous attractors in disordered neural circuits’. *Zenodo* <https://doi.org/10.5281/zenodo.21827768> (2026).
- Clark, D. G., Abbott, L. F. & Sompolsinsky, H. Code for ‘Symmetries and continuous attractors in disordered neural circuits’. *Zenodo* <https://doi.org/10.5281/zenodo.21829584> (2026).

Acknowledgements

We thank A. Litwin-Kumar, J. Zavatone-Veth, J. Cotler and H. Grier for helpful discussions. Anthropic's Claude and OpenAI's ChatGPT were used to edit and organize the text, to generate code for analyses and plotting and to check the paper for errors, including errors in the mathematical expressions.

Author contributions

D.G.C., L.F.A. and H.S. conceived of the study. D.G.C. developed the core theoretical framework, with input from L.F.A. and H.S., and performed all analyses, simulations and figure production. D.G.C. wrote the paper. L.F.A. and H.S. provided input on the paper at all stages. L.F.A. and H.S. supervised the project. All authors discussed the results and approved the final paper.

Funding

D.G.C. was supported during the early stages of this work by the Gatsby Charitable Foundation and the Kavli Foundation and, during the later stages, by a Kempner Institute Research Fellowship. L.F.A. was supported by the Gatsby Charitable Foundation and the Kavli Foundation. H.S. was supported by the Kempner Institute and the Gatsby Charitable Foundation. This work has been made possible in part by a gift from the Chan Zuckerberg Initiative Foundation to establish the Kempner Institute at Harvard University. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41593-026-02452-4>.

Correspondence and requests for materials should be addressed to David G. Clark.

Peer review information *Nature Neuroscience* thanks Tatiana Engel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection. No new experiments were performed for this study.
Data analysis	Python 3.12 with NumPy 1.26.4, SciPy 1.13.1, Matplotlib 3.8.4, PyTorch 2.4.0, pynwb 2.6, Astropy 6.1 and statsmodels 0.14. Custom analysis code is distributed as the Python package 'ring' together with 11 Jupyter notebooks that reproduce every figure, with the full environment specification included. Available at https://github.com/davidclark1/symmetries-continuous-attractors and archived on Zenodo at https://doi.org/10.5281/zenodo.21829584 .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw electrophysiological and behavioral recordings analyzed in this study were collected by Duszkievicz et al. [16] and are publicly available from the DANDI

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

n/a

Reporting on race, ethnicity, or other socially relevant groupings

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical method was used to predetermine sample size. The dataset size was fixed by the original study of Duzkiewicz et al. (2024), comprising 31 recording sessions, one per mouse, and 2,691 units, of which 1,533 were classified as head-direction cells and analyzed here.

Data exclusions

No recording sessions were excluded. Analyses use the subset of units that Duzkiewicz et al. (2024) classified as head-direction cells, on the basis of significant head-direction information relative to a time-reversed control. That criterion was pre-established by the original study and the labels are distributed with the dataset. For each session, only the epoch recorded during exploration of the square environment was analyzed

Replication

This is a theoretical and computational study. All simulations and analyses are deterministic given a fixed random seed, and every notebook sets its seed explicitly. All figures were successfully reproduced from the released code and cached results in a clean environment. All attempts at replication were successful.

Randomization

No new experiments were performed and the study analyses a previously published dataset with no experimental groups, conditions or interventions to allocate.

Blinding

No new experiments were performed and there were no group allocations to be blinded to. For the original recordings, see the Methods of Duzkiewicz et al. (2024).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a