



Article

<https://doi.org/10.1038/s41593-024-01766-5>

Maintaining and updating accurate internal representations of continuous variables with a handful of neurons

Received: 5 February 2023

Marcella Noorman , Brad K. Hulse, Vivek Jayaraman , Sandro Romani & Ann M. Hermundstad

Accepted: 14 August 2024

Published online: 3 October 2024

Check for updates

Many animals rely on persistent internal representations of continuous variables for working memory, navigation, and motor control. Existing theories typically assume that large networks of neurons are required to maintain such representations accurately; networks with few neurons are thought to generate discrete representations. However, analysis of two-photon calcium imaging data from tethered flies walking in darkness suggests that their small head-direction system can maintain a surprisingly continuous and accurate representation. We thus ask whether it is possible for a small network to generate a continuous, rather than discrete, representation of such a variable. We show analytically that even very small networks can be tuned to maintain continuous internal representations, but this comes at the cost of sensitivity to noise and variations in tuning. This work expands the computational repertoire of small networks, and raises the possibility that larger networks could represent more and higher-dimensional variables than previously thought.

The brain is thought to rely on persistent internal representations of continuous variables for a wide range of computations, from working memory^{1–4} to navigation^{5–9} to motor control^{10–12}. Such internal representations have been described in terms of manifolds along which population activity evolves (Fig. 1a, top), and they have been studied theoretically within the framework of continuous attractor networks^{2,3,5,7,11,13}; see refs. 14–16 for recent reviews. This framework for continuous attractor networks has historically relied on large numbers of neurons to ensure that these internal representations are approximately continuous and accurate, and this requirement becomes even more crucial in multiple dimensions and to represent multiple variables. Theories of navigation, for example, rely on large numbers of neurons to explain how continuous attractors could underlie the activity of head direction (HD), place, and grid cells in multiple dimensions^{17–19}, and how the hippocampus might build multiple continuous attractors corresponding to different environments that an animal has visited^{5,20,21}. Here, we ask whether such continuous representations can be maintained in much smaller networks.

One prominent example of a continuous attractor network is the ring attractor network, which can maintain an internal representation of a periodic variable such as orientation^{13,22}, and has been proposed as a model of the HD system^{9,23–26}. Ring attractor networks derive their name from the one-dimensional ring manifold on which activity evolves. This manifold emerges in the limit that an infinitely large population of orientation-tuned neurons maintains sustained and localized activity through positive feedback^{13,22,24}; this can be achieved through recurrent connectivity by which neurons with similar tuning excite one another, and neurons with dissimilar tuning inhibit one another (Fig. 1a, bottom, and refs. 13,22,24,27, but also see ref. 28). The resulting population dynamics can generate a localized bump of activity that persists at the same orientation in the absence of input and traverses the ring manifold through the integration of self-motion inputs^{23–25}. As a result of their infinite size, ring attractor networks achieve infinite precision in maintaining and accurately updating the bump of activity. Large networks have been used to approximate this infinite precision^{2,4,7};

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. e-mail: noormanm@janelia.hhmi.org; hermundstada@janelia.hhmi.org

small networks, in contrast, exhibit notable failures that are indicative of finite, rather than infinite, precision^{29–31}. Consistent with these studies, we work under the a priori assumption that achieving infinite precision in representing periodic variables requires infinitely large networks (see the Supplementary Note for further discussion).

Although ring attractor networks were proposed theoretically several decades ago, it has been difficult to identify ring-like architectures in brains. Ring attractor networks have been used to explain bell-shaped tuning curves of mammalian HD neurons that display persistent firing in the absence of input and whose activity is updated by self-motion even in darkness^{6,32}, but it has not yet been possible to measure patterns of connectivity between these neurons. Mammalian HD neurons have been observed to coherently change their tuning when animals are placed in different settings⁶, and recent works suggests that HD population dynamics traverse a one-dimensional ring-like manifold³³. In the fly *Drosophila melanogaster*, a recurrent network of neurons in a brain region called the central complex (CX; Fig. 1b) was recently shown to exhibit the functional and structural connectivity^{34–36} (Fig. 1c), as well as the dynamics^{8,30,34,37} (Fig. 1d,e), of a ring-like attractor network. These dynamics are observable as a bump of population activity in so-called EPG or ‘compass’ neurons in a toroidal structure of the CX called the ellipsoid body (EB). This bump of activity tracks the fly’s orientation during turns and persists when the fly stops moving (Fig. 1e). These dynamics are driven both by localizing sensory cues and by the integration of self-motion cues, which enables the bump to track the fly’s movements even in darkness^{8,30,37}. The underlying circuit architecture features two subpopulations of ‘shift’ neurons that are jointly tuned to orientation and angular velocity and that receive input from and project back to the compass neurons^{30,35–37}, as previously hypothesized²³ (Fig. 1a, bottom). Thus, both physiological and anatomical considerations suggest that this circuit exhibits the key features of a ring-like attractor network, with one major exception: the fly circuit has far fewer computational units—sets of neurons with the same HD tuning—than are thought necessary to approximate an accurate ring attractor³⁰. This low number is likely conserved across many insects, including those that are considered more accomplished navigators, such as bees³⁸, suggesting that it does not limit navigational performance. Motivated by these observations, we sought to characterize the capabilities of small networks to represent and integrate an analog, periodic variable. In what follows, we dissect the functional properties of discrete ring-like attractor networks, and show how small circuits might overcome limitations of discreteness to achieve functional performance thought to emerge only in the limit of large systems.

Results

The computational properties that make ring attractor networks such appealing models of the HD system arise in the limit of large system sizes. Specifically, in the limit that the number of neurons approaches

infinity (what we term a ‘continuous’ system), a ring attractor network generates a continuum of configurations that define the ring attractor manifold^{13,22,24} (Fig. 1f, top). These configurations are marginally stable, such that perturbations along the manifold will be maintained, and perturbations off the manifold will be driven back to it. These properties allow us to express the manifold as a flat dimension in the energy landscape of the system⁷; all points along this flat dimension have equal and minimum energy; thus, the system can stably sit at any of these points in the absence of input (Fig. 1g, second column, dark blue). Moreover, small changes in input can drive the system along this flat dimension without obstruction, such that the population activity accurately tracks these changes^{23–25} (Fig. 1g, third and fourth columns, dark blue). This flat energy dimension gives the system infinite precision in encoding and updating an internal representation of a one-dimensional circular variable such as HD.

However, when the system is small (what we term a ‘discrete’ system), these properties are thought to break down, thereby limiting how precisely the internal HD representation can be stored and updated. Instead of exhibiting a flat dimension, the energy landscape is assumed to exhibit a set of discrete basins (Fig. 1f, bottom) that attract the population activity in the absence of input³⁹ (Fig. 1g, second column, light blue), prevent the integration of small inputs¹⁴ (Fig. 1g, third column, light blue), and prevent the accurate integration of large inputs (Fig. 1g, fourth column, light blue). For a small network such as the fly compass network, we would thus expect to observe three distinct signatures of discreteness: (1) drift in the absence of input, in which the HD bump drifts to stereotyped orientations around the EB when the fly stops turning; (2) failure to integrate small angular velocities, in which the HD bump does not move continuously when the fly makes slow turns; and (3) variable responses to larger angular velocities, in which the HD bump moves faster or slower relative to the fly’s movements, depending on its orientation within the EB.

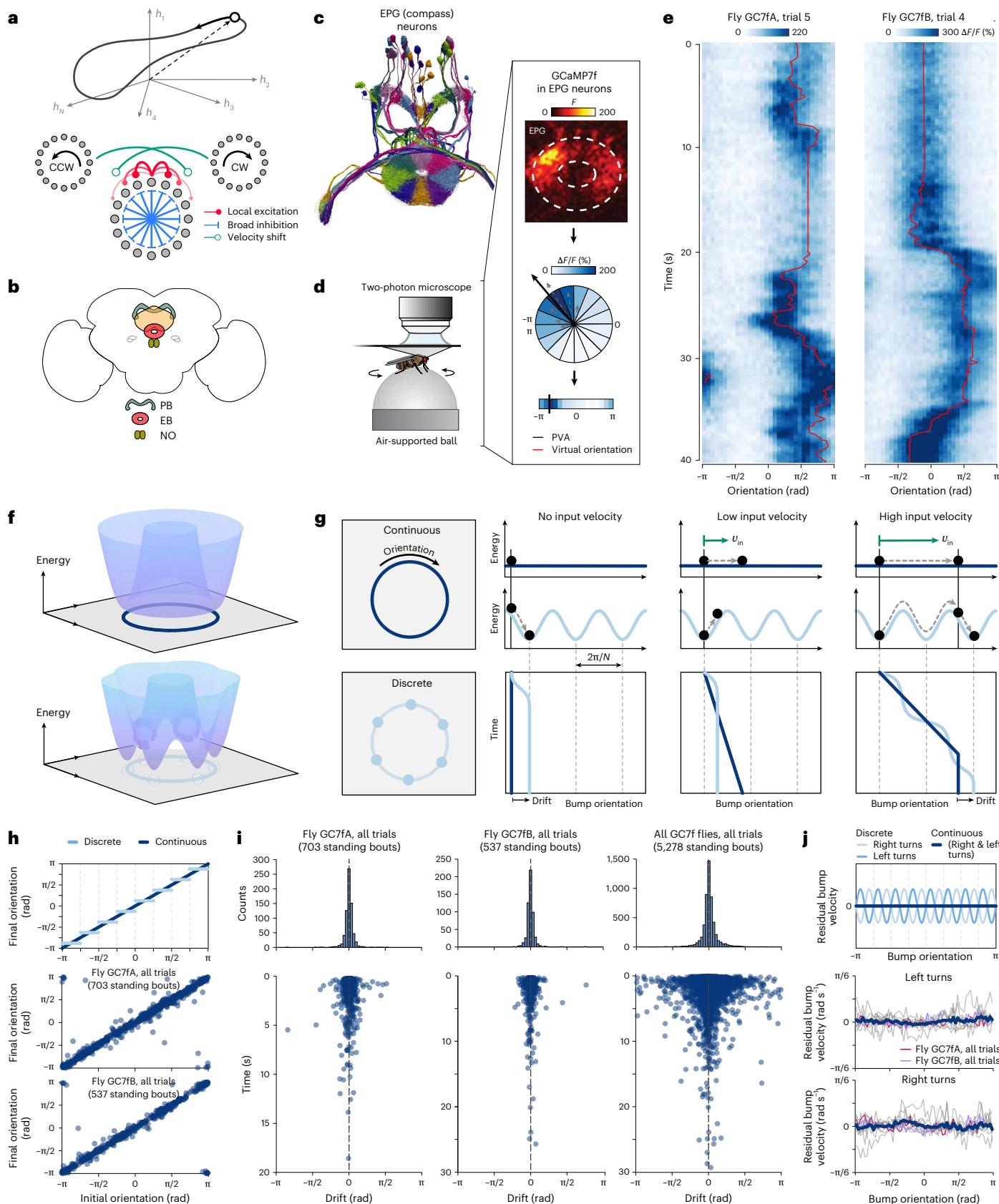
To assess whether the fly circuit can overcome these expected limitations, we performed two-photon calcium imaging of compass neurons in the EB while head-fixed flies walked on an air-supported ball in darkness (Fig. 1d,e,h–j and Methods). While fly-to-fly variability in the accuracy of integration may be due, in part, to limitations of the fly-on-a-ball system (Methods), several flies showed a remarkable ability to track changes in their angular orientation in darkness. We first measured bump drift in the absence of input⁸ by comparing the bump orientation when the fly stopped moving to when the fly began walking again. The distributions of initial and final bump orientations were similar (Extended Data Fig. 1), and there were no apparent signatures that the bump drifted to a discrete number of stereotypical orientations (Fig. 1h). The distribution of drifts was strongly peaked at zero (Fig. 1i, top row), and included epochs in which the bump persisted at the same orientation for several seconds⁸ (Fig. 1i, bottom row). We then analyzed the average bump velocity at different orientations as

Fig. 1 | A biological attractor network overcomes hypothesized limitations of discreteness. **a**, Top: ring-like manifold of neural activity. Bottom: a ring attractor network maintains an internal representation of orientation through local excitation (red) and broad inhibition (blue). Two side rings use angular velocity input to shift this representation (green). CW, clockwise; CCW, counterclockwise. **b**, Schematic of the fly CX. ‘Compass’ neurons innervate the EB and maintain an internal representation of orientation. ‘Shift’ neurons innervate the protocerebral bridge (PB) and shift the representation through angular velocity input from the noduli (NO). **c**, Electron microscopy reconstruction of compass neurons. **d**, Two-photon imaging setup for tethered walking flies. Box: 32 regions of interest (ROIs) are used to compute the population vector average (PVA) of the change in fluorescence ($\Delta F/F$). **e**, Compass neurons maintain a localized bump of activity (heatmap) that tracks the fly’s orientation (red line). **f**, In the absence of input, network dynamics evolve toward the minima of an energy landscape. Infinitely large networks generate flat landscapes (top); small networks generate bumpy landscapes (bottom; illustrated for $N = 6$ neurons).

g, In continuous networks (dark blue), a flat landscape allows activity to persist at the same orientation in the absence of input (second column) and to integrate velocity input linearly (third and fourth columns). In discrete networks (light blue), local minima cause drift in the absence of input (second column), prevent continuous integration of small inputs (third column), and cause nonlinear integration of large inputs (fourth column). **h**, Bump orientations in the EB before and after stopping periods that exceeded 300 ms, schematized for discrete versus continuous networks (top) and shown for the same flies from **e** (middle and bottom). **i**, Distribution of bump drifts (top histograms) accumulated across stopping periods (bottom scatterplots), shown for the same two flies (left and middle columns) and accumulated across flies (right column). **j**, Residual bump velocities during left versus right turns as a function of bump orientation in the EB, schematized for discrete versus continuous networks (top) and shown for individual flies (middle and bottom; dark blue lines show population averages). Bump velocities were normalized for gain differences before computing residuals (Methods).

a function of the fly's average turning velocity. Again, across several flies, the bump velocity was consistent across orientations, with no apparent signatures of nonlinear integration nor apparent failures to track small velocities (Fig. 1j and Extended Data Fig. 2). Thus, despite

the imperfections of measuring the accuracy of the HD representation in head-fixed flies on a ball, we found that the peak performance of the HD system belied its small size both in its low drift and in its accurate integration.



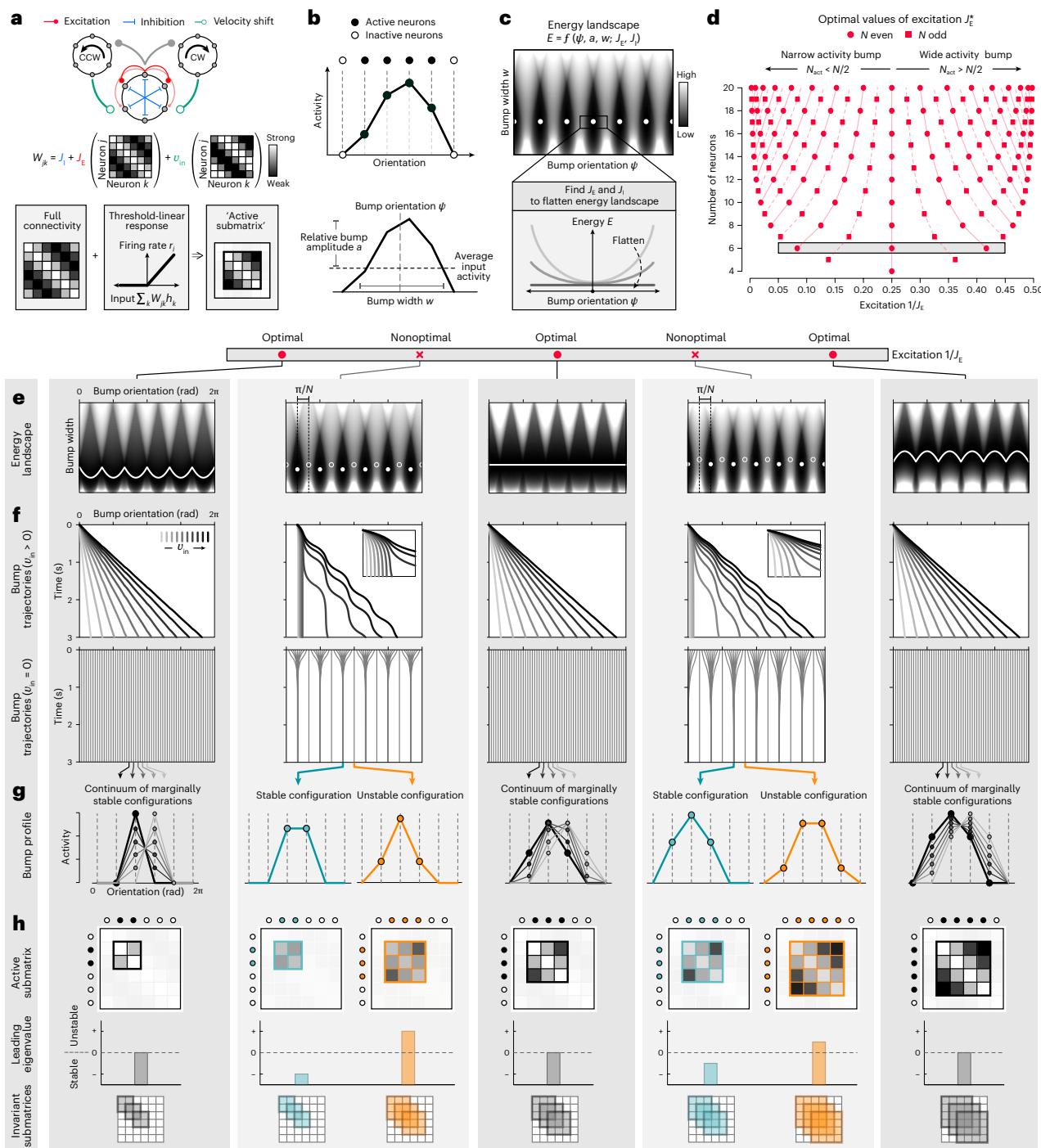


Fig. 2 | Optimally tuned local excitation can recover a ring attractor manifold. **a**, Schematic of the network model and connectivity W_{jk} . Top: a population of neurons is recurrently connected through local excitation (J_E) and broad inhibition (J_I). Two side rings receive input from and project back to the center ring with shifted, velocity-dependent connections. Bottom: a threshold-linear response function ensures that a subset of N_{act} neurons is active at any time; their dynamics are governed by an 'active submatrix' of the full connectivity. **b**, Top: J_E and J_I can be selected to maintain a persistent bump of population activity. Bottom: characterization of the bump configuration (Methods). **c**, Top: energy of different bump configurations for naive choices of J_E and J_I . The resulting landscape is bumpy, with local minima (white points) separated by barriers. Bottom: we sought parameters that 'flatten' the energy landscape by minimizing local curvature. **d**, For a network of size N , there are $N - 3$ optimal values of J_E that flatten the energy. Shaded bar: optimal values of excitation for a network size of $N = 6$ (see e-h). **e-h**, We evaluate the performance (rows) of networks of size $N = 6$

with different values of J_E (columns; $J_E^* = [12, 4, 2.4]$ (optimal); $J_E = [6, 3]$ (nonoptimal)). **e**, Same as c, for different values of J_E . Optimal energy landscapes are flat (white line); nonoptimal landscapes have local minima (filled markers) separated by barriers (open markers). **f**, Bump trajectories in response to a constant input (top row) and in the absence of input (bottom row). Insets show zoomed-in portions of trajectories, which highlight the failure to integrate small inputs. **g**, Same as b, shown for bump configurations at the endpoints in f. **h**, Top row: same as heatmaps in a, shown for active submatrices corresponding to the bump configurations in g. Filled markers denote active neurons. Middle row: the leading eigenvalue of each submatrix governs the dynamics of active neurons. Bottom row: in optimal networks, the bump is always maintained by the same number of active neurons (gray); in nonoptimal networks, the bump is maintained by different numbers of active neurons depending on whether the bump configuration is stable (turquoise) or unstable (orange).

Small networks generate a continuum of stable configurations

The previous results suggest that small networks can, in practice, integrate angular velocity without suffering the performance failures expected of discrete systems. To explore how this might be achieved in principle, we studied the performance of small attractor networks (Fig. 2a and Methods).

We considered networks of N orientation-tuned neurons whose preferred orientations θ_j uniformly tile orientation space, with an angular separation of $\Delta\theta = 2\pi/N$ radians (rad). These neurons can be arranged topologically in a ring according to their preferred orientations, with neurons locally exciting and broadly inhibiting their neighbors. We capture this with a symmetric cosine weight matrix $W_{jk}^{\text{sym}} = J_l + J_E \cos(\theta_j - \theta_k)$, where J_E and J_l respectively control the strength of the tuned and untuned components of recurrent connectivity between neurons with preferred orientations θ_j and θ_k . We will refer to these components as local excitation and broad inhibition, respectively (but note that the tuned component takes on both positive and negative values, and thus is not strictly excitatory; within the parameter regimes that we consider, the untuned component is strictly inhibitory). The network receives angular velocity input ν_{in} through asymmetric, velocity-modulated weights $W_{jk}^{\text{asym}} = \sin(\theta_j - \theta_k)$ (see also ref. 24); this input could be implemented through two linear side rings whose time constants are much smaller than that of neurons in the center ring (Supplementary Note). Each neuron transforms its inputs through a nonlinear transfer function $\phi(\cdot)$. The total input activity h_j of each neuron is then governed by

$$t\dot{h}_j = -h_j + \frac{1}{N} \sum_k (W_{jk}^{\text{sym}} + \nu_{\text{in}} W_{jk}^{\text{asym}}) \phi(h_k) + c_{\text{ff}}, \quad j = 1, \dots, N, \quad (1)$$

where c_{ff} is a constant feedforward input to all neurons in the network. In what follows, we take $\phi(\cdot)$ to be threshold linear; this ensures that only a subset of all neurons will be active at any time. As a result, the dynamics of active neurons will be governed by an ‘active submatrix’ of the full connectivity (Fig. 2a, bottom). We derive our theoretical results for networks of arbitrary size $N < \infty$; unless otherwise noted, we illustrate these results using a network of size $N = 6$ because this is the smallest network that exhibits the range of dynamics observed across parameter tunings.

For sufficiently strong local excitation and broad inhibition, this network generates a stable bump of activity (Fig. 2b (top), Extended Data Fig. 3a and Methods). We characterize the bump by the Fourier modes of the population activity (given by equation (1)). For the network connectivity chosen here, which varies sinusoidally with the difference between preferred orientations, the population activity is fully specified by the zeroeth- and first-order Fourier modes. This allows us to characterize the ‘configuration’ of the activity bump in terms of its relative amplitude a , angular width w , and angular orientation ψ (Fig. 2b (bottom) and Supplementary Note). These quantities vary continuously over time, and thus, the same number of active neurons can maintain bump configurations with different relative amplitudes, widths, and orientations.

We began by characterizing the manifold of stable bump configurations in the absence of angular velocity input (Extended Data Fig. 3b–i and Methods). To this end, we constructed a landscape that describes the energy of different bump configurations for a given set of parameters J_E and J_l (refs. 40,41 and Methods). For most parameter settings, the energy landscape is bumpy, with discrete minima separated by barriers (Fig. 2c, top), as expected for small networks³⁹. The landscape is highly curved about these minima, indicating that the bump would be highly attracted to these particular orientations. To weaken this attraction, we analytically determined the values of J_E and J_l that would locally minimize this curvature, and thus locally flatten the energy landscape (Fig. 2c, bottom). Surprisingly, we found that specific values of local excitation drive the curvature to zero, resulting in an energy

landscape that is completely flat as a function of orientation (Extended Data Fig. 4). For a network of size N , there are $N - 3$ such ‘optimal’ values of local excitation J_E^* (Fig. 2d). Figure 2e illustrates the corresponding optimal energy landscapes for a network of size $N = 6$, and contrasts these with two nonoptimal landscapes generated with intermediate values of local excitation.

To verify that these optimally tuned networks could overcome the failure modes highlighted in Fig. 1g, we simulated the response of each network to a constant velocity input (Fig. 2f and Methods). As expected, we found that optimal networks accurately integrated angular velocity input, such that the bump orientation changed linearly over time (Fig. 2f, top row). When this velocity input was removed (Fig. 2f, bottom row), the bump persisted at the same orientation and did not drift (we also observed this in networks with different nonlinearities and connectivity profiles in one and two dimensions; Extended Data Fig. 5 and Methods). In contrast, nonoptimal networks failed to integrate small velocities (Fig. 2f, top row insets), and they nonlinearly integrated larger velocities (Fig. 2f, top row main panels). When this velocity input was removed, the bump drifted toward the set of discrete orientations corresponding to the local minima of their energy landscapes (Fig. 2f, bottom row).

In the absence of velocity input, optimal networks generate a continuum of marginally stable configurations in which the bump can persist (Fig. 2g). These configurations share one striking feature: the bump is always maintained by the same number of active neurons despite variations in relative amplitude, width, and orientation. This feature has important consequences for network dynamics: when a fixed subset of neurons is active, equation (1) for $h_j > 0$ reduces to a linear dynamical system that depends only on an ‘active submatrix’ of the full connectivity W (Fig. 2h, top row; note that we take the full connectivity to be $W = (W^{\text{sym}}/N - I)/t$). Moreover, because the connectivity is rotationally invariant, this active submatrix—and thus the resulting network dynamics—will be identical for any contiguous subset of N_{act} active neurons. To characterize these dynamics, we determined the eigenvalue spectra of these active submatrices (Methods). Each submatrix exhibited a single zero eigenvalue (Fig. 2h, middle row); the real part of all remaining eigenvalues was less than zero. This property gives rise to a so-called line attractor that produces a continuum of marginally stable configurations along a line⁴¹. Thus, in this network, a ring attractor emerges as a discrete set of N line attractors that each governs the dynamics of distinct subsets of active neurons (Fig. 2h, bottom row), and that are ‘stitched together’ at the points where an active subset gains and loses an active neuron.

In contrast, nonoptimal networks can only maintain a discrete set of bump configurations in the absence of input; these configurations correspond to so-called fixed points of the dynamics. One subset of these configurations is stable; the bump will return to these stable fixed points following small perturbations (Fig. 2g, turquoise curves). The other subset is unstable; the bump will move away from these unstable fixed points if perturbed (Fig. 2g, orange curves). In these two configurations—stable and unstable—the bump is maintained by different numbers of active neurons (also called the ‘support’ of the fixed point^{42,43}), and the corresponding active submatrices differ in size (Fig. 2h, top row). The smaller of these submatrices has a leading eigenvalue less than zero and governs network dynamics about the stable fixed point, whereas the larger of these submatrices has a leading eigenvalue greater than zero and governs dynamics about the unstable fixed point (Fig. 2h, middle row). In what follows, we use these active submatrices to dissect the dynamics of nonoptimal networks, and we show how the balance between stable and unstable dynamics shapes performance.

Variations in tuning degrade network performance

The previous results highlight a unique feature of threshold-linear networks: when a fixed subset of neurons is active, the corresponding

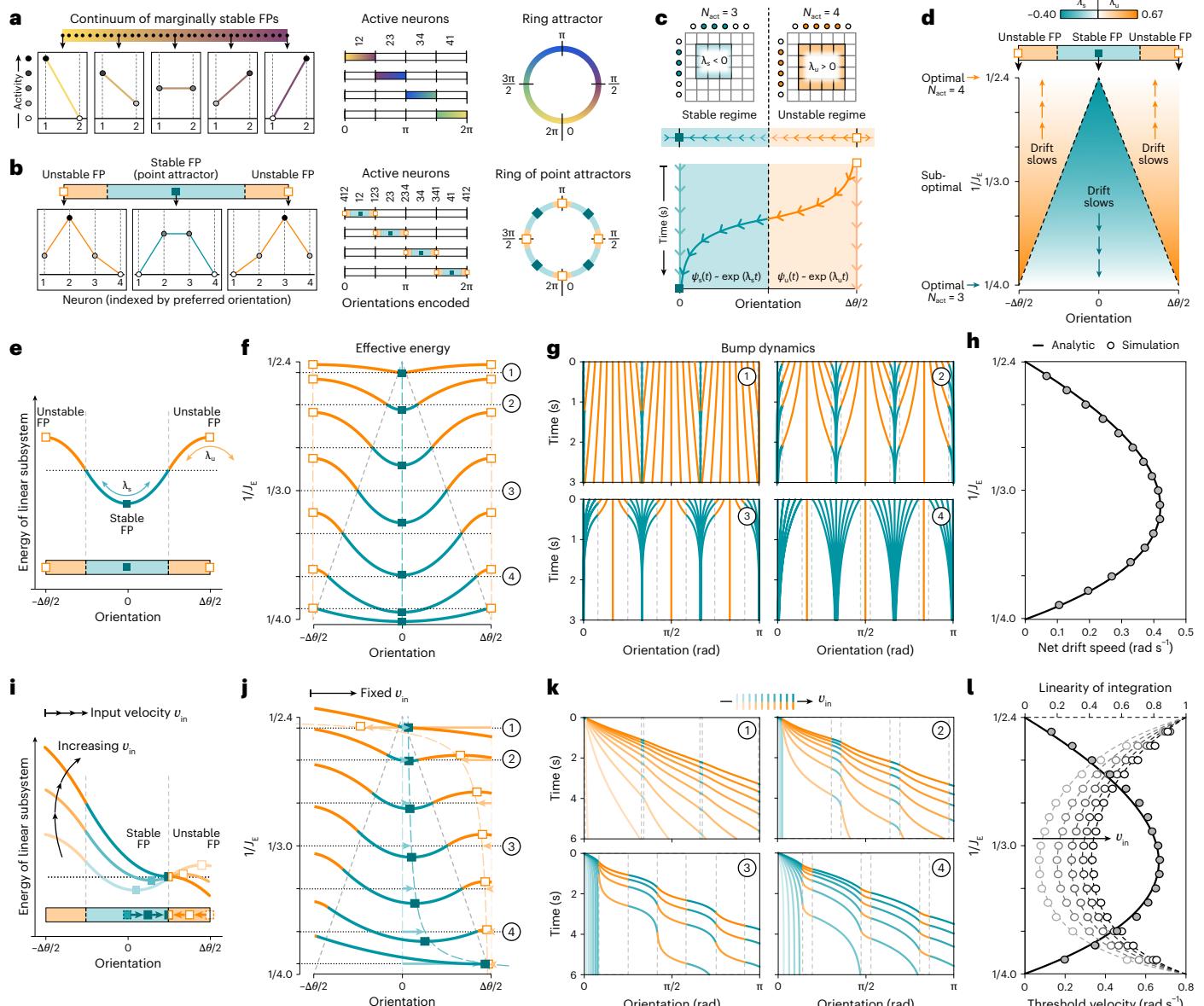


Fig. 3 | Nonoptimal networks balance periods of stability and instability.

a, A linear subsystem of active neurons can be tuned to encode a continuum of orientations over a fixed interval (heatmap; left). Multiple line attractors can be stitched together at orientations where the active subset simultaneously gains and loses an active neuron (middle), thereby generating a ring attractor (right). **b**, Without precise tuning, each linear subsystem (shaded region; left) encodes a single unstable or stable fixed point ('FP'; markers). When stitched together (middle), the set of linear subsystems can stably encode only a finite number of orientations ('point attractors'; right). **c**, Top: the dynamics of each linear subsystem are governed by the leading eigenvalue λ of the active submatrix of the connectivity (Fig. 2h). Bottom: in the unstable regime (orange), the bump accelerates away from an unstable fixed point at rate $\lambda_u > 0$; in the stable regime (turquoise), the bump decelerates toward a stable fixed point at rate $\lambda_s < 0$. **d**, Bump dynamics depend on the fixed-point orientations (square markers), drift rates λ (color map), and angular span of each regime (colored areas). Illustrated

without velocity input. **e–h**, Bump dynamics without velocity input. **e**, Simplified energy landscape. **f**, Same as **e** for different J_E . As J_E approaches an optimal value, one region of the landscape flattens and fills the entire ring; the other sharpens and shrinks in span. **g**, Bump dynamics for energy landscapes in **f**. **h**, Net drift speed, computed analytically (line) and by simulation (markers). **i–l**, Bump dynamics with velocity input. **i**, Small velocities shift the fixed points toward the boundary between stable and unstable regimes, tipping the energy landscape in the direction of the input. At a threshold velocity (equation (5)), the fixed points meet at the boundary, and the bump slides continuously down the landscape. **j**, Same as **i** for different J_E , given a fixed input velocity. J_E affects how quickly the fixed points move through the energy landscape, and, thus, how readily the landscape tips for a given velocity. **k**, Bump dynamics for energy landscapes in **j**. **l**, Threshold velocity (solid curve) and linearity of integration (dashed curves), computed analytically (lines) and by simulation (markers).

dynamical system is linear, and the dynamics of the full network can be viewed as a set of linear subsystems that are stitched together at points where the active subset gains or loses an active neuron. In this way, a ring attractor that encodes a continuum of values on a circle can be constructed by stitching together multiple line attractors that each encode a continuum of values on a line segment. Because a line attractor can be constructed from a network with as few as two neurons, a minimal ring

attractor could, in principle, be constructed using only three neurons. However, our choice of connectivity requires a minimum of four neurons to construct a ring attractor, in which each contiguous pair of neurons encodes a distinct line attractor (Fig. 3a). This requires a precise handoff between linear systems that share active neurons, such that the network dynamics move between line attractors by simultaneously activating and inactivating single neurons at the edges of the active subset.

Achieving this precise handoff requires precise tuning, such that the leading eigenvalue λ of all active submatrices of W is zero. Without a zero eigenvalue, a linear subsystem can, at most, encode a single stable or unstable fixed point. By interleaving linear subsystems that encode stable and unstable fixed points, the network can still cover a circular interval, but the values that can be stably represented are limited to a discrete set (Fig. 3b). In the vicinity of an unstable fixed point (the ‘unstable’ regime), the bump is pushed exponentially quickly away from the fixed point with rate $\lambda_u > 0$ (Fig. 3c, orange). In the vicinity of a stable fixed point (the ‘stable’ regime), the bump is pulled exponentially slowly toward the fixed point with rate $\lambda_s < 0$ (Fig. 3c, turquoise). The bump transitions from the unstable to the stable regime when the active subset loses an active neuron.

This picture highlights how nonlinear computations, such as the integration of angular velocity, can be performed through an orchestrated interaction between multiple linear subsystems that have different fixed-point structures⁴⁴. By decomposing the full dynamical system into linear subsystems, this picture allows us to analytically characterize inaccuracies in nonoptimal networks, and thereby estimate the precision in tuning required to bound these inaccuracies. We measure these inaccuracies using the expected signatures of discreteness highlighted in Fig. 1g (drift in the absence of input, failure to integrate small inputs, and nonlinear integration of large inputs), and we relate these to a simplified description of the energy landscapes shown in Fig. 2e. A complete description of the energy landscape is not attainable in the presence of velocity inputs due to the asymmetry that it introduces in the connectivity matrix (Fig. 2a); to circumvent this, we construct an approximate description that relies on three features of the linear subsystems described above: (1) the orientations of the unstable and stable fixed points, (2) the rates at which the bump is pushed from or pulled toward these fixed points, and (3) the angular span of the regimes governed by each fixed point. As we will show, the local excitation determines the overall curvature of the energy landscape through the rates and angular spans of each regime, which affects the amount of drift. Input velocity shifts the fixed points within this landscape, which influences the accuracy of velocity integration.

Drift in the absence of input. In the absence of velocity input, the stable and unstable fixed points are evenly spaced by $\Delta\theta/2 = \pi/N$ rad regardless of the strength of local excitation. However, the local excitation affects how quickly the bump moves relative to each fixed point, which, in turn, affects the rate of drift in the network. If we vary the local excitation between two optimal values, $J_{E,n}^*$ and $J_{E,n+1}^*$ (corresponding to scenarios in which the bump is always maintained by n or $n+1$ active neurons, respectively), we find that the drift rates λ_s and λ_u depend on how closely tuned the local excitation is to either optimal value (Fig. 3d and Extended Data Fig. 6):

$$\begin{aligned}\lambda_s &= (J_E/J_{E,n}^* - 1)/\tau < 0, \\ \lambda_u &= (J_E/J_{E,n+1}^* - 1)/\tau > 0.\end{aligned}\quad (2)$$

Thus, in the stable regime, where the bump is maintained by n active neurons, the dynamics depend on how closely tuned the excitation is to the value that would be optimal if n neurons maintained the bump. Similarly, in the unstable regime, where the bump is maintained by $n+1$ active neurons, the dynamics depend on how closely tuned the excitation is to the value that would be optimal if $n+1$ neurons maintained the bump. Assuming that the bump orientation transitions smoothly between regimes (as seen in simulations; Fig. 2f, top row), the relative widths $\Delta\theta_{s,u}/\Delta\theta$ of these regimes depend on the ratio of the drift rates (Fig. 3d):

$$\frac{\Delta\theta_s}{\Delta\theta} = \frac{1}{1 + |\lambda_s|/|\lambda_u|} = 1 - \frac{\Delta\theta_u}{\Delta\theta}. \quad (3)$$

Together, these expressions enabled us to construct a simplified landscape that captures the energy of different bump orientations within each linear subsystem (Fig. 3e and Methods). The fixed points determine the locations of extrema within the landscape, the drift rates determine the curvature of the landscape about these extrema, and the angular spans of each regime delineate different regions of the landscape that correspond to stable versus unstable dynamics. This description explains how a ring attractor emerges as the connectivity is tuned toward an optimal value (Fig. 3f): at one extreme ($J_E \rightarrow J_{E,n}^*$), the stable region of the landscape flattens and expands to fill the entire ring ($\lambda_s \rightarrow 0, \Delta\theta_s \rightarrow \Delta\theta$), whereas the unstable region sharpens and shrinks in span; at the other extreme ($J_E \rightarrow J_{E,n+1}^*$), the unstable region of the landscape flattens and expands to fill the entire ring ($\lambda_u \rightarrow 0, \Delta\theta_u \rightarrow \Delta\theta$), whereas the stable region sharpens and shrinks in span. These differences in the shape of the energy landscape affect the drift dynamics (Fig. 3g), an effect that we quantify by measuring the net drift speed of the bump (Fig. 3h):

$$|\lambda_d| = c\Delta\theta_s|\lambda_s| = c\Delta\theta_u|\lambda_u|, \quad (4)$$

where $c = (e - 1)/2e$ is a constant. This speed is related to the overall curvature of the landscape, and will be largest at intermediate values of local excitation for which the landscape is bumpiest.

Inaccuracies in velocity integration. When a sufficiently small velocity input is injected into the network, the local curvature and angular span of the stable and unstable regions of the landscape will remain approximately unchanged (Extended Data Fig. 7). However, the orientations of the fixed points will shift toward the boundary between regions, thereby tipping the landscape in the direction of the velocity input and driving the bump to a new stable fixed point (Fig. 3i and Extended Data Fig. 8c,d). The flatter the overall landscape (that is, the smaller the value of $|\lambda_d|$), the more readily the landscape will tip for a given velocity input (Fig. 3j).

At a particular threshold velocity, v_{thresh} , the fixed points will meet at the boundary between regions, thereby enabling the bump to slide down the landscape without getting stuck. This threshold velocity specifies the minimum input that can be continuously integrated by the network, and depends on the overall curvature of the landscape through the net drift speed $|\lambda_d|$:

$$v_{\text{thresh}} \approx |\lambda_d|/2c. \quad (5)$$

The larger the overall curvature of the landscape, the larger the input velocity needed to continuously move the bump (Fig. 3k). In the limit that the local excitation approaches an optimal value, the overall curvature goes to zero, and the network can integrate infinitesimally small inputs (Fig. 3l, solid curve).

Above this threshold velocity, the fixed points will shift outside of their respective regions of the landscape, but their effect will still be felt through the local landscape curvature. As a result, the bump will speed up and slow down as it moves through the unstable and stable regions of the landscape, but it will never get stuck at a fixed point (Fig. 3k and Extended Data Fig. 8e,f). This manifests as nonlinear integration, which we quantify by measuring the ratio between the slowest and fastest bump velocities, v_{\min} and v_{\max} . This ratio depends only on the relative difference between the threshold and input velocities:

$$\text{linearity}(v_{\text{in}}) = \frac{v_{\min}}{v_{\max}} \approx \frac{v_{\text{in}} - v_{\text{thresh}}}{v_{\text{in}} + v_{\text{thresh}}}. \quad (6)$$

Bumpier energy landscapes lead to larger threshold velocities, which lead to increasingly nonlinear integration. However, because the overall curvature (and thus the threshold velocity) is fixed for a given value of local excitation, its relative impact on integration decreases as input

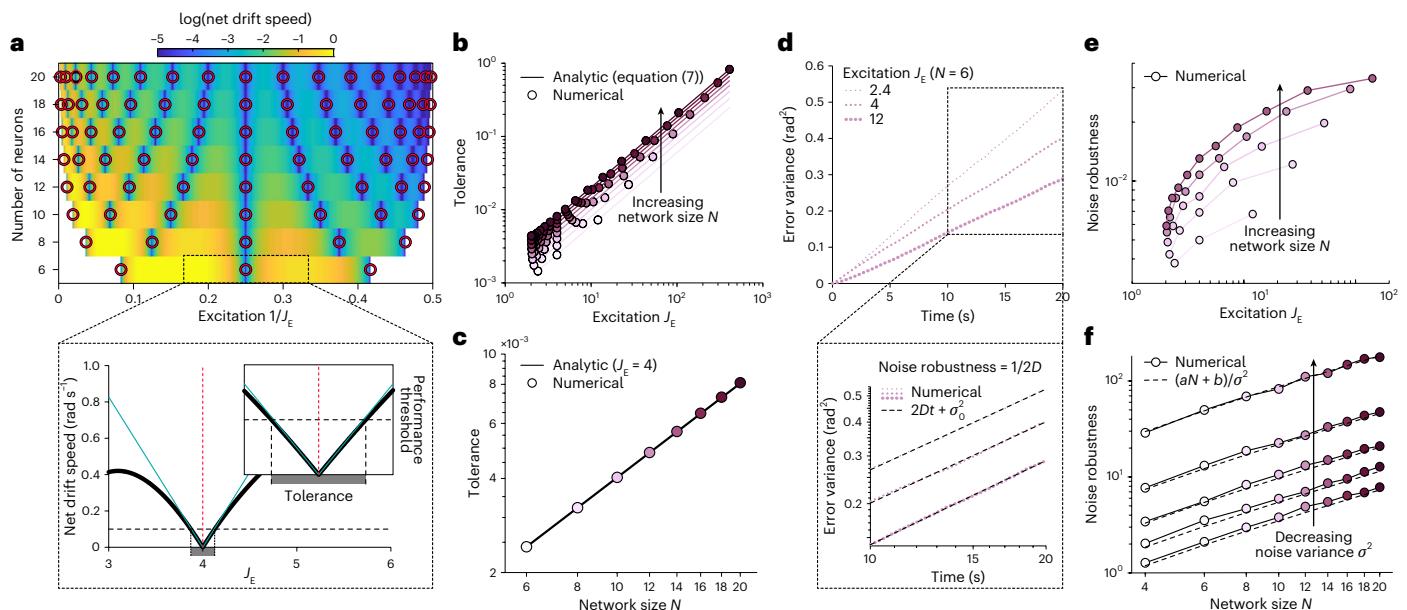


Fig. 4 | Smaller networks require more fine-tuning and are less robust to noise. **a**, Top: log of net drift speed (color map) as a function of J_E and N . Red circular markers indicate optimal values of J_E^* ; darker blue colors indicate slower (that is, better) drift rates. Suboptimal networks achieve better performance as N increases. Bottom: to estimate tolerance around an optimal value of J_E^* , we compute the local change in net drift speed with respect to J_E (turquoise lines) that will achieve performance below some threshold (horizontal dashed black line, illustrated for a threshold of 0.1 rad s^{-1}). **b**, For a given N (different colors), larger values of local excitation require less fine-tuning to achieve the same performance. Solid lines mark the analytic tolerance given in equation (7); filled circles indicate the numerically estimated tolerance about each optimal value of J_E^* . Results were computed for a threshold value of 0.001 rad s^{-1} , and are shown for all evenly sized networks between $N = 6$ and $N = 20$. **c**, Given a fixed value of J_E^* , the tolerance increases linearly with N . Results are shown for $J_E^* = 4$, the only

optimal value of local excitation that remains unchanged with even N . **d**, Top: error variance between the current and initial bump positions in a small, optimally tuned network with additive Gaussian noise. Numerical results are shown for three different optimal values of J_E^* , and with a noise variance $\sigma^2 = (A/6)^2$, where $A = 0.2$ is the bump amplitude. Bottom: beyond 10 s , the error variance grows linearly over time, following a diffusion equation with slope $2D$ (where D is the diffusion coefficient). We use $1/2D$ as a measure of noise robustness, with lower diffusion signifying higher robustness. **e**, Consistent with **d**, larger optimal values of J_E^* lead to higher noise robustness for a fixed N . **f**, Given a fixed value of J_E^* (shown for $J_E^* = 4$), noise robustness increases linearly with N , and is inversely proportional to noise variance σ^2 (shown for $\sigma^2 = (A/6)^2 \times [1, 4, 9, 16, 25]$). Dashed lines indicate best linear fits; see Extended Data Fig. 9 for fit coefficients.

velocity increases (Fig. 3*i*, dashed curves). In the limit that the local excitation approaches an optimal value, the threshold velocity goes to zero, and the bump moves continuously at the rate of the input velocity.

Optimal small networks are less robust

The previous results provide a mechanistic understanding of how small networks can achieve optimal performance through the precise tuning of local excitation. To assess the potential cost of this precision, we used the previous results to characterize how size affects the robustness of optimal networks.

We first characterized robustness to variations in parameter tuning. For a given network size, deviations from optimal tuning degrade performance through more rapid drift, larger threshold velocities, and more nonlinear velocity integration. In larger networks, this degradation is less severe (Fig. 4*a*, top). To quantify this, we asked how precisely the local excitation should be tuned to meet a criterion level of performance (Fig. 4*a*, bottom). For small values of this criterion, we analytically determined the width of the interval about each optimal value of local excitation J_E^* for which a given measure of network performance meets this criterion; we define the width of this interval to be the tolerance $\text{tol}(J_E^*, N)$:

$$\text{tol}(J_E^*, N) \geq c_p J_E^* N, \quad (7)$$

where c_p is a constant that depends on the specific performance measure (net drift rate, threshold velocity, or linearity of integration) and the desired performance criterion. For a given network size, equation (7) shows that larger optimal values of local excitation permit a wider

range of parameter values that meet the same criterion level of performance, and are thus more robust to parameter tuning (Fig. 4*b*). This robustness increases linearly with network size; this can be seen most clearly for $J_E^* = 4$, which is an optimal value of local excitation for all evenly sized networks (Fig. 4*c*). When summed across all optimal values of local excitation, equation (7) allows us to estimate the net volume of parameter space that achieves a desired performance threshold (Methods). Because larger networks permit more values of optimal excitation and exhibit higher tolerances around these values, we find that the net volume of desirable parameter space increases at least quadratically with network size (Extended Data Fig. 9*a*).

We next characterized robustness to noise. We simulated the dynamics of optimally tuned networks with additive Gaussian noise, and measured how quickly the bump diffused in the absence of velocity input (Fig. 4*d*, top). At longer timescales, the difference between the initial and final bump positions is diffusive, with a variance that grows linearly over time (Fig. 4*d*, bottom). The inverse diffusion rate gives a measure of noise robustness; the faster the diffusion, the less robust the network is to noise. For a given network size, larger optimal values of excitation are more robust to noise (Fig. 4*e*), in qualitative agreement with their increased robustness to variations in parameter tuning (Fig. 4*b*). For a given value of excitation, noise robustness increases linearly with network size, and inversely with the noise variance (Fig. 4*f* and Extended Data Fig. 9*b*).

Together, these results highlight that optimally tuned small networks can recover the performance of infinitely large networks. However, in the networks considered here, this comes at the cost of being less robust to variations in parameter tuning and to noise.

Discussion

Continuous attractor networks have provided a common theoretical framework for studying a wide range of computations¹⁶ involved in working memory^{2–4}, navigation^{5,7,9}, and motor control^{11,12}. Across these different task domains, this framework has historically invoked networks of many neurons to ensure smooth and accurate dynamics. However, growing evidence suggests that similar computations might be performed in much smaller brains with far fewer neurons^{8,30,34,35,37,45}. Here, we asked to what extent network size limits the performance of attractor networks^{3,46}, and whether small networks can overcome these limitations. We focused on a class of attractor networks that maintain a persistent internal representation of a single circular variable, such as orientation, and that update this representation by integrating an internal signal, such as angular velocity. In the limit of infinite numbers of neurons, these ring attractor networks generate a continuous ring manifold along which the population activity smoothly and accurately evolves in the absence of noise. Here, we showed that networks with as few as four neurons could recover this continuous ring attractor manifold, so long as the tuned component of the connectivity (what we term local excitation) is precisely chosen. In the threshold-linear networks studied here, this manifold emerges as a set of line attractor manifolds that govern the dynamics of active subsets of neurons, and that are stitched together to generate a complete ring manifold. The resulting population activity can persist at any orientation in the absence of input, and it can smoothly integrate velocity input.

Together, these results suggest that very small networks can achieve levels of performance that were thought to require large networks. However, this performance comes at the cost of finely tuning local excitation to one of a discrete number of optimal values. Our biological inspiration was the small HD circuit of the fruit fly^{8,30,35,37}. Although such networks have been modeled previously^{29–31,47}, studies have not demonstrated persistent encoding of arbitrary orientations in the absence of orienting stimuli. Further, although previous studies^{31,47} have shown that network performance changes as connection strengths vary, our study fully characterizes how network size and connection strength influence performance. It is unclear whether the fly HD system relies on the fine-tuning that we require for optimal performance. To date, this system has only been probed under head fixation on an air-supported ball (Methods); thus, its performance during free behavior is unknown. Moreover, some inaccuracies in its performance may be attributable to errors in the computation of angular velocity, and not errors in its integration. Our main objective was to investigate the performance and capabilities of small ring-like attractor networks rather than to provide a detailed model of the fly HD circuit per se. As such, there are many differences between the fly circuit and the simple model we explore here, some of which may provide as-yet-undescribed mechanisms to overcome potential problems of discreteness. For example, a potential substrate for tuning local excitation may be the synaptic contacts that fly HD neurons make between themselves in different substructures of the CX^{15,35}. Some of these and other fine-scale details of synaptic connectivity have not been incorporated into existing rate models^{30,34} or spiking neuron models^{29,31,47} of the circuit. In addition, these previous modeling efforts have focused on capturing the dynamics of the circuit without incorporating the biophysical properties of its neurons, and, in most cases, with only a subset of the excitatory and inhibitory cell types likely involved in generating the dynamics. Although the receptor and transmitter profiles of the relevant neurons are known³⁵, further experiments are required to assess how intrinsic neuronal properties shape persistent population activity, as reported in the mammalian HD system⁴⁸. Indeed, these intrinsic properties may account for the low drift we observed in the circuit (Fig. 1*i*) relative to that predicted by the model (Fig. 4*d*). Thus, while our work shows that small networks can, with appropriate tuning, implement continuous ring attractors,

further experiments are needed to understand their cellular and synaptic implementation in real circuits.

Importantly, large ring attractor networks also suffer from the problem of fine-tuning, where noise in the connectivity—arising, for example, from heterogeneity in synaptic or cellular properties—can yield bumpy energy landscapes similar to those generated here (Fig. 2*e*). Several mechanisms have been proposed to combat this issue, including homeostatic synaptic scaling⁴⁹ and synaptic facilitation⁵⁰. These mechanisms might also be effective in the small networks studied here, where—in addition to fine-tuning the profile of the connectivity—the overall strength of local excitation must also be fine-tuned. Away from these optimal values, network dynamics are governed by unstable and stable linear regimes in which the population activity is pushed from or pulled toward discrete fixed points. We identified three properties of these regimes that govern network performance: the angular width of each regime, the locations of fixed points within each regime, and the speed at which the bump is pushed from or pulled toward each fixed point. Varying the strength of local excitation alters the balance between the regimes, such that improving performance in one regime worsens performance in the other. However, as the local excitation approaches an optimal value, the overall performance is dominated by the better-performing regime, which, in the same limit, becomes a ring attractor.

This analysis relied on characterizing the behavior of threshold-linear networks in terms of a separation between different linear dynamical regimes. This separation has recently been used to infer the underlying connectivity of biological networks⁵¹, and to design different connectivity motifs that generate distinct dynamical patterns, for example, to keep count or coarsely represent different positions^{52,53}. Here, we showed how the precise tuning of interactions within a single connectivity motif shapes the properties of these linear regimes, and how these properties, in turn, affect performance. We found that certain regions of parameter space reduce drift and improve integration, and among these ‘good’ parameter regions, some are more robust than others. Specifically, we found that larger optimal values of local excitation, which generate narrower activity bumps, are more robust to variations in tuning and to additive noise, consistent with previous studies of noise robustness in attractor networks^{3,46}.

Our results relied on specific assumptions about network connectivity and dynamics. We assumed local cosine-tuned excitation and broad uniform inhibition, but ring attractor manifolds can be generated with different hand-tuned^{22,24,25,27,54} or learned⁵⁵ connectivity structures. Similarly, velocity integration can be performed in multiple ways, for example, using a network of two rings that receive differential velocity input²⁵, or through two side rings that inherit heading activity from and project back to a center ring with velocity-dependent phase shifts²³, as has been observed experimentally^{30,37}. Our formulation approximates this second implementation in the limit that the side rings have fast neural time constants²⁴. Finally, our choice of a threshold-linear response function enabled us to decompose the dynamics into distinct linear regimes^{42,43} that differentially affect performance, and it allowed us to analytically characterize the tuning precision required to achieve a desired level of performance. In such threshold-linear networks, this precision is limited to the tuned component of the connectivity; however, in networks with other nonlinearities, both the tuned and untuned components must be precisely chosen (Extended Data Fig. 5a). We expect such optimal tunings to exist more generally, provided that the energy of the system varies smoothly with the network tuning. In such cases, parameter-dependent changes in the stability of fixed points must be connected through optimal parameter tunings that locally flatten the energy as a function of orientation, as observed in Fig. 3*f* (Supplementary Note). In the absence of such tuning precision, small networks can fail to integrate velocity inputs and can drift in the absence of input. While such performance failures are known to arise in small attractor networks with differing connectivity structures and

neural response functions^{3,46}, it remains an open question how these different design features affect the relationship between tuning precision and performance more broadly.

While these results were motivated by and interpreted in the context of the small HD system of *Drosophila*, they immediately generalize to other scenarios. For example, the ring attractor network can be used to model place fields in circular environments, grid fields in one dimension, persistent-activity-mediated short-term memory of stimuli represented by angular variables¹, and the preparation of motion toward targets on a circle¹⁰. Our results suggest that such representations could be accurately maintained using few neurons, thereby broadening the classes of computations that could be performed by small circuits. Moreover, these results could further generalize to higher-dimensional continuous variables, such as HD, place, and grid fields in two or three dimensions^{9,17–19} (see Extended Data Fig. 5b for proof-of-principle numerical results). More broadly, the ability to represent one continuous variable accurately using small numbers of neurons could more easily enable large systems to represent multiple continuous variables, such as the representation of many environments observed in the rodent hippocampus^{5,20,21}.

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/s41593-024-01766-5>.

References

- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
- Camperi, M. & Wang, X.-J. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci.* **5**, 383–405 (1998).
- Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
- Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).
- Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).
- Taube, J. S. The head direction signal: origins and sensory–motor integration. *Annu. Rev. Neurosci.* **30**, 181–207 (2007).
- Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).
- Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).
- Finkelstein, A. et al. Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015).
- Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
- Seung, H. S. How the brain keeps the eyes still. *Proc. Natl Acad. Sci. USA* **93**, 13339–13344 (1996).
- Goldman, M. S. et al. Linear regression of eye velocity on eye position and head velocity suggests a common oculomotor neural integrator. *J. Neurophysiol.* **88**, 659–665 (2002).
- Hansel, D. & Sompolinsky, H. in *Methods in Neuronal Modeling: From Synapses to Networks* 2nd edn (eds Koch, C. & Segev, I.) 499–567 (MIT Press, 1998).
- Chaudhuri, R. & Fiete, I. Computational principles of memory. *Nat. Neurosci.* **19**, 394–403 (2016).
- Hulse, B. K. & Jayaraman, V. Mechanisms underlying the neural computation of head direction. *Annu. Rev. Neurosci.* **43**, 31–54 (2020).
- Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* **23**, 744–766 (2022).
- Laurens, J. & Angelaki, D. E. The brain compass: a perspective on how self-motion updates the head direction cell attractor. *Neuron* **97**, 275–289 (2018).
- Ginosar, G. et al. Locally ordered representation of 3D space in the entorhinal cortex. *Nature* **596**, 404–409 (2021).
- Grieves, R. M. et al. Irregular distribution of grid cell firing fields in rats exploring a 3D volumetric space. *Nat. Neurosci.* **24**, 1567–1573 (2021).
- Battaglia, F. P. & Treves, A. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **58**, 7738–7753 (1998).
- Monasson, R. & Rosay, S. Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: collective motion of the activity. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **89**, 032803 (2014).
- 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).
- Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. A model of the neural basis of the rat's sense of direction. *Adv. Neural Inf. Process. Syst.* **7**, 173–180 (1995).
- Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
- Xie, X., Hahnloser, R. H. R. & Seung, H. S. Double-ring network model of the head-direction system. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 041902 (2002).
- Song, P. & Wang, X.-J. Angular path integration by moving “hill of activity”: a spiking neuron model without recurrent excitation of the head-direction system. *J. Neurosci.* **25**, 1002–1014 (2005).
- Amaral, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87 (1977).
- Lim, S. & Goldman, M. S. Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. *J. Neurosci.* **34**, 6790–6806 (2014).
- Pisokas, I., Heinze, S. & Webb, B. The head direction circuit of two insect species. *eLife* **9**, e53985 (2020).
- Turner-Evans, D. et al. Angular velocity integration in a fly heading circuit. *eLife* **6**, e23496 (2017).
- Kakaria, K. S. & de Bivort, B. L. Ring attractor dynamics emerge from a spiking model of the entire protocerebral bridge. *Front. Behav. Neurosci.* **11**, 8 (2017).
- Taube, J. S., Müller, R. U. & Ranck, J. B. Jr Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).
- 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).
- Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).

35. Turner-Evans, D. B. et al. The neuroanatomical ultrastructure and function of a biological ring attractor. *Neuron* **108**, 145–163 (2020).

36. 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).

37. Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).

38. Sayre, M. E., Templin, R., Chavez, J., Kempenaers, J. & Heinze, S. A projectome of the bumblebee central complex. *eLife* **10**, e68911 (2021).

39. Brody, C. D., Romo, R. & Kepecs, A. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr. Opin. Neurobiol.* **13**, 204–211 (2003).

40. Cohen, M. A. & Grossberg, S. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. Syst. Man Cybern.* **SMC-13**, 815–826 (1983).

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. Curto, C., Langdon, C. & Morrison, K. Combinatorial geometry of threshold-linear networks. Preprint at <https://arxiv.org/abs/2008.01032> (2020).

43. Hahnloser, R. H. R. & Seung, H. S. in *Advances in Neural Information Processing Systems* Vol. 13 (eds Leen, T. et al.) 217–223 (MIT Press, 2001).

44. Morrison, K., Degeratu, A., Itskov, V. & Curto, C. Diversity of emergent dynamics in competitive threshold-linear networks. *SIAM J. Appl. Dyn. Syst.* **23**, 855–884 (2024).

45. Petrucco, L. et al. Neural dynamics and architecture of the heading direction circuit in zebrafish. *Nat. Neurosci.* **26**, 765–773 (2023).

46. Burak, Y. & Fiete, I. R. Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl Acad. Sci. USA* **109**, 17645–17650 (2012).

47. Su, T.-S., Lee, W.-J., Huang, Y.-C., Wang, C.-T. & Lo, C.-C. Coupled symmetric and asymmetric circuits underlying spatial orientation in fruit flies. *Nat. Commun.* **8**, 139 (2017).

48. Yoshida, M. & Hasselmo, M. E. Persistent firing supported by an intrinsic cellular mechanism in a component of the head direction system. *J. Neurosci.* **29**, 4945–4952 (2009).

49. Renart, A., Song, P. & Wang, X.-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**, 473–485 (2003).

50. Itskov, V., Hansel, D. & Tsodyks, M. Short-term facilitation may stabilize parametric working memory trace. *Front. Comput. Neurosci.* **5**, 40 (2011).

51. Biswas, T. & Fitzgerald, J. E. Geometric framework to predict structure from function in neural networks. *Phys. Rev. Res.* **4**, 023255 (2022).

52. Parmelee, C., Alvarez, J. L., Curto, C. & Morrison, K. Sequential attractors in combinatorial threshold-linear networks. *SIAM J. Appl. Dyn. Syst.* **21**, 1597–1630 (2022).

53. Londono-Alvarez, J., Curto, C. & Morrison, K. TLN counters, position trackers and central pattern generators. *J. Comput. Neurosci.* **49** (Suppl.1), abstr. P128 (2021).

54. Romani, S. & Tsodyks, M. Continuous attractors with morphed/correlated maps. *PLoS Comput. Biol.* **6**, e1000869 (2010).

55. Darshan, R. & Rivkind, A. Learning to represent continuous variables in heterogeneous neural networks. *Cell Rep.* **39**, 110612 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Methods

Experimental setup

Fly preparation for imaging. We expressed the genetically encoded calcium indicator GCaMP7f (ref. 56) in EPG neurons by crossing GCaMP7f flies (w1118;PBac[20XUAS-IVS-Syn21-op1-GCaMP7f-p10] in VK00005) to the EPG GAL4 driver line SS00096 (ref. 57). Flies (females, age 5–9 days, $n = 10$) were prepared for imaging as previously described^{8,58}. Briefly, flies were anesthetized at 4 °C, their proboscis immobilized with wax to reduce brain movements, and their head/thorax fixed to a holder with a recording chamber using ultraviolet glue. To gain optical access to the brain, we removed a section of cuticle between the ocelli and antennae, along with the underlying fat and air sacs. Throughout the experiment, the head was submerged in saline containing NaCl (103 mM), KCl (3 mM), TES (5 mM), trehalose (8 mM), glucose (10 mM), NaHCO₃ (26 mM), NaH₂PO₄ (1 mM), CaCl₂ (2.5 mM) and MgCl₂ (4 mM), with a pH of 7.3 and an osmolarity of 280 mOsm.

Two-photon calcium imaging. Calcium imaging was performed with a custom-built two-photon microscope controlled with ScanImage (version 2022, Vidrio Technologies)⁵⁹. Excitation of GCaMP7f was generated with an infrared (920 nm), femtosecond-pulsed (pulse width ~110 fs) laser (Chameleon Ultra II, Coherent) with 15 mW of power, as measured after the objective ($\times 60$ Olympus LUMPlanFL/IR, 0.9 numerical aperture). Fast Z-stacks (eight planes with 6-μm spacing and three fly-back frames) were collected at 10 Hz by raster scanning (128 × 128 pixels, $-75 \times 75 \mu\text{m}^2$) using an 8-kHz resonant-galvo system and piezo-controlled Z positioning. Focal planes were selected to cover the full extent of EPG processes in the EB. Emitted light was directed (primary dichroic: 735, secondary dichroic: 594), filtered (filter A: 680 SP, filter B: 514/44) and detected with a GaAsP photomultiplier tube (H10770PB-40, Hamamatsu).

Spherical treadmill system. Following dissection, flies were positioned on an air-supported polyurethane foam ball (8-mm diameter, 47 mg) under the two-photon microscope and allowed to walk. Rotations of the ball were tracked at 500 Hz, as described previously⁵⁸. Behavioral data and imaging timestamps were recorded using Wave-Surfer (version 0.947, <http://wavesurfer.janelia.org/>). For each fly, we collected five 20-min trials during which flies walked or stood in darkness.

Data analysis

All data analysis was performed in MATLAB (version 2022a, MathWorks). Some analyses relied on functions from the Circular Statistics Toolbox (version 2012a)⁶⁰. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications^{8,30,61}. Flies were selected at random from their vials; however, as all data were collected from a single experimental condition (flies walking in darkness), no other randomization was performed. Data collection and analysis were not performed blind to the conditions of the experiments. We excluded any data collected beyond 100 min for consistency and to exclude a small number of flies whose behavior and/or imaging degraded in quality, a known limitation of fly-on-a-ball calcium imaging experiments.

Extracting bump orientation and strength. Each Z-stack was reduced to a single frame using a maximum-intensity projection technique. An ellipse was manually drawn around the perimeter of the EB and automatically segmented into 32 equal-area, wedge-shaped ROIs. The number of ROIs was chosen to be twice the number of anatomically defined EB wedges⁶². Activity within each ROI was averaged for each frame, producing 32 ROI time series. For each ROI time series, baseline fluorescence (F_0) was defined as the average of the lowest 10% of samples. $\Delta F/F$ was computed as $100 \times (F - F_0)/(F_0)$, where F is the instantaneous fluorescence from the raw ROI time series. These ROI

time series were then smoothed with a third-order Savitzky–Golay filter over 11 frames as in previous studies^{8,30}. We used the PVA as a measure of bump strength and orientation. PVA was computed by taking the circular mean of vectors whose angles were the ROI's wedge positions and whose length was equal to the ROI's $\Delta F/F$. The magnitude of this mean resultant vector length was normalized to have a maximum possible length of 1.

Characterizing bump drift. To determine bump drift (Fig. 1h,i), we first identified periods when flies were standing still (defined as zero rotational and translational velocity), disregarding periods shorter than 300 ms. Drift was computed as the circular distance between bump orientations (PVA phase) at the beginning and end of these periods of standing. To determine whether the EPG bump drifted from its initial position to preferred discrete locations within the EB when the fly stood still, we compared the distributions of initial and final bump positions across 64 nonoverlapping bins from $-\pi$ to π around the structure (Extended Data Fig. 1a,b). We used Watson's U^2 test^{63,64}, a nonparametric two-sample test, for this comparison, implemented using MATLAB code from P. Mégavand (watsons_u2, https://github.com/pierremegevand/watsons_u2, 2017). We used 500 permutations to compute P values for this test; these P values, together with the test statistic U^2 , are reported in the caption of Extended Data Fig. 1b. Finally, we computed the distribution of drifts for periods between 300 ms and 2 s across 64 nonoverlapping binned initial positions from $-\pi$ to π around the EB, and fit each fly's drift distribution with sinusoidal functions of the form $A \times \sin(\omega \times \psi + \theta) + C$, where $\omega \in \{8, 16\}$ is the frequency of the sinusoid, ψ is the initial bump position during the standing period, and A, θ, C are learned parameters for the amplitude, phase, and DC offset, respectively (Extended Data Fig. 1c,d). Frequencies of 8 and 16 Hz were chosen to match the number of computational units in the fly's compass network, which, in a discrete network, would cause the bump to drift toward 8 (or 16) distinct bump positions (schematized in Fig. 1h, top). For each fly, we computed the R^2 value between the drift, measured as a function of HD, and the sinusoidal fits (Extended Data Fig. 1c); these R^2 values are reported in Extended Data Fig. 1d.

Characterizing bump velocity. To determine whether the EPG bump shows signs of nonlinear integration (Fig. 1j, top), we measured whether the bump moved faster or slower than expected as a function of bump position for both left and right turns (Fig. 1j, middle and bottom). We began by performing a linear regression (ordinary least squares) between the fly's instantaneous angular velocity and the bump's angular velocity (both sampled at 10 Hz) to account for fly-to-fly variability in the gain of angular integration, as observed in previous studies^{8,30,61}. Linear fits were separately performed for left and right turns, and the residuals were taken as a measure of whether the bump was moving faster (or slower) than expected after accounting for each fly's naive gain. Next, we binned data by bump position (64 nonoverlapping bins from $-\pi$ to π) and computed the average residual bump velocity for each bin, producing the curves shown in the middle and bottom panels of Fig. 1j. Lastly, we fit each fly's curve with sinusoidal functions of the form $A \times \sin(\omega \times \psi + \theta) + C$, where $\omega \in \{8, 16\}$ is the frequency of the sinusoid, ψ is the bump position, and A, θ, C are learned parameters for the amplitude, phase, and DC offset, respectively (Extended Data Fig. 2). Frequencies of 8 and 16 Hz were chosen to match the number of computational units in the fly's compass network, which, in a discrete network, would cause the bump to move faster or slower than expected at 8 (or 16) distinct bump positions (schematized in Fig. 1j, top). For each fly, we computed the R^2 value between the residual bump velocity, measured as a function of HD, and the sinusoidal fits (Extended Data Fig. 2a); these R^2 values are reported in Extended Data Fig. 2b.

We note that our fly-on-a-ball calcium imaging setup comes with potential challenges for evaluating the presence or extent of nonlinear

integration, including slow GCaMP dynamics, altered proprioceptive feedback that the fly may experience while walking on a ball heavier than itself, head fixation that may prevent the fly from altering its head–body angle during turns, potential neural propagation delays involved in relaying and integrating the angular velocity signal, and measurement noise inherent to calcium imaging that could corrupt bump velocity estimation.

Model overview

Network equations. We consider an effective single-ring network of N neurons (or, equivalently, of N computational units; see ‘Network equations’ in the Supplementary Note). Neurons are ordered according to their preferred heading θ_j , which we take to be evenly spaced by $\Delta\theta = 2\pi/N$ rad. Neurons are recurrently connected according to their preferred headings through a symmetric weight matrix $W_{jk}^{\text{sym}} = J_E + J_I \cos(\theta_j - \theta_k)$, where J_E and J_I parametrize the strength of local excitation and uniform inhibition, respectively (note that J_E and J_I actually correspond to tuned and untuned components of the connectivity; for ease of language, we use local excitation and broad inhibition here and throughout). Neurons receive velocity input through an asymmetric, velocity-modulated weight matrix $v_{\text{in}} W_{jk}^{\text{asym}} = v_{\text{in}} \sin(\theta_j - \theta_k)$; in the main text, we took $v_{\text{in}} > 0$. Each neuron j receives a constant feedforward input c_{ff} and a net input $1/N \sum_k (W_{jk}^{\text{sym}} + v_{\text{in}} W_{jk}^{\text{asym}}) r_k$ from all other neurons in the network, where the firing rate $r_k = \phi(h_k)$ is a nonlinear function of the total input activity h_k . For all analyses shown in the main text, we took the nonlinear transfer function $\phi(\cdot)$ to be rectified linear (that is, $\phi(\cdot) = [\cdot]_+$, but see also Extended Data Fig. 5 and ‘Robustness to changes in the transfer function and recurrent weights’ in the Methods). The dynamics of each neuron are given by the system of single-neuron equations in equation (1); we chose $\tau = 0.1$ s and $c_{\text{ff}} = 1$.

By applying a discrete Fourier transform to the single-neuron equations, we can express this system of equations in terms of its Fourier modes. After initial transients, only the DC and first-order modes remain, and the resulting dynamical system reduces to a set of three equations that govern the dynamics of the orientation ψ , amplitude a relative to the average input activity, and width w of the bump (‘Order equations’ in the Supplementary Note); we will refer to these as the system of bump equations.

Stable parameter regime. The system of bump equations will generate a stable bump of activity for certain combinations of J_E and J_I (‘Fixed point analysis’ in the Supplementary Note and Extended Data Fig. 3a). For all analyses shown in the main text, we first selected a desired value of $J_E > 2$, and then selected a value of J_I such that it produced a bump of activity whose full amplitude $A = H_0 + a$ (where H_0 is the average input activity) was at least approximately $A = 0.2$. To do so, we first uniformly sampled bump orientations $\psi \in [0, 2\pi]$ and widths $w \in [2\pi/N, 2(N-1)\pi/N]$, and we used these to calculate the contour $J_E f_{\text{even}}(w, \psi) = 1$ using MATLAB’s ‘contourc.m’, where $f_{\text{even}}(w, \psi)$ is given by equation (S19) in the Supplementary Note (see also equation (S30) in the Supplementary Note and Extended Data Fig. 3c). This gave us values $(w, \psi) \in C_{J_E} = \{(w, \psi) | J_E f_{\text{even}}(w, \psi) = 1\}$ that satisfy the contour equation. We then used these values of w and ψ to determine an upper bound on J_I given by

$$J_I^{\text{bound}} = \min_{(w, \psi) \in C_{J_E}} \frac{-\cos(w/2)}{f_0(w, \psi)}, \quad (8)$$

where $f_0(w, \psi)$ is given by equation (S18) (see also equation (S32)) in the Supplementary Note. We then used these same values of w and ψ to determine a value for J_I , given by

$$J_I = \min_{(w, \psi) \in C_{J_E}} \frac{(c_{\text{ff}}/A - 1) \cos(w/2) - c_{\text{ff}}/A}{f_0(w, \psi)}, \quad (9)$$

and verified that $J_I < J_I^{\text{bound}}$. Plugging $A = 0.2$ into equation (9) resulted in a bump of activity whose minimum full amplitude was approximately $A = 0.2$.

Model analytics

Stationary solutions. To determine the configurations to which the system evolves in the absence of velocity input, we characterized the stationary solutions of the system of bump equations (‘Fixed point analysis’ in the Supplementary Note). This allowed us to determine relationships between the bump orientation, relative amplitude, and width that would persistently maintain a stable bump of activity (Extended Data Fig. 3b,c). For a network of N neurons that receive no velocity input, most parameter settings will yield two sets of N fixed points each—one set will be stable, and the other will be unstable. For a given value of J_E , one set will be aligned with the preferred headings $\{\theta_j\}$, and the other set will be aligned precisely between the preferred headings; the second and fourth columns of Fig. 2e highlight examples for which the unstable (second column) and stable (fourth column) sets of fixed points are aligned with the preferred headings. The value of J_E and the parity of N (whether the network consists of an even or odd number of neurons) together specify which of these two configurations the network will adopt. When N is even and $J_E < J_{E,N-2}^*$ (denoting bumps supported by $N-1$ and $N-2$ neurons), the set of fixed points aligned with the preferred headings will be unstable. When N is odd, the reverse will be true: for $J_E < J_{E,N-2}^*$, the set of fixed points aligned with the preferred headings will be stable. For a given network size N , as J_E passes through an optimal value J_E^* , this stability switches (Extended Data Fig. 3d,g). At each of these fixed points, the widths of the stable and unstable bump configurations are determined solely by J_E , whereas their relative amplitudes depend on both J_E and J_I .

Energy landscape. We derived an energy landscape $E(a, w, \psi; J_E, J_I)$ for the system of bump equations in the absence of velocity input^{40,41} (‘Energy landscape’ in the Supplementary Note). This function describes the stable configurations to which the system will evolve in the absence of input.

To minimize the curvature of the energy landscape, we first determined the 3×3 Hessian matrix of the second derivatives of the energy E with respect to a , w , and ψ . When evaluated at the orientations ψ^s of the stable fixed points (see the previous subsection), we found that the Hessian reduced to a block diagonal matrix, with a single eigenvector along ψ whose eigenvalue is given by

$$\frac{\partial^2 E}{\partial \psi^2} \propto 1 - \frac{J_E}{N} \sum_{k \in K_{\text{act}}} \sin^2(\theta_k - \psi^s), \quad (10)$$

where K_{act} denotes the set of indices of the neurons that actively maintain the bump. This eigenvalue quantifies the degree of local curvature of the energy as a function of bump orientation ψ . For a system of size N , there are $N-3$ values of local excitation J_E for which this eigenvalue goes to zero, and thus for which the energy landscape is locally flat as a function of ψ . These correspond to bump configurations for which the bump is maintained by $N_{\text{act}} \in [2, N-2]$ active neurons:

$$\frac{1}{J_{E,N_{\text{act}}}^*} = \frac{1}{4} + \frac{1}{2N} \left(\bar{n} + \frac{\sin(2\pi\bar{n}/N)}{\sin(2\pi/N)} \right); \bar{n} = N_{\text{act}} - \frac{N}{2}. \quad (11)$$

We found that these values of local excitation, which are shown in Fig. 2d, also ensure that the energy landscape is flat for all bump orientations (as shown in Fig. 2e; also see Extended Data Fig. 4).

Leading eigenvalues of active submatrices. In the absence of velocity input, the bump dynamics are governed by the leading eigenvalue λ of a submatrix of the connectivity $(-I + W^{\text{sym}}/N)/\tau$; this eigenvalue determines the rate at which the bump will drift in the absence of input. When

the local excitation J_E is optimally tuned (that is, $J_E = J_{E,N_{act}}^*$), the bump of activity will be maintained by a fixed number of active neurons $N_{act} \in [2, \dots, N-2]$. For each distinct value of N_{act} , there is thus a distinct $N_{act} \times N_{act}$ submatrix of the connectivity whose single leading eigenvalue determines the drift dynamics. Away from these optimal values of local excitation, the bump of activity will be maintained by either n or $n+1$ active neurons (see equation (S50) in the Supplementary Note). The drift dynamics are then governed by the leading eigenvalues of the corresponding $n \times n$ and $(n+1) \times (n+1)$ active submatrices.

To determine these dynamics, we analytically determined the rates of bump drift in the stable and unstable regimes, which are given in equation (2) (see 'Performance of non-optimal solutions: Dynamics in the absence of input velocity', and, in particular, equations (S54) and (S56) in the Supplementary Note). We then compared these analytically derived drift rates to the leading eigenvalues that we computed numerically by directly diagonalizing active submatrices of the connectivity (using the MATLAB function 'eig.m'); this comparison is shown in Extended Data Fig. 6.

Widths of stable and unstable regimes. In the absence of input, the widths of the stable and unstable regimes can be determined analytically by finding the orientation at which the bump transitions from unstable to stable dynamics as it drifts away from an unstable fixed point. This reduces to matching two exponential equations that govern the dynamics of the bump orientation in the two regimes (with drift rates λ_u and λ_s , respectively), and that must tend toward the orientations of the unstable and stable fixed points as $t \rightarrow -\infty$ and $t \rightarrow +\infty$, respectively. The resulting widths of each regime are given by equation (3) and shown in Fig. 3d and Extended Data Fig. 8b, and they are centered on the orientations of the stable and unstable fixed points in the absence of input. Given a stable fixed point at $\psi = \psi^s$ and an unstable fixed point at $\psi = \psi^u = \psi^s + \pi/N$, the resulting equation for the bump can then be written as (see equations (S61) and (S62) in the Supplementary Note):

$$\psi(t) = \begin{cases} \psi^u + (\psi_0 - \psi^u) \exp(\lambda_u t); & 0 < t < t_{\Delta n} \text{ unstable regime} \\ \psi^s + \frac{\Delta\theta_s}{2} \exp(-|\lambda_s|(t - t_{\Delta n})); & t > t_{\Delta n} \text{ stable regime} \end{cases} \quad (12)$$

where $\psi^s + \Delta\theta_s/2 < \psi_0 < \psi^u$ is the initial orientation of the bump, and $t_{\Delta n} = (1/\lambda_u) \log(\Delta\theta_u/(2(\psi^u - \psi_0)))$ is the time when the bump orientation crosses from the unstable regime into the stable regime. See 'Performance of non-optimal solutions: Dynamics in the absence of input velocity' in the Supplementary Note for more details.

Drift in the absence of input. To measure the net bump drift, we analytically computed the time τ_d that it takes for the bump to drift from within ε_u of an unstable fixed point to within ε_s of a stable one. We chose $\varepsilon_u = \Delta\theta_u/2e$ and $\varepsilon_s = \Delta\theta_s/2e$, such that the bump covered an angular distance of $\Delta\psi_d = (1 - 1/e)\Delta\theta/2$ in the time τ_d . We then measured the net drift speed as $\Delta\psi_d/\tau_d$ (see equations (S68)–(S71) in the Supplementary Note).

Small velocity approximation. In the presence of velocity input, the bump dynamics will be governed by the leading eigenvalue λ of a submatrix of the full connectivity $(-I + (W^{\text{sym}} + v_{\text{in}} W^{\text{asym}})/N)/\tau$. The asymmetric component of this connectivity is modulated by the input velocity v_{in} , and introduces a velocity-dependent correction to the eigenvalue λ_0 of the symmetric connectivity $(-I + W^{\text{sym}}/N)/\tau$ (Extended Data Fig. 7):

$$\lambda \approx \lambda_0 + f(J_E)v_{\text{in}}^2 + \mathcal{O}(v_{\text{in}}^3). \quad (13)$$

For sufficiently small input velocities, we can approximate the leading eigenvalues λ_u and λ_s , and thus the corresponding widths of the unstable and stable regimes, as being equal to their values in the absence of velocity input (see 'Leading eigenvalues of active submatrices' and

'Widths of stable and unstable regimes' in the Methods). All analytic results shown in Fig. 3i–l were generated under this assumption. This approximation breaks down as the input velocity increases, and it breaks down more quickly for smaller values of local excitation (as shown in Fig. 3l; see also Extended Data Fig. 7a).

Locations of fixed points in a velocity-driven regime. Although we can approximate the rates and width of the stable and unstable regimes as remaining unchanged for a sufficiently small velocity input, we cannot make the same approximation for the orientations of stable and unstable fixed points. Therefore, we will treat the stable and unstable fixed-point orientations as functions of v_{in} : $\psi^s = \psi^s(v_{\text{in}})$, $\psi^u = \psi^u(v_{\text{in}})$, respectively. The orientation of the stable and unstable fixed points found in the absence of velocity input will then be given by $\psi^s(0)$ and $\psi^u(0)$, respectively. To determine how the orientations of these fixed points shift with velocity, we repeated the analyses described in 'Widths of stable and unstable regimes' in the Methods, but with a different set of initial conditions (see 'Performance of non-optimal solutions: Dynamics in the presence of small input velocity' in the Supplementary Note for details). Given a bump that begins at a stable fixed point $\psi = \psi^s(0)$ in the absence of input, and given an initial velocity v_{in} , the bump will be driven to a new stable fixed point at an orientation $\psi^s(v_{\text{in}}) = \psi^s(0) + v_{\text{in}}/|\lambda_s|$ as $t \rightarrow \infty$. In the limit that $t \rightarrow -\infty$, the bump will be driven to (and hence, in forward time, away from) an unstable fixed point at an orientation $\psi^u(v_{\text{in}}) = \psi^u(0) - v_{\text{in}}/|\lambda_u|$. Over an interval $\psi \in [\psi^s(0) - \Delta\theta_s/2, \psi^u(0) + \Delta\theta_u/2]$, the resulting equation for the bump can be written as (see equations (S78) and (S79) in the Supplementary Note):

$$\psi(t) = \begin{cases} \psi^s(0) + \frac{v_{\text{in}}}{|\lambda_s|} (1 - \exp(-|\lambda_s|t)); & 0 < t < t_c \text{ stable regime} \\ \psi^u(0) - \frac{v_{\text{in}}}{\lambda_u} + \left(\frac{v_{\text{in}}}{\lambda_u} - \frac{1}{2}(\Delta\theta - \Delta\theta_s) \right) \exp(\lambda_u(t - t_c)); & t > t_c \text{ unstable regime} \end{cases} \quad (14)$$

where $t_c = (1/|\lambda_s|) \log(1/(1 - \Delta\theta_s|\lambda_s|/2v_{\text{in}}))$ is the time when the bump orientation crosses from the stable regime into the unstable regime.

At the threshold velocity given in equation (5), the two fixed points will meet at the boundary between regimes; this is the minimum velocity needed for the bump to move continuously. Below this velocity, the bump will be driven away from the unstable fixed point in the unstable regime, and toward a stable fixed point in the stable regime. Above this velocity, the stable and unstable fixed points will still drive the bump dynamics, but their orientations will move outside of their respective regimes. The minimum and maximum bump velocities, v_{min} and v_{max} (given by equation (6)), can be computed analytically from equation (14) by evaluating the time derivative of $\psi(t)$ at the boundary from the stable to the unstable regime, and vice versa. We used these minimum and maximum velocities to define the linearity of integration as $v_{\text{min}}/v_{\text{max}}$. See 'Performance of non-optimal solutions: Dynamics in the presence of small input velocity' in the Supplementary Note for details.

Simplified energy landscape. Having described each linear subsystem in terms of (1) the orientations of the fixed points, (2) the rate at which the bump drifts toward or away from these fixed points, and (3) the angular regime governed by each fixed point, we used these three properties to construct a simplified landscape that describes the energy of different bump orientations. Given a linear system, an energy function can be chosen to be quadratic⁶⁵; we thus choose $E_{u,s}(\psi) = \alpha_{u,s}\psi^2$, where $\alpha_s > 0$ for the stable subsystem, and $\alpha_u < 0$ for the unstable subsystem. To select the appropriate values of $\alpha_{u,s}$, we require that the energy function has extrema at the orientations of the stable and unstable fixed points $\psi^s(v_{\text{in}})$ and $\psi^u(v_{\text{in}})$, and that the energy transitions smoothly between the stable and unstable regimes; this yields

$$E(\psi) = \begin{cases} -\lambda_u(\psi - \psi^u(v_{in}))^2 + C(v_{in}) & \text{unstable regime} \\ -\lambda_s(\psi - \psi^s(v_{in}))^2 & \text{stable regime} \end{cases}, \quad (15)$$

where $C(v_{in}) = \Delta\theta(\lambda_u\Delta\theta_u/4 - v_{in} + v_{in}^2/(\lambda_u\Delta\theta_u))$ and where $\alpha_u = -\lambda_u < 0$ and $\alpha_s = -\lambda_s > 0$ as required. When moving around the ring, each successive pair of stable and unstable regimes will be governed by an energy landscape of this form but with a vertical shift, such that $E(\psi \pm n\Delta\theta) = E(\psi) \mp 2nv_{in}\Delta\theta$. See 'Simplified energy' in the Supplementary Note for more details.

Tolerance in tuning. To determine how precisely the local excitation must be tuned to achieve a criterion level of performance, we first computed the derivative of each performance measure as a function of local excitation, evaluated at an optimal value; we denote this $m_p(J_E^*)$ (see equations (S96)–(S99) in the Supplementary Note). This slope gives us a local linear estimate of how quickly the performance degrades away from an optimal value of local excitation. Because each performance measure can be expressed as a function of the net drift speed $|\lambda_d|$, computing this slope reduced to computing $\partial|\lambda_d|/\partial J_E|_{J_E^*}$. Given a criterion for the system to be within $\varepsilon_p^{\text{tol}}$ of optimal performance for a performance measure P , the tolerance about a given optimal value J_E^* can then be computed as $\text{tol}_P(J_E^*) \geq 2\varepsilon_p^{\text{tol}}/|m_p(J_E^*)|$ (where \geq indicates that this is a lower bound on the tolerance, as the linear slope will overestimate the rate of degradation of performance; see equation (S113) in the Supplementary Note).

To determine the volume of parameter space that can meet this desired performance, we summed the tolerance across all optimal values of local excitation for a given network size N (see equation (S120) in the Supplementary Note). We then approximated this sum by its largest value, which reduces to

$$V(N) \geq c_p \frac{N^2}{1 - \cos(2\pi/N)}. \quad (16)$$

See 'Degradation of performance as a function of local excitation' in the Supplementary Note for more details.

Model simulations

Overview. All simulations that we performed used MATLAB's ODE solver 'ode45.m' with an integration timestep of $\Delta t = 0.01$ s. We first initialized the network to generate a bump of activity at a given orientation ψ . Using this as the initial condition for the network, we then simulated the single-neuron dynamics in equation (1), and we performed a discrete Fourier transform using MATLAB's 'fft.m' function to extract the bump dynamics as a function of the single-neuron dynamics (see equation (S16) in the Supplementary Note). When simulating angular velocity integration, we first determined the velocity scaling that would generate a comparable rate of bump movement for a given (constant) velocity input (see 'Velocity-driven dynamics' in the Methods). We then simulated the network dynamics in response to this scaled input.

Parameter choices. All results shown in Figs. 2 and 3 were generated using networks of size $N = 6$. When illustrating network properties for different values of local excitation, we used the following values of J_E (evenly spaced in $1/J_E$): $J_E = [12, 6, 4, 3, 2.4]$ (Fig. 2e–h); $J_E = [3.89, 3.6, 3.3, 3, 2.77, 2.57, 2.44]$ (Fig. 3f,j); $J_E = [3.6, 3, 2.57, 2.44]$ (Fig. 3g,k); 17 evenly spaced values of $1/J_E$ between 1/3.86 and 1/2.45 (Fig. 3h,l). When simulating network dynamics in the presence of velocity input, we used the following values of velocity input v_{in} : ten evenly spaced velocity values between 0.2 and 2.0 rad s⁻¹ (Fig. 2f); ten evenly spaced values between 0.1 and 1.0 rad s⁻¹ (Fig. 3k); five evenly spaced values between 0.8 and 1.6 rad s⁻¹ (Fig. 3l). In all cases, we scaled the velocity input as described below (see 'Velocity-driven dynamics' in the Methods).

Drift in the absence of input. For simulations of bump drift, we simulated the network with the velocity input set to zero. To illustrate drift trajectories for different values of J_E (as shown in the bottom row of Fig. 2f and in Fig. 3g), we initialized the bump at six evenly spaced orientations between (and including) 0 and π/N , and we simulated the evolution of the bump for 3 s. We repeated this for repeating angular units between 0 and 2π .

Measuring net drift speed. To measure the net drift speed (as described in 'Drift in the absence of input' in 'Model analytics' in the Methods), we initialized the bump at an orientation $\psi^u - \varepsilon_u$ (where ψ^u is the orientation of an unstable fixed point; for the values of J_E used in Fig. 3, $\psi^u = \pi/N$; see 'Parameter choices' in the Methods). We then simulated the network dynamics until the bump reached an orientation ε_s . We set $\varepsilon_u = \Delta\theta_u/2e$ and $\varepsilon_s = \Delta\theta_s/2e$, where $\Delta\theta_{u,s}$ were computed as described in 'Widths of stable and unstable regimes' in the Methods. We used the time it took for the bump to reach this orientation as the measure of the net drift timescale τ_d , and we used $\Delta\psi_d/\tau_d$ as a measure of net drift speed, where $\Delta\psi_d = (1 - 1/e)\Delta\theta/2$ is the angular distance traveled by the bump in the time τ_d . Fig. 3h compares the net drift speed from simulations to that obtained analytically for different values of J_E .

Velocity-driven dynamics. For simulations of angular velocity integration, we injected a constant velocity input throughout the simulation. To permit a comparison to analytic predictions, we scaled the input velocity such that the rate of movement of the bump matched the input velocity at an input of $v_{in} = 50$ rad s⁻¹. To this end, we determined the best-fitting linear trajectory that minimized the absolute deviation from the bump trajectory over a time window of $t = 6$ s, and we used the slope of this linear trajectory to scale all other input velocities injected into the network. We performed this scaling separately for each set of network parameters (that is, for each choice of (J_E, J_I)). All velocity values described in simulations were scaled in this way.

Measuring threshold velocity. To measure the threshold velocity required to move the bump continuously (as shown in Fig. 3l), we first analytically computed the threshold velocity as described in 'Locations of fixed points in a velocity-driven regime' in the Methods. We then chose 50 evenly spaced input velocity values between (and including) $v_{\text{thresh}} - 0.05$ rad s⁻¹ and $v_{\text{thresh}} + 0.05$ rad s⁻¹. We initialized the bump at the orientation of a stable fixed point (here, at $\psi^s = 0$), and we then simulated the network dynamics in response to each velocity individually. We determined the minimum of these velocities that would move the bump beyond an orientation of π/N within a time interval of 10 s. Fig. 3l compares this simulated value to the value obtained analytically.

Measuring the linearity of integration. To measure the linearity of integration from simulations, we simulated the bump trajectory for different constant input velocities (as described above in 'Overview'). For each input velocity, we determined the time t_c when the bump orientation ψ crossed from the stable into the unstable regime or vice versa; these times were used to compute the minimum and maximum velocities, respectively (note that we used the analytically derived boundaries between regimes to determine these crossing times; see 'Widths of stable and unstable regimes' in the Methods). We then determined the bump velocity as $v = (\psi(t_c + \Delta t) - \psi(t_c - \Delta t))/2\Delta t$, where $\Delta t = 0.1$ s is the integration timestep used in all simulations. Fig. 3l compares this simulated value to the value derived analytically (see 'Locations of fixed points in a velocity-driven regime' in the Methods).

Robustness to variations in parameter tuning. To summarize performance as a function of network size (shown in Fig. 4a), we analytically computed the net drift speed (as described in 'Drift in the absence of input' in 'Model analytics' in the Methods) as a function of local excitation in the range $J_E \in [J_{E,N-2}^*, J_{E,2}^*]$ (that is, between the minimum

and maximum optimal values of local excitation, maintained by $N_{\text{act}} = N - 2$ and $N_{\text{act}} = 2$ active neurons, respectively). For each optimal value of local excitation, we numerically estimated the tolerance as the range of local excitation values about an optimum for which the net drift speed would be consistently below a fixed performance threshold (we used a threshold value of 0.001 rad s^{-1}). We considered only those values of local excitation above the minimum optimal value or below the maximum optimal value to estimate this tolerance; thus, to estimate the tolerance about the minimum and maximum optimal values, we measured the tolerance in only one direction ($J_E \leq J_{E,2}^*$ or $J_E \geq J_{E,N-2}^*$), and we doubled this value to use as our estimate. We then compared these tolerance estimates to the analytic lower bound given in equation (7), as shown in Fig. 4b,c (also see equations (S113)–(S119) in the Supplementary Note). Finally, we summed these tolerance values (computed numerically or analytically) for each network size N to estimate the net volume of parameter space that meets this threshold level of performance, as shown in Extended Data Fig. 9a.

Robustness to noise. To measure noise robustness, we added independent Gaussian noise with variance σ^2 to each neuron in our optimal networks, and we simulated network dynamics in the absence of velocity input. We ran 10,000 simulations in which we tracked the orientation of the bump over a total time of 20 s, and we used this to measure the variance of the difference between the initial and final bump positions over time: $\langle (\psi(t) - \psi_0)^2 \rangle$. For short timescales, the dynamics of this quantity are affected by the finite integration timescale τ ; at longer timescales, this quantity follows a diffusion equation with diffusion constant D : $\langle (\psi(t) - \psi_0)^2 \rangle = \sigma_0^2 + 2Dt$. We used the bump trajectories for $t > 10$ s to fit a value for $2D$, as shown in Fig. 4d, and we took $1/2D$ as a measure of noise robustness. Figure 4e,f measures this robustness for optimally tuned networks of varying J_E^* and N , and for varying noise levels: $\sigma^2 = (A/6)^2 \times [1, 4, 9, 16, 25]$, where $A = 0.2$ is the bump amplitude. To extract the dependence on N and σ^2 , as shown in Fig. 4f, we found the best-fitting coefficients a, b for the linear relationship $2D = (aN + b)/\sigma^2$ (see Extended Data Fig. 9b for a visualization of these coefficients).

Robustness to changes in the transfer function and recurrent weights. We examined the robustness of the continuous attractor regime to changes in the number of Fourier modes of the recurrent connections in W^{sym} , the neuron input–output relationship ϕ , and an increase in the dimensionality of the attractor. To this aim, we numerically solved the dynamics of equation (1) with $v_{\text{in}} = 0$ in two different scenarios. First, we used (1) a von Mises connectivity profile with concentration parameter κ for the recurrent weights $W_{jk}^{\text{sym}} = J_1 + J_E \exp(\kappa \cos(\theta_j - \theta_k))/(2\pi I_0(\kappa))$, where $I_0(\kappa)$ is the modified Bessel function of order 0; (2) a smooth nonlinear transfer function, $\phi(x) = \log(1 + e^x)$. We numerically solved the dynamics of a network with $N = 8$ units and $J_1 = -30$, with cosine-shaped initial conditions centered at 50 uniformly spaced orientations on the ring (Extended Data Fig. 5a). We evaluated the dispersion (circular variance) between the initial and final orientations on the ring for different values of J_E after numerically solving the dynamics for a total time of 500τ , where τ is the single-neuron time constant. We observed the presence of optimal values of J_E (Extended Data Fig. 5a, red), where the network behaved like a continuous attractor, as opposed to other values of J_E (Extended Data Fig. 5a, purple, blue) where the discreteness of the solution was evident. The specific values of optimal excitation depend on both the value of J_1 (Extended Data Fig. 5a, empty circles), and on the strength of constant feedforward input c_{ff} .

We next examined the dynamics in equation (1) with a recurrent weight profile storing a two-dimensional toroidal attractor with $N = 16$ neurons, $W_{jk}^{\text{sym}} = J_1 + \frac{J_E}{2}(\cos(\theta_j^1 - \theta_k^1) + \cos(\theta_j^2 - \theta_k^2))$, $J_1 = -20$, where the preferred orientations (θ_j^1, θ_j^2) of the units were uniformly spaced on the torus (Extended Data Fig. 5b). We similarly observed the presence

of an optimal value of J_E for which the dispersion between subthreshold bumps initialized at 100 different orientations on the torus and the final orientations were close to 0.

Reporting summary

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

Data availability

All data collected for this study are freely available via figshare at <https://doi.org/10.25378/janelia.26169355> (ref. 66).

Code availability

All custom code written for this study is freely available via Zenodo at <https://doi.org/10.5281/zenodo.12789923> (ref. 67) and is maintained on GitHub at <https://github.com/HermundstadLab/DiscreteRingAttractor>.

References

56. Dana, H. et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).
57. Dionne, H., Hibbard, K. L., Cavallaro, A., Kao, J.-C. & Rubin, G. M. Genetic reagents for making split-GAL4 lines in *Drosophila*. *Genetics* **209**, 31–35 (2018).
58. Seelig, J. D. et al. Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nat. Methods* **7**, 535–540 (2010).
59. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
60. Berens, P. CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* <https://doi.org/10.18637/jss.v031.i10> (2009).
61. Hulse, B. K., Stanoev, A., Turner-Evans, D. B., Seelig, J. D. & Jayaraman, V. A rotational velocity estimate constructed through visuomotor competition updates the fly's neural compass. Preprint at bioRxiv <https://doi.org/10.1101/2023.09.25.559373> (2023).
62. Wolff, T., Iyer, N. A. & Rubin, G. M. Neuroarchitecture and neuroanatomy of the *Drosophila* central complex: a GAL4-based dissection of protocerebral bridge neurons and circuits. *J. Comp. Neurol.* **523**, 997–1037 (2015).
63. Watson, G. S. Goodness-of-fit tests on a circle. II. *Biometrika* **49**, 57–63 (1962).
64. Landler, L., Ruxton, G. D. & Malkemper, E. P. Advice on comparing two independent samples of circular data in biology. *Sci. Rep.* **11**, 20337 (2021).
65. Gajic, Z. & Qureshi, M. T. J. *Lyapunov Matrix Equation in System Stability and Control* (Dover Publications, 2008).
66. Noorman, M., Hulse, B. K., Jayaraman, V., Romani, S. & Hermundstad, A. M. 2P calcium imaging from compass neurons of tethered flies walking on a ball in darkness. *figshare* <https://doi.org/10.25378/janelia.26169355> (2024).
67. Noorman, M., Hulse, B. K., Jayaraman, V., Romani, S. & Hermundstad, A. M. HermundstadLab/DiscreteRingAttractor: v1.0. Zenodo <https://doi.org/10.5281/zenodo.12789923> (2024).

Acknowledgements

We thank J.E. Fitzgerald for useful discussions about characterizing threshold-linear dynamics in terms of linear subspaces and D. Turner-Evans for early conversations about the circuitry of the fly HD system. We are grateful to A. Stanoev for helpful pilot experiments. We thank members of the Hermundstad and Jayaraman laboratories for their useful feedback during

this project. This work was funded by the Howard Hughes Medical Institute.

Author contributions

M.N., S.R. and A.M.H. conceptualized the problem, with input from V.J. B.K.H. performed all experiments and data processing. B.K.H. and V.J. performed data analysis, with input from M.N., S.R. and A.M.H. M.N. performed the bulk of the analytics on the full nonlinear system, with contributions from S.R. and A.M.H. M.N. and A.M.H. performed the analyses on the linear subsystems. M.N., S.R. and A.M.H. performed simulations. M.N. and A.M.H. wrote the paper, with input and editing from all authors.

Competing interests

The authors declare no competing interests.

Additional information

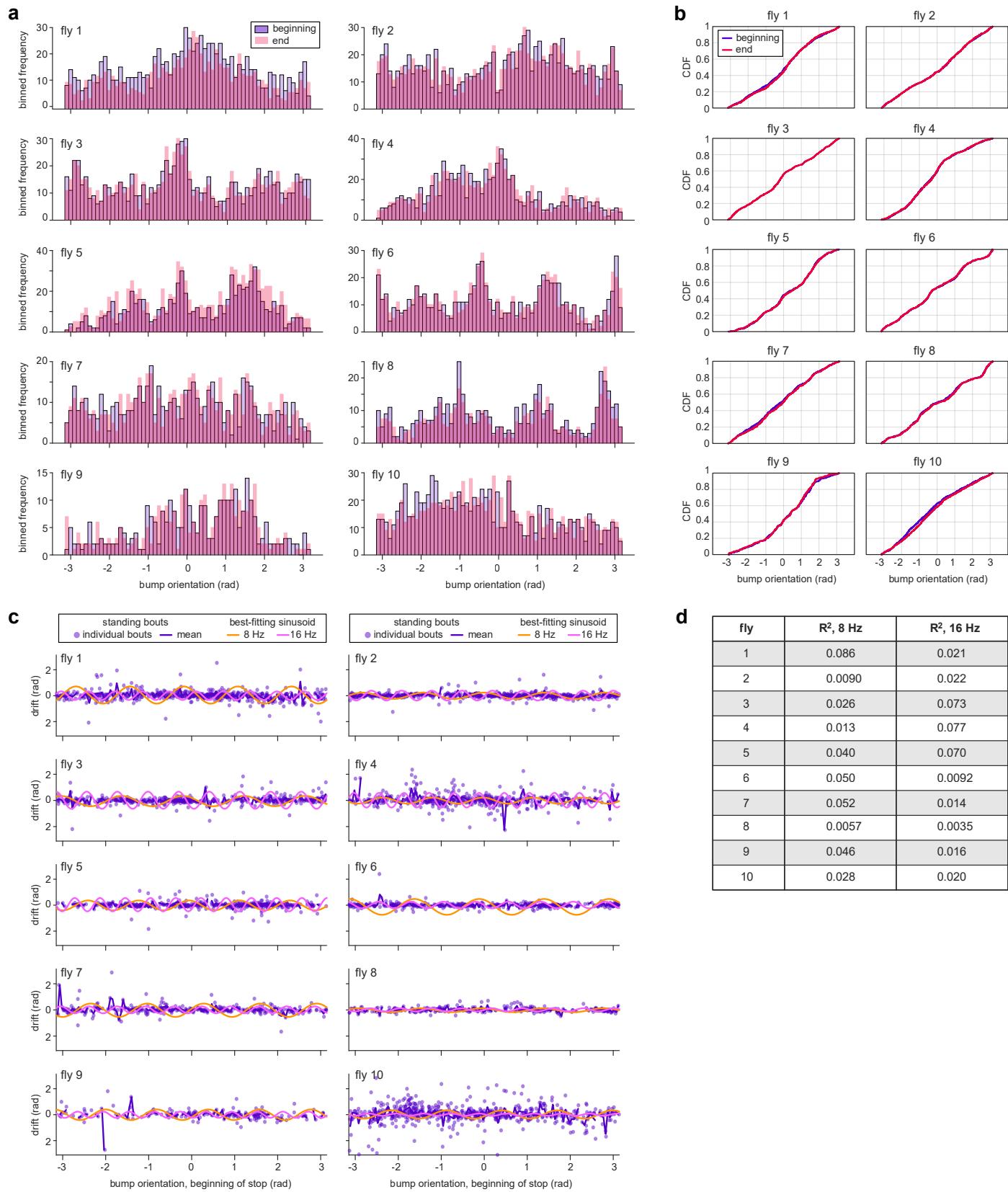
Extended data is available for this paper at <https://doi.org/10.1038/s41593-024-01766-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41593-024-01766-5>.

Correspondence and requests for materials should be addressed to Marcella Noorman or Ann M. Hermundstad.

Peer review information *Nature Neuroscience* thanks Chenglin Miao 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.

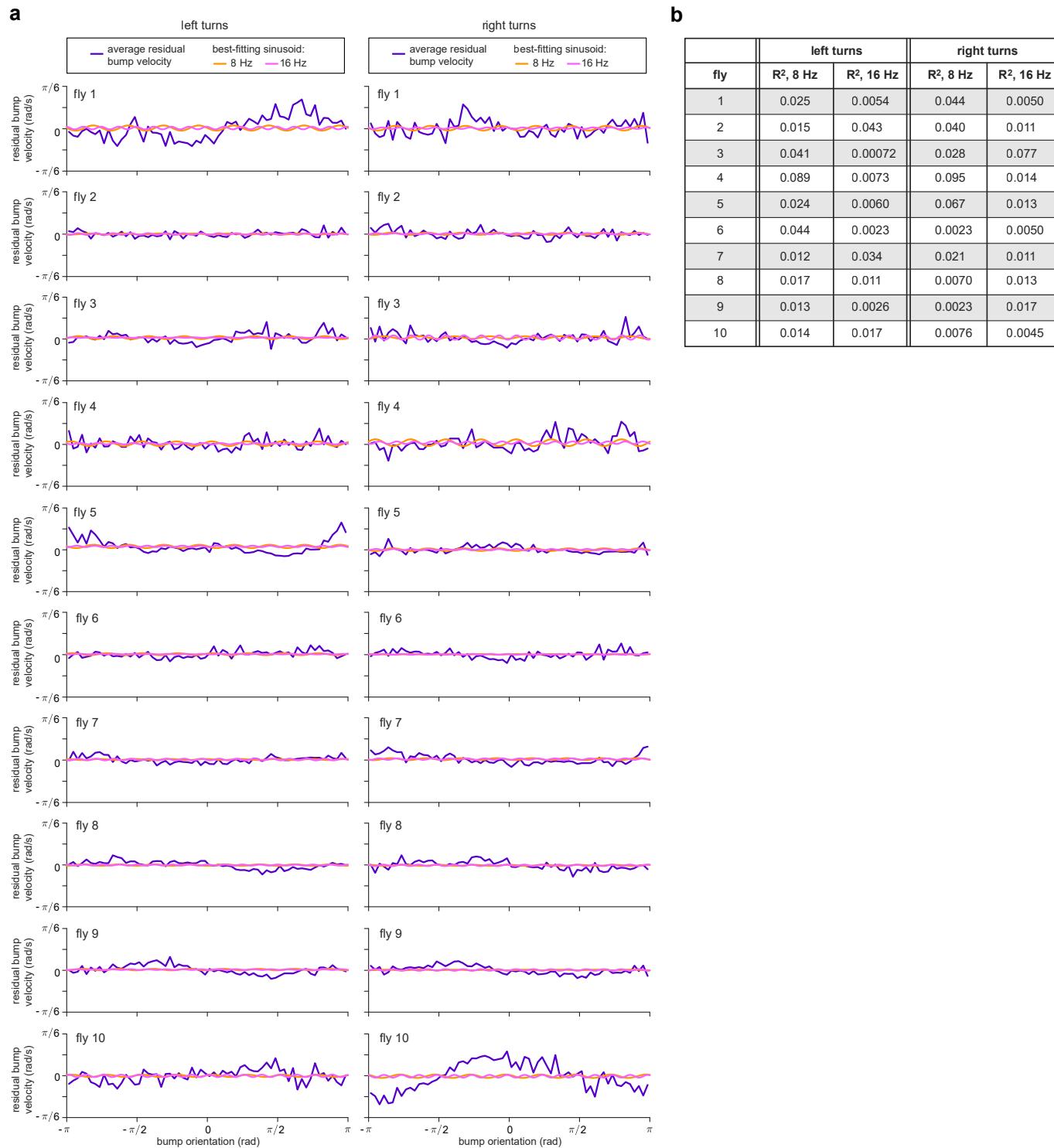


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Analysis of bump drift during standing bouts.

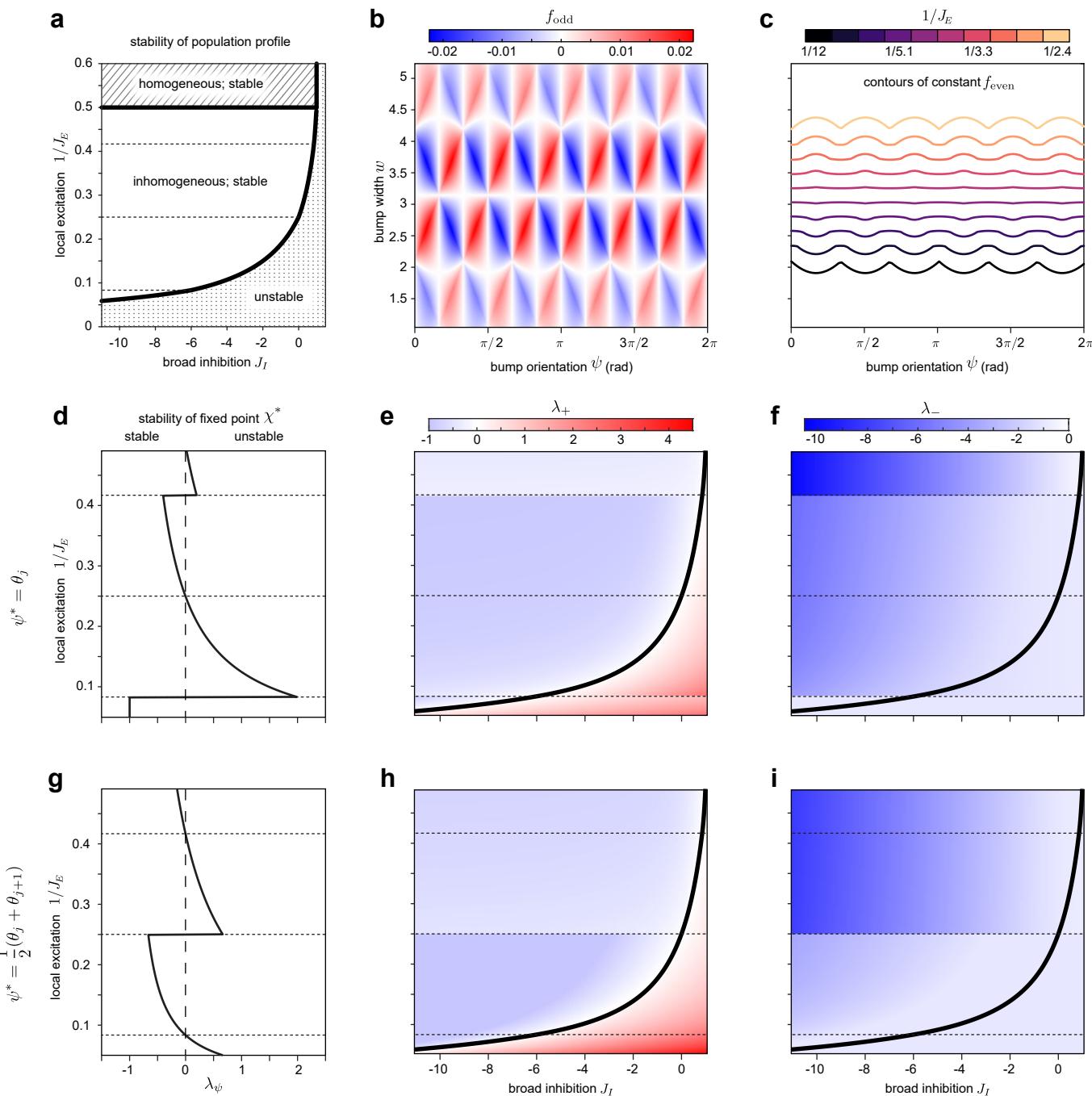
a, Histograms of orientations in the ellipsoid body occupied by the compass bump at the beginning (blue) and end (red) of standing bouts for all ten flies. Note degree of overlap in the distributions, with no sign of an increase in specific orientations from beginning to restart. **b**, Cumulative distributions of orientations in the ellipsoid body occupied by the compass bump at the beginning (blue) and end (red) of standing bouts for all ten flies. Differences between the two distributions are not statistically significant. P-values for Watson's U^2 test (flies 1-10): 0.5560, 1.0000, 1.0000, 0.9920, 0.9980, 1.0000, 0.9580, 1.0000, 0.9860, 0.1180. U^2 test statistic (flies 1-10): 0.0660, 0.0085, 0.0095, 0.0193, 0.0157, 0.0070, 0.0295, 0.0128, 0.0221, 0.1394. **c**, Drift during

standing bouts for all ten flies, measured at different starting orientations of the compass bump. 8- and 16-Hz sinusoids were fit to drifts for each fly. One signature of discreteness in the performance of the compass system would be lower drift when the bump starts at stable orientations during standing bouts and higher drift when the bump starts outside of those orientations. We did not see such fluctuations in the data (see panel **d**). **d**, R^2 values for sinusoidal fits in panel **c**. In panels **a-d**, only those standing bouts that were greater than 0.3 s and less than 2 s were used for analyses. This resulted in the following numbers of standing bouts for flies 1-10: 980, 1005, 835, 826, 723, 714, 573, 527, 312, 949. Flies 2 and 6 correspond to flies GC7fA and GC7fB, respectively, in Fig. 1e,h-j.



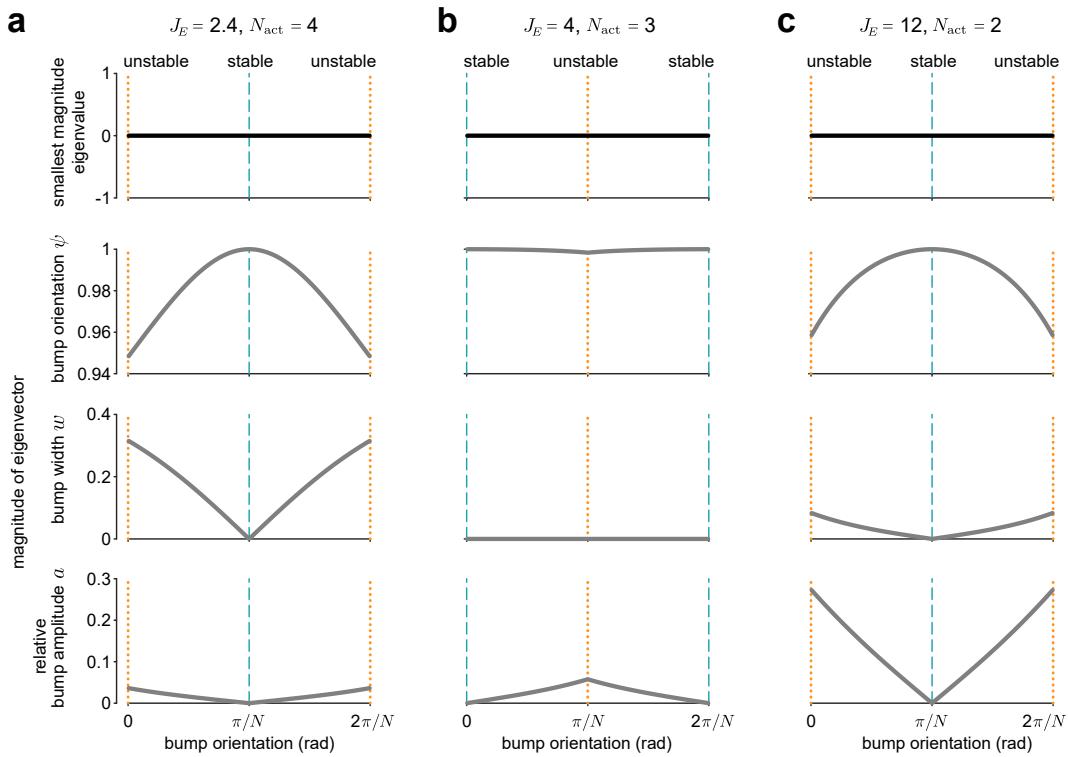
Extended Data Fig. 2 | Analysis of residual bump velocity. **a**, Average residual bump velocities measured at different bump orientations, shown separately for left and right turns for all ten flies. 8- and 16-Hz sinusoids were fit to these average

residuals. One signature of discreteness would be systematically higher or lower residual velocities at specific bump orientations; we did not see such fluctuations in the data (see panel **b**). **b**, R^2 values for sinusoidal fits in panel **a**.



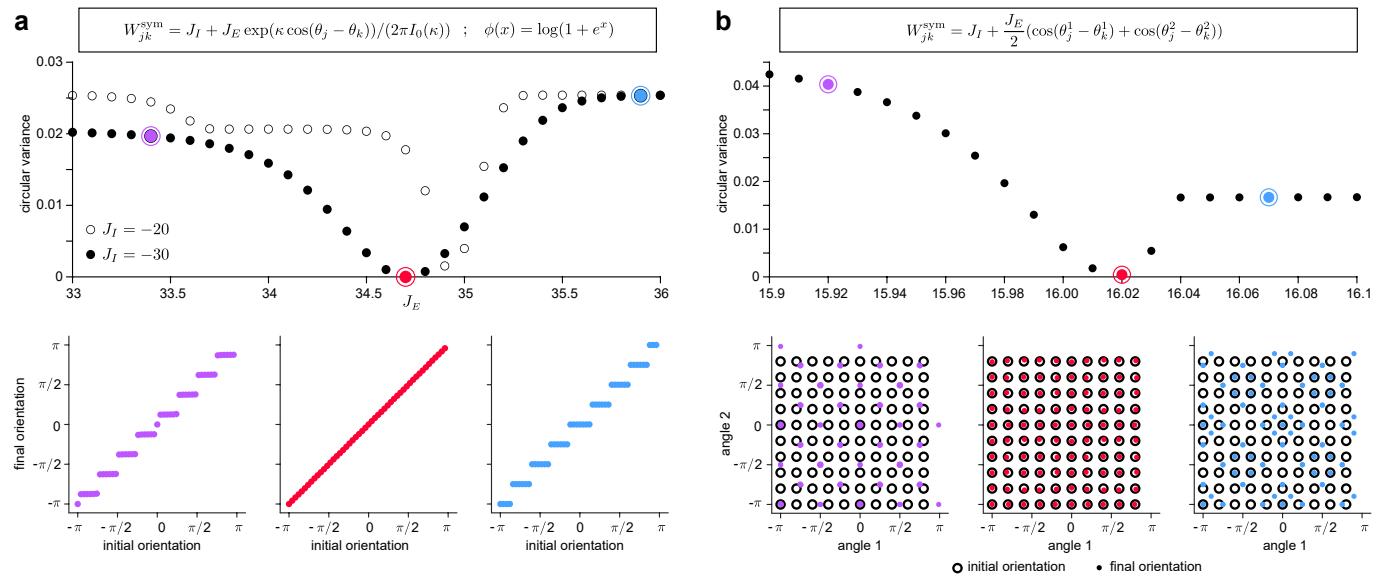
Extended Data Fig. 3 | Stability of population profile and fixed-point analysis. **a**, The stability of the shape of the population profile depends on J_E and J_I (shown for $N=6$). ‘Unstable’ regime: the population activity diverges over time. ‘Homogeneous’ regime: the network generates a stable activity profile that is uniform across the entire network. ‘Inhomogeneous’ regime: the network generates a stable bump of activity that persists at a discrete set of orientations in the absence of input. Dashed lines indicate optimal values of J_E for which the network generates a set of marginally stable solutions that can persist at any orientation in the absence of input. **b-c**, Fixed point conditions from the equations for bump orientation (panel **b**; $f_{\text{odd}}=0$) and relative amplitude (panel **c**; $f_{\text{even}}=1/J_E$). See Supplementary Note for details. **b**, Heatmap of f_{odd} (w, ψ) for densely sampled bump widths $w \in [2\pi/N, 2(N-1)\pi/N]$ and orientations $\psi \in [0, 2\pi]$. Red and blue regions correspond to $f_{\text{odd}} > 0$ and $f_{\text{odd}} < 0$ (which drive the bump orientation to the right and left, respectively). White regions indicate $f_{\text{odd}} = 0$, which correspond to potential fixed points $\chi^* = (a^*, w^*, \psi^*)$ at which the bump can stably persist. Note that at $\psi = (\theta_c + \theta_d)/2$, $d = c, c+1$, $f_{\text{odd}}(w, \psi) = 0$ regardless of the value of w . **c**, Contours of constant $f_{\text{even}}(w, \psi)$, shown for 10 evenly spaced

values of $1/J_E$ between and including $1/12$ and $1/2.4$. These contours indicate a necessary (but not sufficient) relationship between w and ψ for stationary bump solutions. **d-i**, Eigenvalues of linearized system about fixed points with orientation $\psi = \theta_j$ (panels **d-f**) or $\psi = (\theta_j + \theta_{j+1})/2$ (panels **g-i**), $j=1 \dots N$. See Supplementary Note for details. **d, g**, Eigenvalue λ_ψ depends on J_E . This eigenvalue corresponds to changes in orientation near the fixed points and is the sole determinant of stability of the fixed points. Note that when the set of fixed points corresponding to $\psi = \theta_j, j=1 \dots N$, is stable, the other set of fixed points corresponding to $\psi = (\theta_j + \theta_{j+1})/2, j=1 \dots N$, is unstable, and vice-versa. The remaining two eigenvalues λ_+ (panels **e, h**) and λ_- (panels **f, i**) depend on both J_E and J_I but are always negative in the parameter regime that generates bump-like profiles (region above black line; compare to ‘inhomogeneous; stable’ in panel **a**). Panels **b, c, e, f, h, i** were generated using redblueu.m (<https://www.mathworks.com/matlabcentral/fileexchange/74791-redblue-colormap-generator-with-zero-as-white-or-black>) and magma.m (<https://www.mathworks.com/matlabcentral/fileexchange/51986-perceptually-uniform-colormaps>).



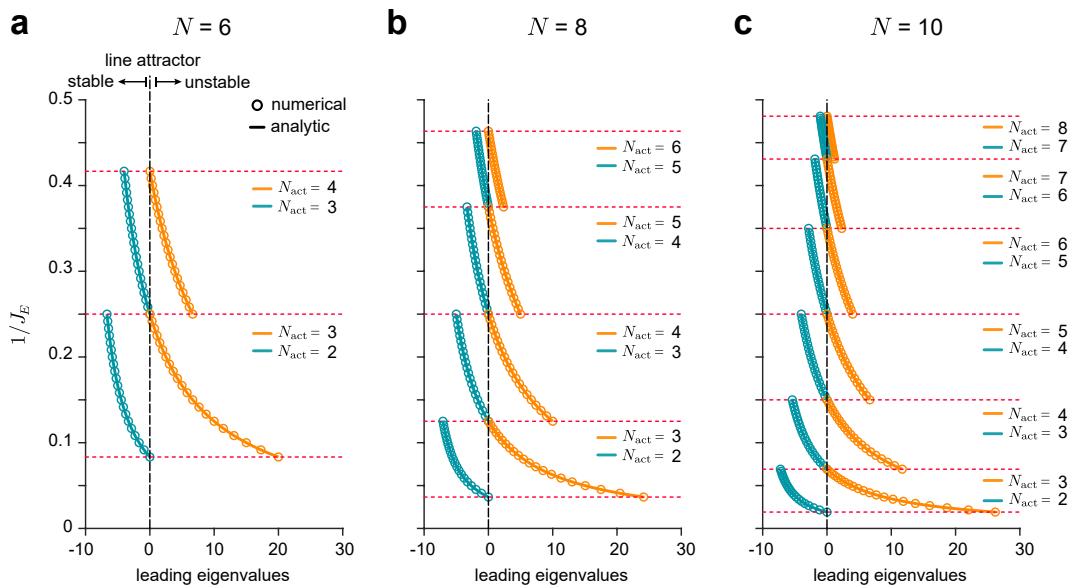
Extended Data Fig. 4 | Flat directions in the energy landscape. Smallest magnitude eigenvalues (top row) and corresponding eigenvector components (lower three rows) for the Hessian matrix of the energy, computed for all three optimal values of local excitation in a network of size $N=6$: **a**, $J_E^*=2.4$; **b**, $J_E^*=4$; **c**, $J_E^*=12$. For each optimal value of local excitation, the Hessian has a single zero

eigenvalue, indicating the existence of a zero-curvature direction within the energy landscape. The corresponding eigenvectors are purely aligned along ψ (second row) at the orientations of the stable fixed points (teal dashed lines). Away from these orientations, the corresponding eigenvectors involve contributions from w and a (third and fourth rows, respectively).



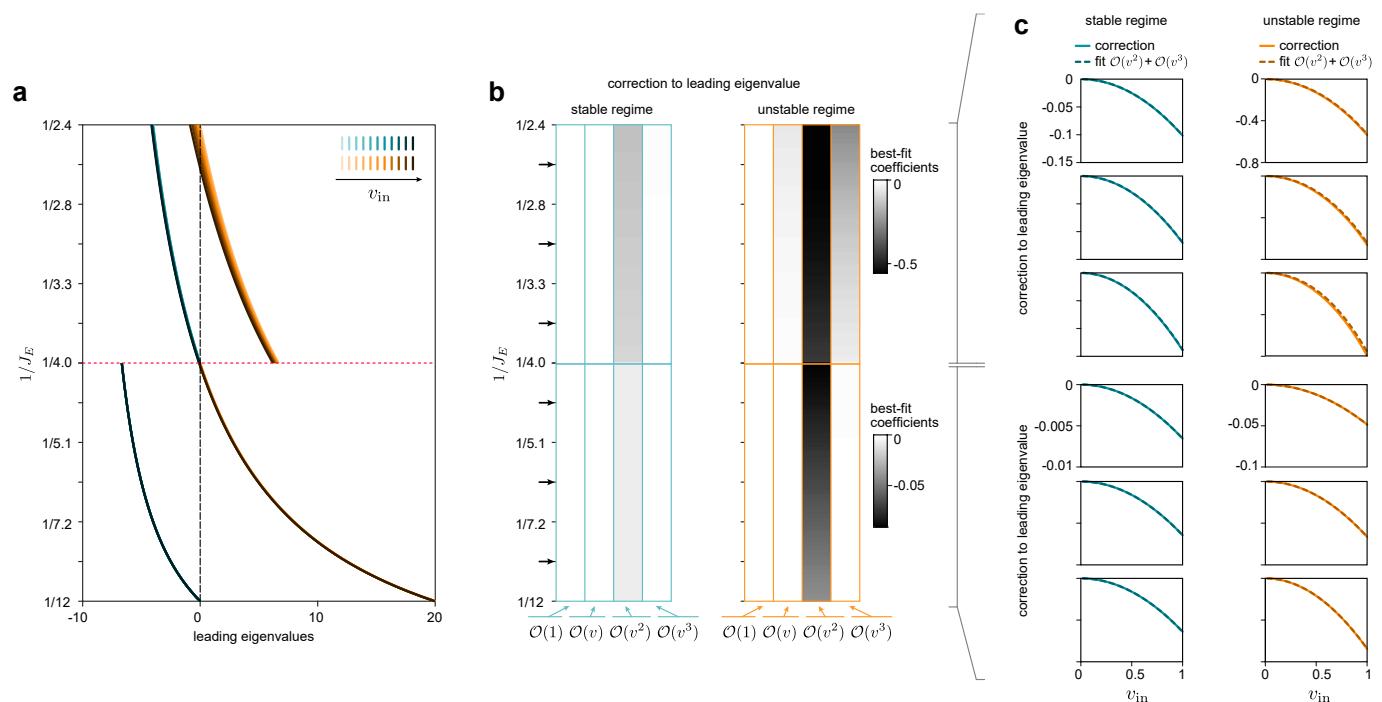
Extended Data Fig. 5 | Robustness to changes in the single neuron transfer function and recurrent synaptic weights. Comparison between initial and final bump orientations as a function of J_E for **a**, a network of $N=8$ neurons with a Von Mises weight profile and a smooth nonlinear transfer function, and **b**, a network of $N=16$ neurons with a recurrent weight profile storing a 2-dimensional toroidal attractor. In both cases, there is an optimal value of J_E for which the circular

variance between the initial and final orientations is close to zero (top, red markers), and the bump does not drift (bottom, center panels). Away from these values of J_E , the circular variance increases (top, purple/blue markers), and the bump drifts from its original orientation (bottom left/right panels). See Methods for simulation details.


Extended Data Fig. 6 | Leading eigenvalues of active submatrices.

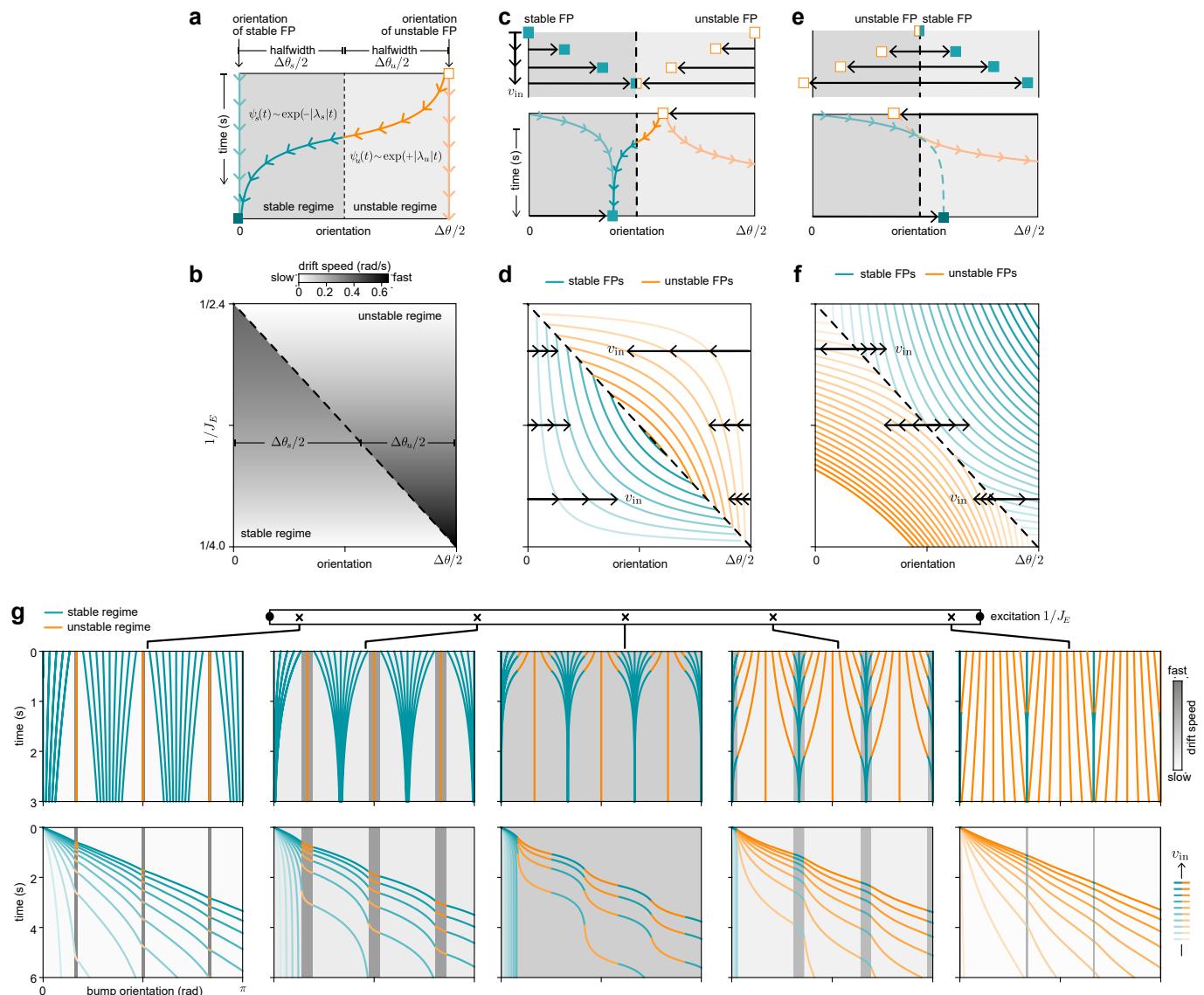
Comparison of analytically- versus numerically- derived eigenvalues (solid lines versus markers, respectively), computed from the active submatrices of the

full connectivity $W = (W^{\text{sym}}/N - I)/\tau$ in the absence of velocity input. Shown for network sizes **a**, $N = 6$, **b**, $N = 8$, and **c**, $N = 10$. Red dotted lines mark optimal values of local excitation for each network size.



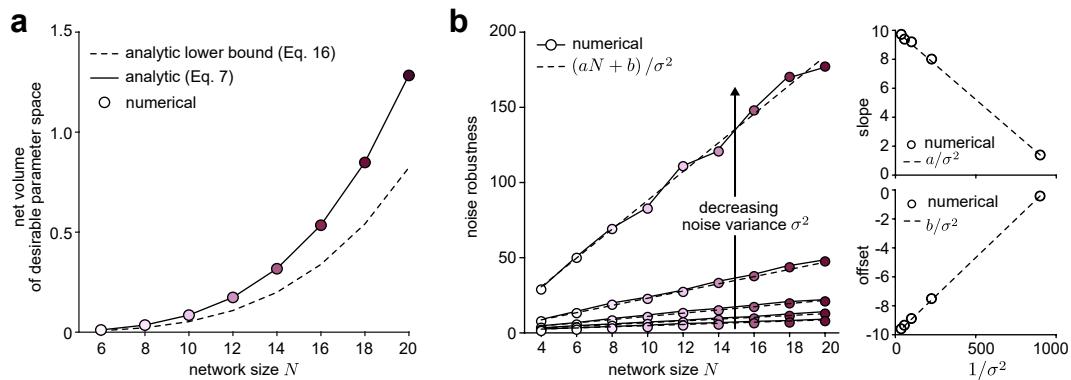
Extended Data Fig. 7 | Velocity correction to leading eigenvalues of active submatrices. **a**, Leading eigenvalues λ of active submatrices as a function of input velocity v_{in} for a network of size $N = 6$. Shown for 11 velocity values evenly spaced between (and including) $v_{in} = 0$ and $v_{in} = 1 \text{ rad s}^{-1}$ (darker colors indicate higher velocities). Eigenvalues were obtained by numerically diagonalizing active submatrices of the full connectivity $W = ((W^{\text{sym}} + v_{in} W^{\text{asym}})/N - I)/\tau$. Red dashed line

marks an optimal value of local excitation. **b**, Coefficients of the best-fitting 3rd order polynomial of the velocity correction $\lambda - \lambda_0$ versus input velocity v_{in} , where λ_0 is the leading eigenvalue of the full connectivity in the absence of velocity input. **c**, Comparison of the velocity correction $\lambda - \lambda_0$ (solid lines) and the best-fitting polynomial (dashed lines), including terms of order $\mathcal{O}(v^2)$ and $\mathcal{O}(v^3)$. Shown for 6 different values of local excitation marked by arrows in panel **b**.



Extended Data Fig. 8 | Impact of stable and unstable fixed points on drift and velocity integration. **a-b**, Repeated from Fig. 3c-d, with regimes colored according to drift speed (grayscale). **c**, Input velocity shifts the orientations of the fixed points. Top: as the input velocity increases from 0, the orientations of the stable and unstable fixed points shift toward the boundary between regimes. At a threshold velocity (Eq. (5)), the two fixed points will meet at the boundary; this threshold velocity is the minimum input velocity needed to move the bump continuously. Bottom: For velocities below this threshold, the bump will be driven to the stable fixed point, regardless of its initial orientation. **d**, The orientations of stable and unstable fixed points (turquoise and orange lines, respectively) shift with increasing velocity (darker shades). The rate of these shifts is set by the drift speeds in the stable and unstable regimes (see panel b): lower drift speeds lead to faster shifts (marked by the large spacing between turquoise lines at the left of the panel, and between orange lines at the right of

the panel). The precise values of these drift speeds ensure that the pair of stable and unstable fixed points will meet at the boundary between regimes at the same threshold velocity, given a fixed value of J_E . **e**, Top: as the input velocity increases above the threshold velocity, the stable and unstable fixed points move beyond their respective regimes. Bottom: When in the stable regime, the bump is pulled from ahead toward the stable fixed point. However, before reaching the stable fixed point, the bump transitions into the unstable regime, and is pushed from behind by the unstable fixed point. This push and pull causes the bump to slow down and speed up as it moves through the stable and unstable regimes, respectively; the closer the fixed points are to the boundary, the stronger this effect. **f**, Above the threshold velocity, the stable and unstable fixed points move beyond their respective regimes, and they continue to shift with velocity at the same rate as shown in panel d. **g**, Example bump trajectories in the absence (top row) and presence (bottom row) of velocity input.


Extended Data Fig. 9 | Analysis of robustness as a function of network size.

a, The net volume of parameter space that achieves a desired performance threshold (estimated by summing the tolerance across all optimal values of local excitation for a given network size N) increases faster than N^2 . Computed analytically via Eq. (7) by summing over all optimal values of local excitation

(solid black line), and estimated numerically by summing over all values shown in Fig. 4b). The analytic lower bound given in Methods Eq. (16) is shown for comparison (gray dashed line). **b**, Left: noise robustness increases linearly with network size. Right: the coefficients of the best linear fit vary inversely with the noise variance σ^2 .

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	All two-photon calcium imaging experiments were performed using a custom microscope controlled with ScanImage (version 2022; Vidrio Technologies). Behavioral data and imaging timestamps were collected using WaveSurfer (version 0.947).
Data analysis	All code was written in Matlab R2020a. The code is freely available at https://doi.org/10.5281/zenodo.12789923 and is maintained at https://github.com/HermundstadLab/DiscreteRingAttractor .

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](#)

Data is freely available at <https://doi.org/10.25378/janelia.26169355>.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

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

Life sciences study design

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

Sample size

Calcium imaging experiments in behaving flies involve tethering, dissections and positioning on the ball. These steps are each subject to some variability. In addition, fly behavior on the ball is variable. These factors, rather than statistical methods, motivated us to perform experiments in 10 flies with long trial durations (5 trials of 20 minutes each). We recorded 100 minutes of data without visual stimuli for each fly, which we assume to be sufficient for the types of analyses we performed. A sample size of 10 flies was chosen to be consistent with previous studies of this circuit (e.g. Seelig and Jayaraman, 2015).

Data exclusions

We excluded any data collected beyond 100 minutes for consistency and to exclude a small number of flies whose behavior and/or imaging degraded in quality, a known limitation of fly-on-a-ball calcium imaging experiments.

Replication

We specifically focused on the peak performance of the internal compass of flies walking in unnatural settings (on a ball), highlighting the top two performers. We additionally report performance across all flies as well. Given the variability and constraints of the experiment discussed above, we do not expect the performance of the internal compass to be perfectly accurate for all flies. In addition, in response to reviewer comments, we collected data from a new set of 10 flies with longer trial durations than were presented to flies featured in our initial submission: the main results replicated across these two different groups.

Randomization

Not relevant. Data was only collected from one genotype of flies.

Blinding

Not relevant. Data was only collected from one genotype of flies.

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	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	All experiments featured <i>Drosophila melanogaster</i> females (age 5-9d) expressing GCaMP6f in EPG neurons.
Wild animals	No wild animals were used in this study.
Reporting on sex	Female flies are larger, making them more suitable for calcium imaging experiments in tethered behaving flies.
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	Studies in <i>Drosophila</i> do not require ethical approval. Treatment of flies was in accordance with standard lab procedures.

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