bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Multiple bumps can enhance robustness to noise in continuous attractor networks

Raymond Wang[1,2] and Louis Kang*[2]

[1]Redwood Center for Theoretical Neuroscience, University of California, Berkeley
[2]Neural Circuits and Computations Unit, RIKEN Center for Brain Science

27 June 2022

Abstract

A central function of continuous attractor networks is encoding coordinates and accurately updating their values through path integration. To do so, these networks produce localized bumps of activity that move coherently in response to velocity inputs. In the brain, continuous attractors are believed to underlie grid cells and head direction cells, which maintain periodic representations of position and orientation, respectively. These representations can be achieved with any number of activity bumps, and the consequences of having more or fewer bumps are unclear. We address this knowledge gap by constructing 1D ring attractor networks with different bump numbers and characterizing their responses to three types of noise: fluctuating inputs, spiking noise, and deviations in connectivity away from ideal attractor configurations. Across all three types, networks with more bumps experience less noise-driven deviations in bump motion. This translates to more robust encodings of linear coordinates, like position, assuming that each neuron represents a fixed length no matter the bump number. Alternatively, we consider encoding a circular coordinate, like orientation, such that the network distance between adjacent bumps always maps onto 360 degrees. Under this mapping, bump number does not significantly affect the amount of error in the coordinate readout. Our simulation results are intuitively explained and quantitatively matched by a unified theory for path integration and noise in multi-bump networks. Thus, to suppress the effects of biologically relevant noise, continuous attractor networks can employ more bumps when encoding linear coordinates; this advantage disappears when encoding circular coordinates. Our findings provide motivation for multiple bumps in the mammalian grid network.

Introduction

Continuous attractor networks (CANs) sustain a set of activity patterns that can be smoothly morphed from one to another along a low-dimensional manifold (Amari, 1977; Ermentrout and Cowan, 1979; Milnor, 1985). Network activity is typically localized into attractor bumps, whose positions along the manifold can represent the value of a continuous variable. These positions can be set by external stimuli, and their persistence serves as a memory of the stimulus value. Certain CAN architectures are also capable of a feature called path integration. Instead of receiving the stimulus value directly, the network receives its changes and integrates over them by synchronously moving the attractor bump (Cannon et al., 1983; McNaughton et al., 1991; Seung, 1996). Path integration allows systems to estimate an external state based on internally perceived changes, which is useful in the absence of ground truth.

Path-integrating CANs have been proposed as a mechanism through which brains encode various physical coordinates. Head direction cells in mammals and compass neurons in insects encode spatial orientation by

*louis.kang@riken.jp

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

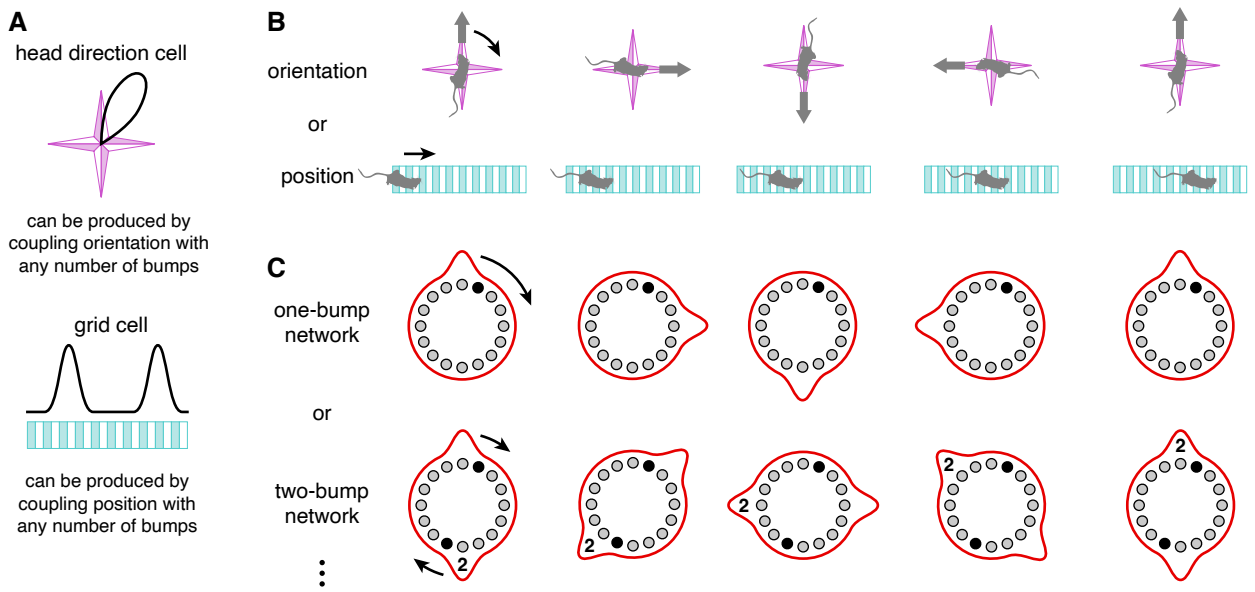

Figure 1: Continuous attractor networks with any number of bumps can produce head direction cells and grid cells. (**A**) Desired tuning curves of a head direction cell and a 1D grid cell. (**B**) Orientation and position coordinates whose changes drive bump motion. (**C**) One- and two-bump ring attractor networks. Each black neuron produces the desired tuning curves in **A**. In the two-bump network, the coupling to coordinate changes is half as strong, and the second bump is labeled for clarity.

preferentially firing when the animal faces a particular direction relative to landmarks (Fig. 1A, top; Taube et al., 1990; Seelig and Jayaraman, 2015). They achieve this as members of 1D CANs whose attractor manifolds have ring topologies (Skaggs et al., 1995; Zhang, 1996). For the case of compass neurons, a ring structure also exists anatomically, and its demonstration of continuous attractor dynamics is well-established (Seelig and Jayaraman, 2015; Kim et al., 2017; Turner-Evans et al., 2017; Green et al., 2017). Grid cells in mammals encode position by preferentially firing at locations that form a triangular lattice in 2D space (1D analogue in Fig. 1A, bottom; Hafting et al., 2005). They are thought to form a 2D CAN with toroidal topology (McNaughton et al., 2006; Fuhs and Touretzky, 2006; Guanella et al., 2007; Burak and Fiete, 2009), and mounting experimental evidence supports this theory (Yoon et al., 2013; Gu et al., 2018; Gardner et al., 2019, 2022). The ability for head direction cells, compass neurons, and grid cells to maintain their tunings in darkness without external cues demonstrates that these CANs can path integrate (Goodridge et al., 1998; Seelig and Jayaraman, 2015; Hafting et al., 2005).

CANs also appear in studies of other brain regions and neural populations. Signatures of continuous attractor dynamics have been detected in the prefrontal cortex during spatial working memory tasks (Constantinidis and Wang, 2004; Edin et al., 2009; Wimmer et al., 2014). Theorists have further invoked CANs to explain place cells (Tsodyks and Sejnowski, 1995; Samsonovich and McNaughton, 1997), hippocampal view cells (Stringer et al., 2005), eye tracking (Cannon et al., 1983; Seung, 1996), visual orientation tuning (Ben-Yishai et al., 1995; Somers et al., 1995), and perceptual decision making (Brody et al., 2003; Machens et al., 2005). Thus, CANs are a crucial circuit motif throughout the brain, and better understanding their performance would provide meaningful insights into neural computation.

One factor that strongly affects the performance of CANs in path integration is biological noise. To accurately represent physical coordinates, attractor bumps must move in precise synchrony with the animal's trajectory. Hence, the bump velocity must remain proportional to the driving input that represents

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coordinate changes (Burak and Fiete, 2009). Different sources of noise produce different types of deviations from this exact relationship, all of which lead to path integration errors. While noisy path-integrating CANs have been previously studied (Zhang, 1996; Stringer et al., 2002; Wu et al., 2008; Burak and Fiete, 2009), these works did not investigate of role of bump number. CANs with different connectivities can produce different numbers of attractor bumps, which are equally spaced throughout the network and perform path integration by moving in unison (Stringer et al., 2004; Fuhs and Touretzky, 2006; Burak and Fiete, 2009). Two networks with different bump numbers have the same representational capability (Fig. 1). They can share the same attractor manifold and produce neurons with identical tuning curves, as long as the coupling strength between bump motion and driving input scales appropriately. The computational advantages of having more or fewer bumps are unknown.

Our aim is to elucidate the relationship between bump number and robustness to noise. We first develop a rigorous theoretical framework for studying 1D CANs that path integrate and contain multiple bumps. Our theory predicts the number, shape, and speed of bumps. We then introduce three forms of noise. The first is Gaussian noise added to the total synaptic input, which can represent fluctuations in a broad range of cellular processes occurring at short timescales. The second is Poisson spiking noise. The third is noise in synaptic connectivity strengths; the ability for bumps to respond readily to driving inputs is generally conferred by a precise network architecture. We add Gaussian noise to the ideal connectivity and evaluate path integration in this setting. The first two forms of noise are independent over time and neurons, in contrast to the third. We find that networks with more bumps can better resist all three forms of noise under certain encoding assumptions. These observations are explained by our theoretical framework with simple scaling arguments. The following Results section presents all simulation findings and major theoretical conclusions; complete theoretical derivations are found in the Theoretical model section.

Results

Bump formation in a ring attractor network

We study a 1D ring attractor network that extends the model of Xie et al. (2002) to allow for multiple attractor bumps. It contains two neural populations $\alpha \in \{\mathrm{L}, \mathrm{R}\}$ at each network position x, with N neurons in each population (Fig. 2A). Each neuron is described by its total synaptic input g that obeys the following dynamics:

$$\tau \frac{\mathrm{d}g_\alpha(x,t)}{\mathrm{d}t} + g_\alpha(x,t) = \sum_\beta \int \mathrm{d}y\, W_\beta(x,y) s_\beta(y,t) + A \pm_\alpha \gamma b(t) + \zeta_\alpha(x,t), \tag{1}$$

where $\pm_\mathrm{L}$ means $-$ and $\pm_\mathrm{R}$ means $+$. Aside from spiking simulations, firing rates s are given by

$$s_\alpha(x,t) = \phi[g_\alpha(x,t)], \tag{2}$$

where ϕ is a nonlinear activation function. For all simulations in this Results section, we take ϕ to be the rectified linear unit (ReLU) activation function (Eq. 35). Our theoretical formulas for diffusion coefficients and velocities in this section also assume a ReLU ϕ. In the Appendix, we consider a logistic ϕ instead and find that all major conclusions are preserved (Fig. 9), and in the Theoretical methods section, we derive most expressions for general ϕ. W is the synaptic connectivity and only depends on the presynaptic population β. It obeys a standard continuous attractor architecture based on local inhibition that is strongest at an inhibition distance l. Each population has its synaptic outputs shifted by a small distance $\xi \ll l$ in opposite

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

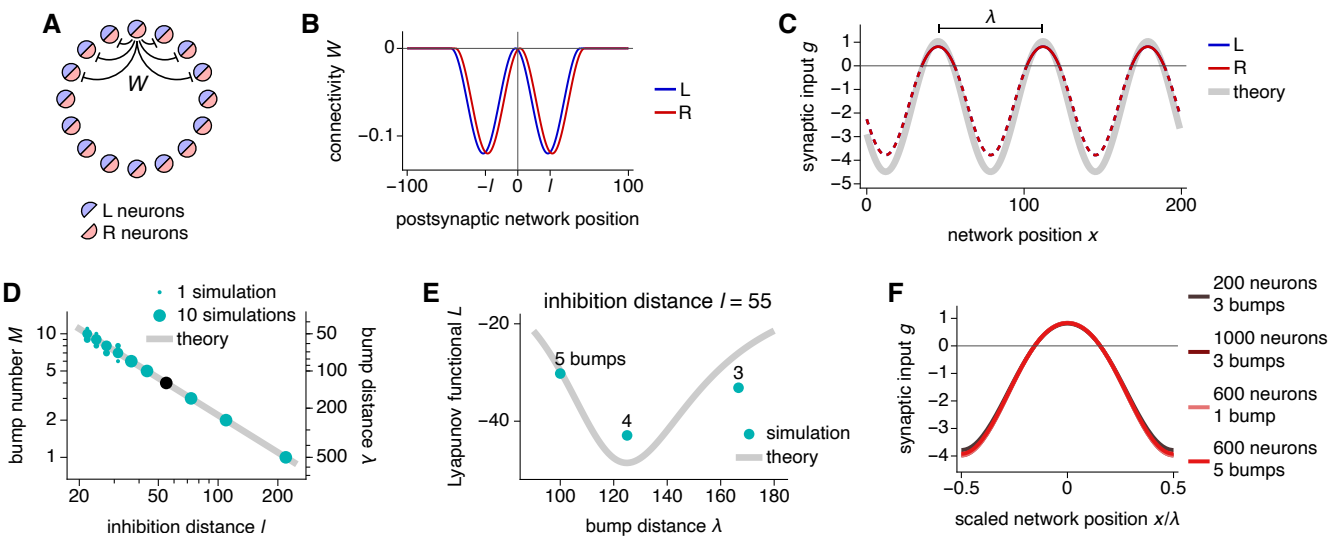

Figure 2: Bump formation in a ring attractor network. (**A**) Network schematic with populations L and R and locally inhibitory connectivity W. (**B, C**) Networks with 200 neurons and 3 bumps. (**B**) Connectivity weights for a neuron at the origin. The inhibition distance is $l = 29$ and the connectivity shift is $\xi = 2$. (**C**) Steady-state synaptic inputs. Curves for both populations lie over each other. With a ReLU activation function, the firing rates follow the solid portions of the colored lines and are 0 over the dashed portions. The bump distance is $\lambda = 200/3$. Thick gray line indicates Eq. 4. (**D, E**) Networks with 500 neurons. (**D**) More bumps and shorter bump distances are produced by smaller inhibition distances. Points indicate data from 10 replicate simulations. Line indicates Eq. 5. (**E**) The inhibition distance $l = 55$ corresponds to the black point in **D** with $\lambda = 125$ and $M = 4$. These values also minimize the Lyapunov functional (Eq. 6), which varies smoothly across λ for infinite networks (line) and takes discrete values for finite networks (points). (**F**) The scaled bump shape remains invariant across network sizes and bump numbers, accomplished by rescaling connectivity strengths according to Eq. 7. Curves for different parameters lie over one another.

directions. We use the connectivity profile described in Fig. 2B and Eq. 38 for all simulations, but all theoretical expressions in this Results section are valid for any W. A is the resting input to all neurons. The driving input, or drive, b is proportional to changes in the coordinate encoded by the network; for the physical coordinates in Fig. 1B, it represents the animal's velocity obtained from self-motion cues. In our results, b is constant in time. It is coupled to the network with strength γ. We will consider various forms of noise ζ. Finally, τ is the neural time constant.

With no drive $b = 0$ and no noise $\zeta = 0$, the network dynamics in Eqs. 1 and 2 can be simplified to

$$\tau \frac{\mathrm{d}g(x,t)}{\mathrm{d}t} + g(x,t) = 2 \int \mathrm{d}y\, W(x-y)\phi[g(y,t)] + A, \tag{3}$$

where $2W(x-y) = \sum_\beta W_\beta(x,y)$ and the synaptic inputs g are equal between the two populations. This baseline equation evolves towards a periodic steady-state g with approximate form (see also Widloski, 2015)

$$g(x) = a \cos \frac{2\pi(x - x_0)}{\lambda} + d. \tag{4}$$

Expressions for a and d are given in the Theoretical model section (Eq. 60). The firing rates $s(x) = \phi[g(x)]$ exhibit attractor bumps with periodicity λ, a free parameter that we call the bump distance (Fig. 2C). x_0 is the arbitrary position of one of the bumps. It parameterizes the attractor manifold with each value corresponding to a different attractor state up to λ.

The bump number $M = N/\lambda$ is determined through λ. It can be predicted by the fastest-growing mode

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in a linearized version of the dynamics (Eq. 43; Sorscher et al., 2019; Khona et al., 2022). The mode with wavenumber q and corresponding wavelength $2\pi/q$ grows at rate $(2\tilde{W}(q) - 1)/\tau$, where $\tilde{W}(q)$ is the Fourier transform of $W(x)$. Thus,

$$\frac{2\pi}{\lambda} = \underset{q}{\operatorname{argmax}}\, \tilde{W}(q). \tag{5}$$

Figure 2D shows that simulations follow the predicted λ and M over various inhibition distances l. Occasionally for small l, a different mode with a slightly different wavelength will grow quickly enough to dominate the network. A periodic network enforces an integer bump number, which discretizes the allowable wavelengths and prevents changes in λ and M once they are established. In an aperiodic or infinite system, the wavelength can smoothly vary from an initial value to a preferred length over the course of a simulation (Burak and Fiete, 2009; Kang and Balasubramanian, 2019). To determine this preferred λ theoretically, we notice that the nonlinear dynamics in Eq. 3 obey the Lyapunov functional

$$L = -\iint \mathrm{d}x\, \mathrm{d}y\, W(x-y)s(x)s(y) + \int \mathrm{d}x \int_0^{s(x)} \mathrm{d}\rho\, \phi^{-1}[\rho] - A \int \mathrm{d}x\, s(x). \tag{6}$$

In the Theoretical model section, we find for ReLU ϕ that L is minimized when $q = 2\pi/\lambda$ maximizes $\tilde{W}(q)$ (Eq. 66). This is the same condition as for the fastest-growing mode in Eq. 5 (Fig. 2E). In other words, the wavelength λ most likely to be established in a periodic network is the preferred bump distance in an aperiodic or infinite system, up to a difference of one fewer or extra bump due to discretization.

We now understand how to produce different bump numbers M in networks of different sizes N by adjusting the inhibition distance l. To compare networks across different values of M and N, we scale the connectivity strength W according to

$$W_\beta(x,y) \propto \frac{M}{N}. \tag{7}$$

This keeps the total connectivity strength per neuron $\int \mathrm{d}x\, W_\beta(x,y)$ constant over M and N. In doing so, the shape of each attractor bump as a function of scaled network position x/λ remains invariant (Fig. 2F). Thus, Eq. 7 isolates our comparisons across M and N to those variables themselves and removes any influence of bump shape. In the Appendix, we consider the alternative without this scaling and find that many major results are preserved (Fig. 10).

Bump dynamics: path integration and diffusion

The drive b produces coherent bump motion by creating an imbalance between the two neural populations. A positive b increases input to the R population and decreases input to the L population (Fig. 3A). Because the synaptic outputs of the former are shifted to the right, the bump moves in that direction. Similarly, a negative b produces leftward bump motion. The bump velocity v_{drive} can be calculated in terms of the baseline firing rates $s(x)$ obtained without drive and noise (see also Xie et al., 2002; Mosheiff and Burak, 2019):

$$v_{\text{drive}} = -\frac{\gamma b \xi \int \mathrm{d}x\, \frac{\mathrm{d}^2 s}{\mathrm{d}x^2}}{\tau \int \mathrm{d}x \left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2}. \tag{8}$$

As a note, these integrals, as well as subsequent ones, do not include the singular points at the edges of attractor bumps. Equation 8 states that bump velocity is proportional to drive b and connectivity shift ξ,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

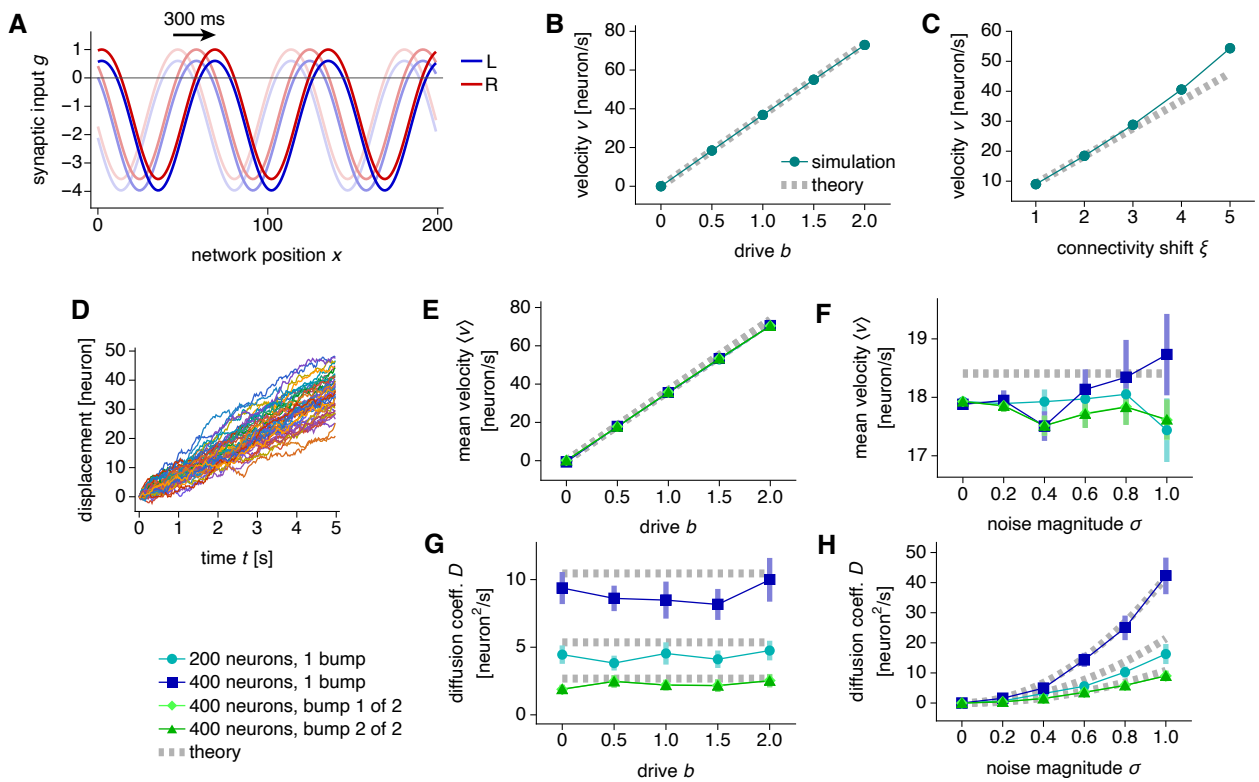

Figure 3: Dynamics in a ring attractor network. (**A–C**) Networks with 200 neurons and 3 bumps. (**A**) Synaptic inputs for populations L and R under drive $b = 2$. Snapshots taken at $150\,\mathrm{ms}$ intervals demonstrate rightward motion. (**B**) Bump velocity is proportional to drive. The connectivity shift is $\xi = 2$. (**C**) Bump velocity is largely proportional to connectivity shift. The drive is $b = 0.5$. (**D–H**) Networks with synaptic input noise. (**D**) Bump displacements for 48 replicate simulations demonstrating diffusion with respect to coherent motion. Networks with 200 neurons and 1 bump. (**E, F**) Mean bump velocity is proportional to drive and remains largely independent of network size, bump number, and noise magnitude. (**G, H**) Bump diffusion coefficient scales quadratically with noise magnitude, remains largely independent of drive, and varies with network size and bump number. The noise magnitude is $\sigma = 0.5$ in **D**, **E**, and **G**, and the drive is $b = 0.5$ in **D**, **F**, and **H**. Values for both bumps in two-bump networks lie over each other. Points indicate data from 48 replicate simulations and bars indicate bootstrapped standard deviations. Dotted gray lines indicate Eqs. 8 and 10.

which is reflected in our simulations, with some deviation at larger ξ (Fig. 3B, C). The strict proportionality between v and b is crucial because it implies faithful path integration (Burak and Fiete, 2009). If $b(t)$ represents coordinate changes (such as angular or linear velocity in Fig. 1B), then the bump position $\theta(t)$ will accurately track the coordinate itself (orientation or position).

In contrast to drive, uncorrelated noise ζ produces bump diffusion. To illustrate this effect, we introduce one form of ζ that we call synaptic input noise. Suppose ζ is independently sampled for each neuron at each simulation timestep from a Gaussian distribution with mean 0 and variance σ^2. Loosely, it can arise from applying the central limit theorem to the multitude of noisy synaptic and cellular processes occurring at short timescales. Then,

$$\langle \zeta_\alpha(x,t) \rangle = 0, \qquad \langle \zeta_\alpha(x,t)\zeta_\beta(y,t') \rangle = \sigma^2 \Delta t\, \delta(t-t')\delta_{\alpha\beta}\delta(x-y), \tag{9}$$

where the timestep Δt sets the resampling rate of ζ, and angle brackets indicate averaging over an ensemble of replicate simulations. Input noise causes bumps to diffuse away from the coherent driven motion (Fig. 3D).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The mean velocity $\langle v \rangle$ remains proportional to drive b, which means that the network still path integrates on average (Fig. 3E). Since $\langle v \rangle$ is largely independent of noise magnitude σ, and the bump diffusion coefficient D is largely independent of b, drive and input noise do not significantly interact within the explored parameter range (Fig. 3F, G). D can be calculated in terms of the baseline firing rates (see also Wu et al., 2008; Burak and Fiete, 2012):

$$D_{\text{input}} = \frac{\sigma^2 \Delta t}{4\tau^2 \int \mathrm{d}x \left(\frac{\mathrm{d}s}{\mathrm{d}x} \right)^2}. \tag{10}$$

The quadratic dependence of D on σ is confirmed by simulation (Fig. 3H).

We now turn our attention to bump number M and network size N. The mean bump velocity $\langle v \rangle$ is independent of these parameters (Fig. 3E, F), which can be understood theoretically. Bump shapes across M and N are simple rescalings of one another (Fig. 2F), so derivatives of s with respect to x are simply proportional to M (more bumps imply faster changes) and inversely proportional to N (larger networks imply slower changes). Similarly, integrals of expressions containing s over x are simply proportional to N. In summary,

$$\frac{\mathrm{d}s}{\mathrm{d}x} \propto \frac{M}{N}, \qquad \frac{\mathrm{d}^2 s}{\mathrm{d}x^2} \propto \frac{M^2}{N^2}, \qquad \int \mathrm{d}x \propto N. \tag{11}$$

Applying these scalings to Eq. 8, we indeed expect v_{drive} to be independent of M and N. In contrast, Fig. 3G, H reveals that the diffusion coefficient D varies with these parameters. When a one-bump network is increased in size from 200 to 400 neurons, D increases as well, which implies greater path integration errors. This undesired effect can be counteracted by increasing the bump number from 1 to 2, which lowers D below that of the one-bump network with 200 neurons. These initial results suggest that bump number and network size are important factors in determining a CAN's resilience to noise. We will explore this idea in greater detail.

Mapping network coordinates onto physical coordinates

Before further comparing networks with different bump numbers M and sizes N, we should scrutinize the relationship between bump motion and the physical coordinate encoded by the network. After all, the latter is typically more important in biological settings. First, we consider the trivial case in which each neuron represents a fixed physical interval across all M and N; this is equivalent to using network coordinates without a physical mapping (Fig. 4A). It is suited for encoding linear variables like position that lack intrinsic periodicity, so larger networks can encode wider coordinate ranges. However, with more bumps or fewer neurons, the range over which the network can uniquely encode different coordinates is shortened. We assume that ambiguity among coordinates encoded by each bump can be resolved by additional cues, such as local features, that identify the true value among the possibilities (O'Keefe and Burgess, 2005; Sreenivasan and Fiete, 2011; Stemmler et al., 2015); this process will be examined in detail below. We leave quantities with dimensions of network distance in natural units of neurons.

Multi-bump networks are intrinsically periodic, especially those with a ring architecture. A natural way for them to encode a circular coordinate like orientation would be to match network and physical periodicities. For example, the bump distance may always represent 360° across different M and N so that neurons always exhibit unimodal tuning (Fig. 4B). This relationship implies that quantities with dimensions of network

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

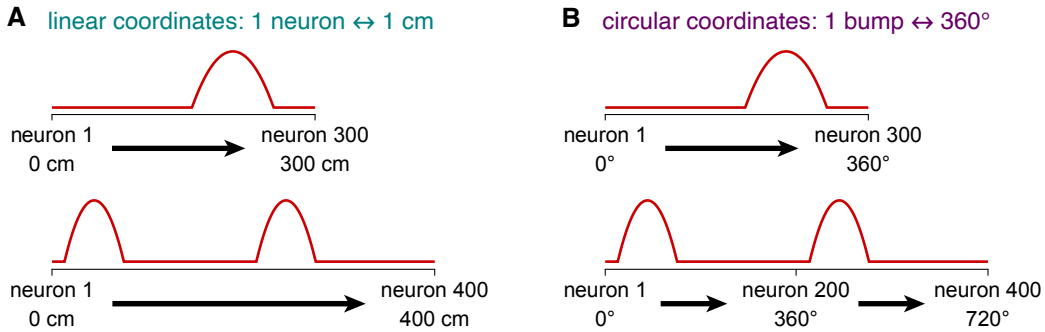

Figure 4: Possible mappings between network coordinates and two types of physical coordinates. (**A**) In networks encoding linear coordinates such as position, one neuron always represents a fixed physical interval. This mapping is trivial and identical to using network coordinates. (**B**) In networks encoding circular coordinates such as orientation, the bump distance always represents $360°$.

distance should be multiplied by powers of the conversion factor

$$\frac{360° \cdot M}{N}, \tag{12}$$

which converts units of neurons to degrees.

For circular mapping, we must also ensure that networks with different bump numbers M and sizes N path integrate consistently with one another. The same drive b should produce the same bump velocity v in units of degree/s. To do so, we rescale the coupling strength γ only under circular mapping:

$$\gamma \propto \frac{N}{M}. \tag{13}$$

This effectively compensates for the factor of M/N in Eq. 12. To see this explicitly, recall that v_{drive} does not depend on M and N in units of neuron/s, as shown in Fig. 3E, F and previously explained through scaling arguments. Under circular mapping, v_{drive} would be multiplied by one power of the conversion factor in Eq. 12. Since its formula contains γ in the numerator (Eq. 8), v_{drive} receives an additional power of the rescaling factor in Eq. 13. The two factors cancel each other, so v_{drive} does not depend on M and N under either mapping:

$$v_{\text{drive}} \propto 1 \quad \text{linear}, \qquad v_{\text{drive}} \propto 1 \quad \text{circular}. \tag{14}$$

Thus, a consistent relationship between b and v_{drive} is preserved in units of both neurons/s and degrees/s.

Of course, there are other possible mappings between network and physical coordinates across bump numbers and network sizes, but for the rest of our paper, we will consider these two. To be clear, networks with the same ring architecture are used for both linear and circular mappings. We will see how noise affects encoding quality in either case.

More bumps improve robustness to input and spiking noise under linear mapping

We now revisit the effect of input noise on bump diffusion, as explored in Fig. 3D–H. We measure how the diffusion coefficient D varies with bump number M and network size N under linear and circular mappings. Under linear mapping, D decreases as a function of M but increases as a function of N (Fig. 5A, B). Thus, more bumps attenuate diffusion produced by input noise, which is especially prominent in large networks.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

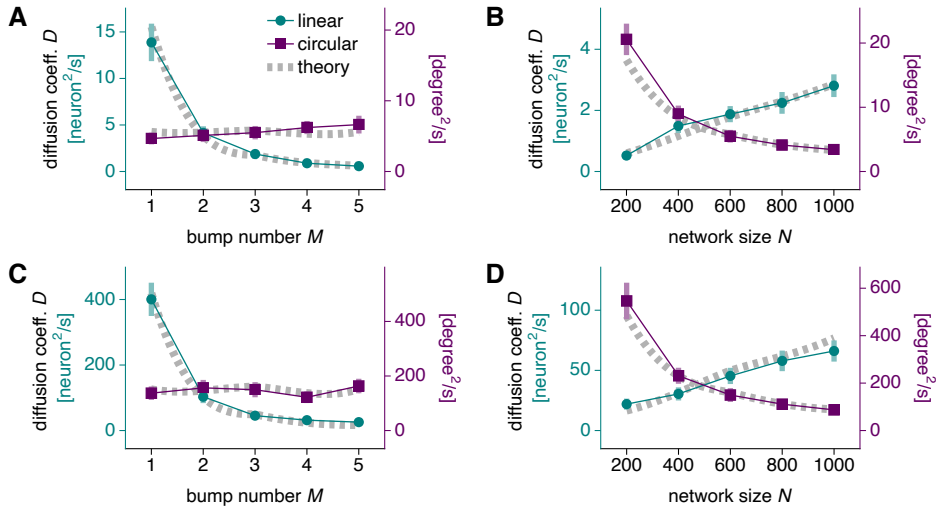

Figure 5: Bump diffusion due to input and spiking noise. (**A, B**) Networks with synaptic input noise of magnitude $\sigma = 0.5$ and drive $b = 0.5$. Dotted gray lines indicate Eq. 10. (**A**) Diffusion decreases with bump number under linear mapping and remains largely constant under circular mapping. Networks with 600 neurons. (**B**) Diffusion increases with network size under linear mapping and decreases under circular mapping. Networks with 3 bumps. (**C, D**) Same as **A**, **B**, but for networks with Poisson spiking noise instead of input noise. Dotted gray lines indicate Eq. 20. Points indicate data from 48 replicate simulations and bars indicate bootstrapped standard deviations.

However, for circular coordinates, D remains largely constant with respect to M and decreases with respect to N (Fig. 5A, B). Increasing the number of bumps provides no benefit. These results can be understood through Eqs. 10, 11, and 12, which predict

$$D_{\text{input}} \propto \frac{N}{M^2} \quad \text{linear}, \qquad D_{\text{input}} \propto \frac{1}{N} \quad \text{circular}. \tag{15}$$

Two powers of the conversion factor in Eq. 12 account for the differences between the two mappings.

Next, we investigate networks with spiking noise instead of input noise. To do so, we replace the deterministic formula for firing rate in Eq. 2 with

$$s_\alpha(x,t) = \frac{c_\alpha(x,t)}{\Delta t}. \tag{16}$$

Here, s is a stochastic, instantaneous firing rate given by the number of spikes c emitted in a simulation timestep divided by the timestep duration Δt. We take the c's to be independent Poisson random variables driven by the deterministic firing rate:

$$c_\alpha(x,t) \sim \text{Pois}\big[\phi[g_\alpha(x,t)]\Delta t\big]. \tag{17}$$

As fully explained in the Theoretical model section (Eq. 99), we can approximate this spiking process by the rate-based dynamics in Eqs. 1 and 2 with the noise term

$$\zeta_\alpha(x,t) = \sum_\beta \int \mathrm{d}y\, W_\beta(x,y) \sqrt{\frac{\phi[g_\beta(y,t)]}{\Delta t}}\, \eta_\beta(y,t). \tag{18}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The η's are independent random variables with zero mean and unit variance:

$$\langle \eta_\alpha(x,t)\rangle = 0, \qquad \langle \eta_\alpha(x,t)\eta_\beta(y,t')\rangle = \Delta t\, \delta(t-t')\delta_{\alpha\beta}\delta(x-y). \tag{19}$$

As for Eq. 9, the simulation timestep Δt sets the rate at which η is resampled. This spiking noise produces bump diffusion with coefficient (see also Burak and Fiete, 2012)

$$D_{\text{spike}} = \frac{\displaystyle\int \mathrm{d}x\, s(x)\left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2}{4\tau^2\left[\displaystyle\int \mathrm{d}x\left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2\right]^2}. \tag{20}$$

As before, s is the baseline firing rate configuration without noise and drive. Through the relationships in Eqs. 11 and 12, D_{spike} scales with M and N in the same way as D_{input} does:

$$D_{\text{spike}} \propto \frac{N}{M^2} \quad \text{linear}, \qquad D_{\text{spike}} \propto \frac{1}{N} \quad \text{circular}. \tag{21}$$

These findings are presented in Fig. 5C, D along with simulation results that confirm our theory. Spiking noise behaves similarly to input noise. Increasing bump number improves robustness to noise under linear mapping but has almost no effect under circular mapping. Bump diffusion in larger networks is exacerbated under linear mapping but suppressed under circular mapping. For both input noise and spiking noise, the conversion factor in Eq. 12 produces the differences between the two mappings. Coupling strength rescaling in Eq. 13 does not play a role because γ does not appear in Eqs. 10 and 20.

To evaluate noise robustness a different way, we perform mutual information analysis of networks with input noise. Mutual information describes how knowledge of one random variable can reduce the uncertainty in another, and it serves as a general metric for encoding quality. Here, we calculate the mutual information I between the physical coordinate encoded by the noisy network, represented by the random variable U with discretized sample space $\mathcal{U}$, and the activity of a single neuron at network position x, represented by the random variable S with discretized sample space $\mathcal{S}$ (see Simulation methods):

$$I[S;U] = \sum_{s\in\mathcal{S},u\in\mathcal{U}} p(s|u)p(u)\log\frac{p(s|u)}{p(s)}. \tag{22}$$

We then average I across neurons. Larger mean mutual information implies more precise coordinate encoding and greater robustness to noise. Thus, we expect that networks with less diffusion in Fig. 5A, B should generally contain more information. Note that the joint activities of all the neurons confer much more coordinate information than single-neuron activities do, but since estimating high-dimensional probability distributions over the former is computationally very costly, we use the latter as a metric for network performance.

The physical coordinate U is either position or orientation and obeys the mappings described in Fig. 4 across bump numbers M and network sizes N. To obtain the probability distributions in Eq. 22 required to calculate I, we initialize multiple replicate simulations at evenly spaced coordinate values u (Fig. 6A). We do not apply an input drive, so the networks should still encode their initialized coordinates at the end of the simulation. However, they contain input noise that degrades their encoding. Collecting the final firing rates $s(x)$ produces $p(s|u)$ for each neuron x. For both position and orientation, we consider narrow and

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

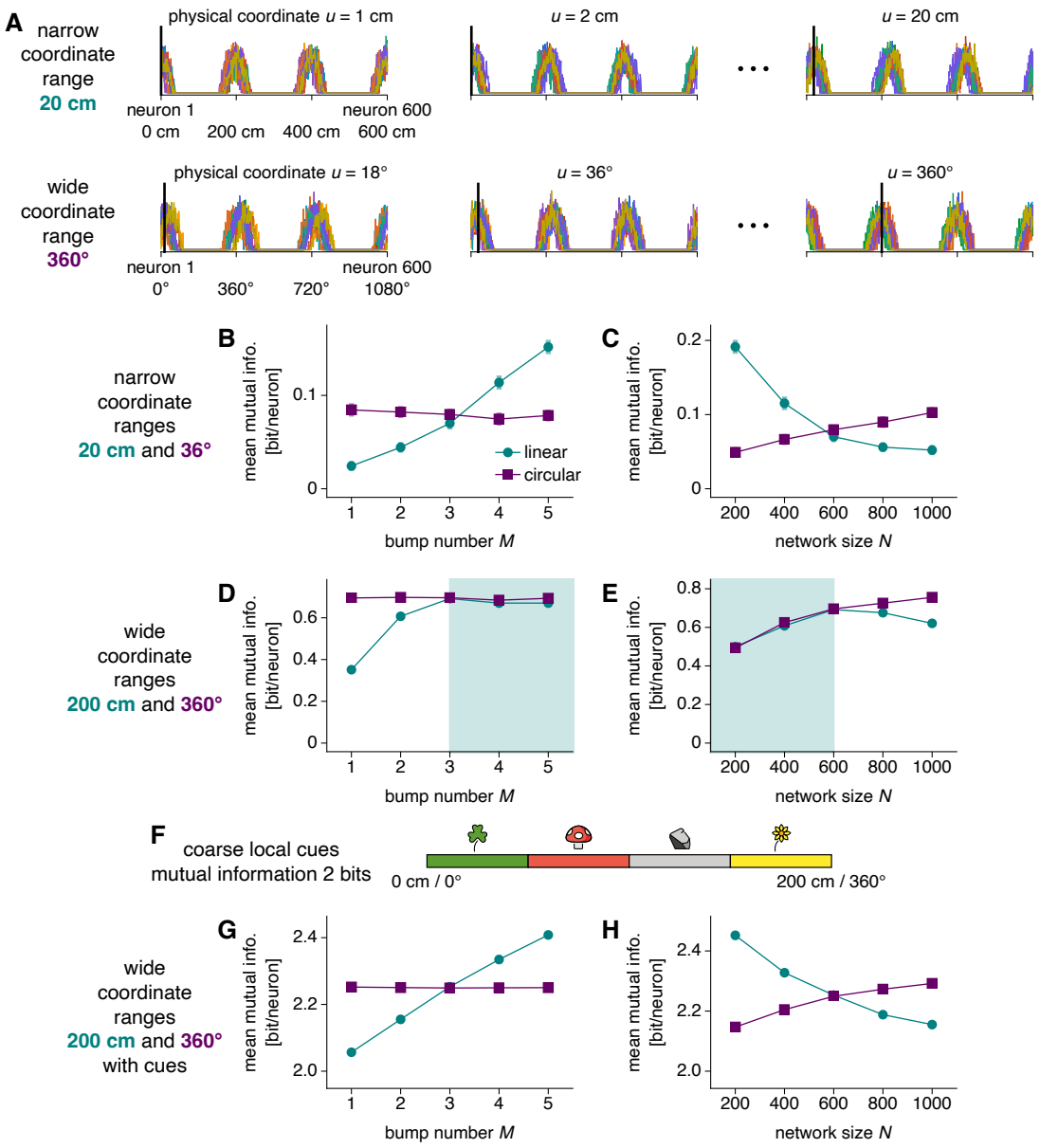

Figure 6: Mutual information between neural activity and physical coordinates with input noise of magnitude $\sigma = 0.5$. (**A**) To calculate mutual information, we initialize replicate simulations without input drive at different coordinate values (thick black lines) and record the final neural activities (thin colored lines). The physical coordinate can be linear or circular and its range can be narrow or wide; here, we illustrate two possibilities for networks with 600 neurons and 3 bumps. (**B, C**) Mutual information between physical coordinate and single-neuron activity under narrow coordinate ranges. (**B**) Information decreases with bump number for linear coordinates and remains largely constant for circular coordinates. Networks with 600 neurons. (**C**) Information decreases with network size for linear coordinates and increases for circular coordinates. Networks with 3 bumps. (**D, E**) Mutual information between physical coordinate and single-neuron activity under wide coordinate ranges. The trends in **B**, **C** are preserved for circular coordinates. They are also preserved for linear coordinates, except for the shaded regions in which the coordinate range exceeds the bump distance. (**F**) Coarse local cues are active over different quadrants of the wide coordinate ranges. (**G, H**) Mutual information between physical coordinate and the joint activities of a single neuron with the four cues in **F** under wide coordinate ranges. The trends in **B**, **C** are preserved for both linear and circular coordinates. Points indicate data from 96 replicate simulations at each coordinate value averaged over neurons and bars indicate bootstrapped standard errors of the mean.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

wide coordinate ranges to assess network performance in both regimes.

We first consider narrow coordinate ranges. For linear coordinates, information increases as a function of M but decreases as a function of N; for circular coordinates, it does not strongly depend on M and increases as a function of N (Fig. 6B, C). These results exactly corroborate those in Fig. 5A, B obtained for bump diffusion, since we expect information and diffusion to be inversely related.

We next consider wide coordinate ranges. Our ring networks can only uniquely represent coordinate ranges up to their bump distances (converted to physical distances by Fig. 4). Beyond these values, two physical coordinates separated by the converted bump distance cannot be distinguished by the network. Our mutual information analysis captures this phenomenon; for linear coordinates, the increase in information with larger M or smaller N as observed in Fig. 6B, C disappears once the converted bump distance drops below the physical range of 200 cm (green shaded regions of Fig. 6D, E). In this regime, the benefits of more bumps and smaller networks toward decreasing diffusion are counteracted by bump ambiguity. In contrast, the circular mapping in Fig. 4 lacks bump ambiguity since the bump distance is always converted to the maximum physical range of 360°, so the same qualitative trends in mutual information are observed for any coordinate range (Fig. 6D, E).

For linear coordinates with wide ranges, the advantages of increasing bump number can be restored by coarse local cues. We illustrate this process by introducing four cues, each of which is active over a different quadrant and is otherwise inactive (Fig. 6F). They can be conceptualized as two-state sensory neurons or neural populations that fire in the presence of a local stimulus. By themselves, the cues do not encode precise coordinate values. Mutual information calculated with the joint activity of each neuron with these cues recovers the behavior observed for narrow ranges across all M and N (Fig. 6G, H). Ring attractor neurons provide information beyond the 2 bits conveyed by the cues alone, and for position, this additional information increases with more bumps and fewer neurons without saturating.

In summary, our conclusions about robustness to input noise obtained by diffusion analysis are also supported by mutual information analysis. Moreover, the latter explicitly reveals how networks encoding wide, linear coordinate ranges can leverage coarse local cues to address ambiguities and preserve the advantages of multiple bumps.

More bumps improve robustness to connectivity noise under linear mapping

Another source of noise in biological CANs is inhomogeneity in the connectivity W. Perfect continuous attractor dynamics requires W to be invariant to translations along the network (Skaggs et al., 1995; Zhang, 1996; Samsonovich and McNaughton, 1997; Fuhs and Touretzky, 2006; Burak and Fiete, 2009), a concept related to Goldstone's theorem in physics (Nambu, 1960; Goldstone, 1961). We consider the effect of replacing $W \to W + V$, where V is a noisy connectivity matrix whose entries are independently drawn from a zero-mean Gaussian distribution. V disrupts the symmetries of W. This noise is quenched and does not change over the course of the simulation, in contrast to input and spiking noise which are independently sampled in time. V creates a noise term

$$\zeta_\alpha(x,t) = \sum_\beta \int \mathrm{d}y\, V_{\alpha\beta}(x,y) s_\beta(y,t). \tag{23}$$

Equation 23 shows that V produces correlated ζ's across neurons, which also differs from input and spiking noise. Because of these distinctions, the dominant effect of connectivity noise is not diffusion, but drift. V

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

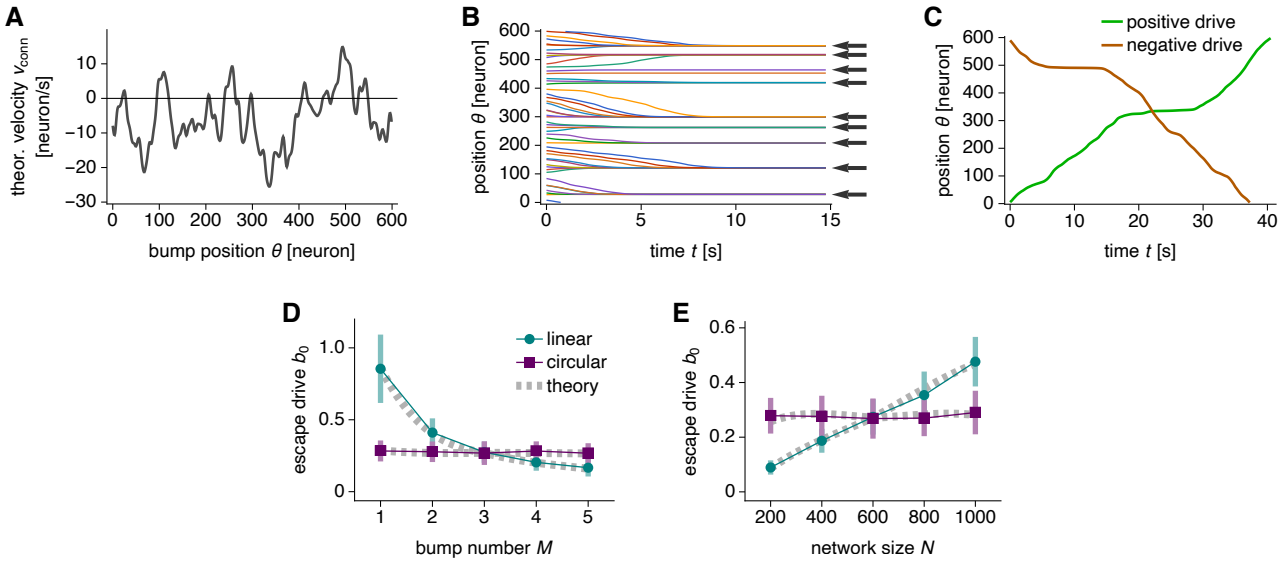

Figure 7: Bump trapping due to connectivity noise at low drive. (**A–C**) Networks with 600 neurons, 1 bump, and the same realization of connectivity noise of magnitude 0.002. (**A**) Theoretical values for drift velocity as a function of bump position using Eq. 24. (**B**) Bumps drift towards trapped positions over time. The drive is $b = 0$. Arrows indicate predictions from $v_{\text{conn}}(\theta)$ crossing 0 with negative slope in **A**. Lines indicate simulations with different starting positions. (**C**) Bump trajectories with smallest positive and negative drive required to travel through the entire network. Respectively, $b = 0.75$ and $b = -0.52$. The larger of the two in magnitude is the escape drive $b_0 = 0.75$. Note that positions with low bump speed exhibit large velocities in the opposite direction in **A**. (**D, E**) Networks with multiple realizations of connectivity noise of magnitude 0.002. (**D**) Escape drive decreases with bump number under linear mapping and remains largely constant under circular mapping. Networks with 600 neurons. (**E**) Escape drive increases with network size under linear mapping and remains largely constant under circular mapping. Networks with 3 bumps. Points indicate simulation means over 48 realizations and bars indicate standard deviations. Dotted gray lines indicate Eq. 26 averaged over 96 realizations.

induces bumps to move with velocity $v_{\text{conn}}(\theta)$, even without drive b:

$$v_{\text{conn}}(\theta) = -\frac{\sum_{\alpha\beta} \iint \mathrm{d}x\,\mathrm{d}y\, V_{\alpha\beta}(x,y) \frac{\mathrm{d}s(x-\theta)}{\mathrm{d}x} s(y-\theta)}{2\tau \int \mathrm{d}x \left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2}. \tag{24}$$

The movement is coherent but irregular, as it depends on the bump position θ (Fig. 7A). Itskov et al. (2011) and Seeholzer et al. (2019) refer to $v_{\text{conn}}(\theta)$ as the drift velocity.

Connectivity noise traps bumps at low drive b. We first consider $b = 0$, so bump motion is governed solely by drift according to $\mathrm{d}\theta/\mathrm{d}t = v_{\text{conn}}(\theta)$. The bump position θ has stable fixed points wherever $v_{\text{conn}}(\theta)$ crosses 0 with negative slope (Itskov et al., 2011; Seeholzer et al., 2019). Simulations confirm that bumps drift towards these points (Fig. 7B). The introduction of b adds a constant v_{drive} that moves the curve in Fig. 7A up for positive b or down for negative b:

$$v_{\text{total}}(\theta) = v_{\text{drive}} + v_{\text{conn}}(\theta). \tag{25}$$

If $v_{\text{total}}(\theta)$ still crosses 0, bumps would still be trapped. The absence of bump motion in response to coordinate changes encoded by b would be a catastrophic failure of path integration. To permit bump motion through

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

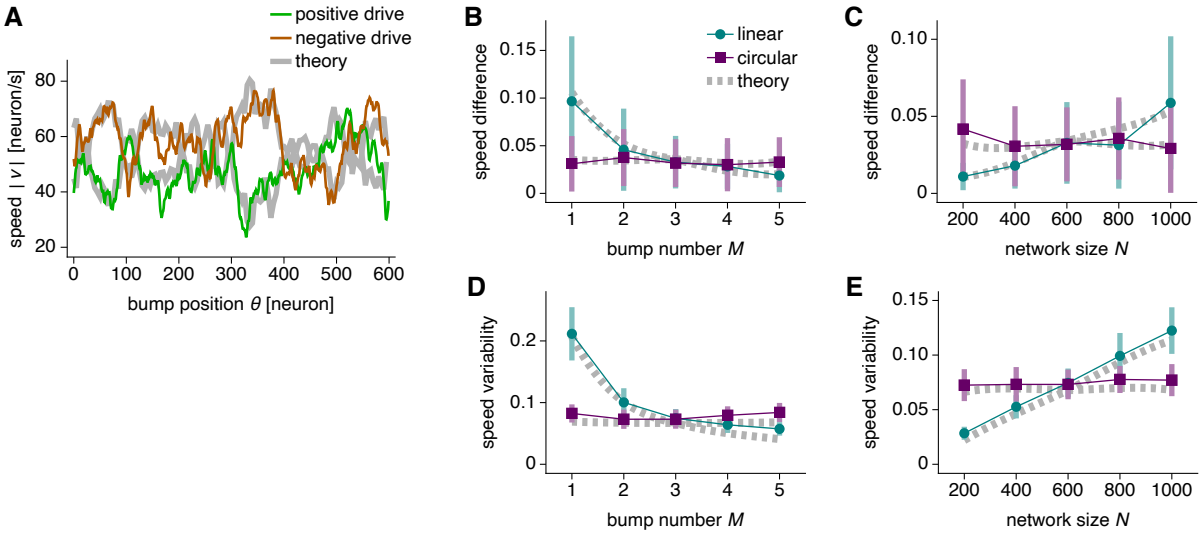

Figure 8: Bump speed irregularity due to connectivity noise at high drive. (**A**) Bump speed as a function of bump position with connectivity noise of magnitude 0.002 and drive $b = 1.5$. Network with 600 neurons, 1 bump, and the same realization of connectivity noise as in Fig. 7A–C. Thick gray lines indicate Eq. 25. (**B–E**) Networks with multiple realizations of connectivity noise of magnitude 0.002 and drive $b = 1.5$. (**B**) Speed difference between directions decreases with bump number under linear mapping and remains largely constant under circular mapping. Networks with 600 neurons. (**C**) Speed difference increases with network size under linear mapping and remains largely constant under circular mapping. Networks with 3 bumps. (**D, E**) Same as **B**, **C**, but for speed variability within each direction. Points indicate simulation means over 48 realizations and bars indicate standard deviations. Dotted gray lines indicate Eqs. 30 and 31 averaged over 96 realizations.

the entire network, the drive must be strong enough to eliminate all zero-crossings. Figure 7C shows bump motion at this drive for both directions of motion. The positive b is just large enough for the bump to pass through the region with the most negative $v_{\text{conn}}(\theta)$ in Fig. 7A; likewise for negative b and positive $v_{\text{conn}}(\theta)$. We call the larger absolute value of these two drives the escape drive b_0. Simulations show that b_0 decreases with bump number M and increases with network size N under linear mapping (Fig. 7D, E). A smaller b_0 implies weaker trapping, so smaller networks with more bumps are more resilient against this phenomenon. Under circular mapping, however, b_0 demonstrates no significant dependence on M or N. We can predict b_0 by inverting the relationship in Eq. 8 between b and v:

$$b_0 = -\frac{\max_\theta |v_{\text{conn}}(\theta)| \cdot \tau \int \mathrm{d}x \left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2}{\gamma\xi \int \mathrm{d}x \, \frac{\mathrm{d}^2 s}{\mathrm{d}x^2}}. \tag{26}$$

This theoretical result agrees well with values obtained by simulation (Fig. 7D, E). In the Theoretical model section, we present a heuristic argument (Eq. 124) that leads to the observed scaling of escape drive on M and N:

$$b_0 \propto \frac{N}{M} \quad \text{linear}, \qquad b_0 \propto 1 \quad \text{circular}. \tag{27}$$

At high drive $|b| > b_0$, attractor bumps are no longer trapped by the drift velocity $v_{\text{conn}}(\theta)$. Instead, the drift term produces irregularities in the total velocity $v_{\text{total}}(\theta)$ (Fig. 8A). They can be decomposed into two components: irregularities between directions of motion and irregularities within each direction. Both imply errors in path integration because v and b are not strictly proportional. To quantify these components, we

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

call $|v_+(\theta)|$ and $|v_-(\theta)|$ the observed bump speeds under positive and negative b. We define speed difference as the unsigned difference between mean speeds in either direction, normalized by the overall mean speed:

$$\text{speed difference} = \frac{2\left|\operatorname*{mean}_\theta |v_+(\theta)| - \operatorname*{mean}_\theta |v_-(\theta)|\right|}{\operatorname*{mean}_\theta |v_+(\theta)| + \operatorname*{mean}_\theta |v_-(\theta)|}. \tag{28}$$

We then define speed variability as the standard deviation of speeds within each direction, averaged over both directions and normalized by the overall mean speed:

$$\text{speed variability} = \frac{\operatorname*{std}_\theta |v_+(\theta)| + \operatorname*{std}_\theta |v_-(\theta)|}{\operatorname*{mean}_\theta |v_+(\theta)| + \operatorname*{mean}_\theta |v_-(\theta)|}. \tag{29}$$

Speed difference and speed variability follow the same trends under changes in bump number M and network size N (Fig. 8B–E). Under linear mapping, they decrease with M and increase with N. Under circular mapping, they do not significantly depend on M and N. These are also the same trends exhibited by the escape drive b_0 (Fig. 7D, E). In terms of theoretical quantities, the formulas for speed difference and variability become

$$\text{speed difference} = \frac{2\left|\operatorname*{mean}_\theta v_{\text{conn}}(\theta)\right|}{|v_{\text{drive}}|} \tag{30}$$

and

$$\text{speed variability} = \frac{\operatorname*{std}_\theta v_{\text{conn}}(\theta)}{|v_{\text{drive}}|}. \tag{31}$$

These expressions match the observed values well (Fig. 8B–E). In the Theoretical methods section, we calculate the observed dependence of speed difference (Eq. 113) and variability (Eq. 120) on M and N:

$$\text{speed difference and variability} \propto \frac{N}{M} \quad \text{linear}, \qquad \text{speed difference and variability} \propto 1 \quad \text{circular}. \tag{32}$$

For all results related to connectivity noise, the coupling strength rescaling in Eq. 13 produces the differences between the two mappings via the γ in Eq. 8. The conversion factor in Eq. 12 does not play a role because escape drive, speed difference, and speed variability do not have dimensions of network distance.

To summarize, CANs with imperfect connectivity benefit from more attractor bumps when encoding linear coordinates. This advantage is present at all driving inputs and may be more crucial for larger networks. On the other hand, connectivity noise has largely the same consequences for networks of all bump numbers and sizes when encoding circular coordinates.

Discussion

We demonstrated how CANs capable of path integration respond to three types of noise. Additive synaptic input noise and Poisson spiking noise cause bumps to diffuse away from the coherent motion responsible for path integration (Figs. 3 and 5). This diffusion is accompanied by a decrease in mutual information between neural activity and encoded coordinate (Fig. 6). Connectivity noise produces a drift velocity field that also impairs path integration by trapping bumps at low drive and perturbing bump motion at high drive (Figs. 7 and 8).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For all three types of noise, CANs with more attractor bumps exhibit less deviation in bump motion in network units. This is observed across network parameters (Figs. 9 and 10 in the Appendix). Thus, CANs can more robustly encode linear variables whose mapping inherits network units and does not rescale with bump number (Fig. 4A). If grid cell networks were to encode spatial position in this manner, then multiple attractor bumps would be preferred over a single bump. Gu et al. (2018) report experimental evidence supporting multi-bump grid networks obtained by calcium imaging of mouse medial entorhinal cortex. Our work implies that the evolution of such networks may have been partially encouraged by biological noise. Additional bumps introduce greater ambiguity among positions encoded by each bump, but this can be resolved by a rough estimate of position from additional cues, such as local landmarks (O'Keefe and Burgess, 2005; Krupic et al., 2014; Bush and Burgess, 2014; Hardcastle et al., 2015), another grid module with different periodicity (O'Keefe and Burgess, 2005; McNaughton et al., 2006; Stensola et al., 2012; Stemmler et al., 2015; Kang and Balasubramanian, 2019; Khona et al., 2022), or a Bayesian prior based on recent memory (Sreenivasan and Fiete, 2011). In this way, grid modules with finer resolution and more attractor bumps could maintain a precise egocentric encoding of position, while coarser modules and occasional allocentric cues would identify the true position out of the few possibilities represented. We explicitly explored one realization of this concept and observed how cues enable networks to improve their information content by increasing bump number, despite a concomitant increase in bump ambiguity (Fig. 6F–H).

In contrast, CANs encoding circular variables may rescale under different bump numbers to match periodicities (Fig. 4B), which eliminates any influence of bump number on encoding accuracy for all three types of noise. If head direction networks were to encode orientation in this manner, then they would face less selective pressure to evolve beyond the single-bump configuration observed in *Drosophila* (Seelig and Jayaraman, 2015). Moreover, without the assumption of bump shape invariance accomplished by Eq. 7, robustness to all three types of noise decreases with bump number, which actively favors single-bump orientation networks (Fig. 10 in the Appendix). Further experimental characterization of bump number in biological CANs, perhaps through techniques proposed by Widloski et al. (2018), would test the degree to which the brain can leverage the theoretical advantages identified in this work.

Under linear mapping, larger CANs exhibit more errors in path integration from all three types of noise. The immediate biological implication is that larger brains face a dramatic degradation of CAN performance, accentuating the importance of suppressing error with multi-bump networks. However, this simple rule that one neuron always represents a fixed physical interval does not need to be followed. Furthermore, larger animals may tolerate greater absolute errors in path integration because they interact with their environments over larger scales. Nevertheless, our results highlight the importance of considering network size when studying the performance of noisy CANs. Under circular mapping, bump diffusion decreases with network size for input and spiking noise, and the magnitude of errors due to connectivity noise is independent of network size. This implies that head direction networks can benefit from incorporating more neurons; the observed interactions between such networks across different mammalian brain regions may act in this manner to suppress noise (Taube, 2007).

The computational advantages of periodic over nonperiodic encodings has been extensively studied in the context of grid cells (Fiete et al., 2008; Sreenivasan and Fiete, 2011; Mathis et al., 2012; Wei et al., 2012; Almeida et al., 2015; Wei et al., 2015; Stemmler et al., 2015). Our results extend these findings by demonstrating that some kinds of periodic encodings can perform better than others. Our results also contribute to a rich literature on noisy CANs. Previous studies have investigated additive input noise (Compte et al., 2000; Wu et al., 2008; Burak and Fiete, 2012; Kilpatrick and Ermentrout, 2013; Seeholzer et al., 2019),

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

multiplicative input noise (Krishnan et al., 2018), spiking noise (Burak and Fiete, 2009, 2012; Wei et al., 2012; Almeida et al., 2015; Bouchacourt and Buschman, 2019; Seeholzer et al., 2019), and quenched noise due to connectivity or input inhomogeneities (Zhang, 1996; Itskov et al., 2011; Kilpatrick and Ermentrout, 2013; Seeholzer et al., 2019; Can and Krishnamurthy, 2021). Among these works, the relationship between bump number and noise has only been considered in the context of multiple-item working memory, in which each network can be loaded with various numbers of bumps (Wei et al., 2012; Almeida et al., 2015; Krishnan et al., 2018; Bouchacourt and Buschman, 2019). Interestingly, they find that robustness to noise decreases with bump number, which is opposite to our results (cf. Almeida et al., 2015, who report no dependence between bump number and encoding accuracy under certain conditions). It appears that CANs designed for path integration with fixed bump number and CANs designed for multiple-item working memory with variable bump number differ crucially in their responses to noise. Further lines of investigation that compare these two classes would greatly contribute to our general understanding of CANs.

Beyond our concrete results on CAN performance, our work offers a comprehensive theoretical framework for studying path-integrating CANs. We derive a formula for the multi-bump attractor state and a Lyapunov functional that governs its formation. We calculate all key dynamical quantities such as velocities and diffusion coefficients in terms of firing rates. Our formulas yield scaling relationships that facilitate an intuitive understanding for their dependence on bump number and network size. Much of our theoretical development does not assume a specific connectivity matrix or nonlinear activation function, which allows our results to have wide significance. For example, we expect them to hold for path-integrating networks that contain excitatory synapses. Other theories have been developed for bump shape (Wu et al., 2002; Xie et al., 2002; Itskov et al., 2011; Kilpatrick and Ermentrout, 2013; Widloski, 2015; Krishnan et al., 2018), path integration velocity (Xie et al., 2002; Mosheiff and Burak, 2019), diffusion coefficients (Wu et al., 2008; Burak and Fiete, 2012; Kilpatrick and Ermentrout, 2013; Krishnan et al., 2018; Seeholzer et al., 2019), and drift velocity (Zhang, 1996; Itskov et al., 2011; Seeholzer et al., 2019). Our work unifies these studies through a simple framework that features path integration, multiple bumps, and a noise term that can represent a wide range of sources. It can be easily extended to include other components of theoretical or biological significance, such as slowly-varying inputs (Tsodyks and Sejnowski, 1995; Fung et al., 2010; Kilpatrick and Ermentrout, 2013), synaptic plasticity (Stringer et al., 2002; Renart et al., 2003), neural oscillations (Thurley et al., 2008; Navratilova et al., 2012; Kang and DeWeese, 2019), and higher-dimensional attractor manifolds (Ermentrout and Cowan, 1979; Samsonovich and McNaughton, 1997).

Theoretical model

Architecture

We investigate CAN dynamics through a 1D ring attractor network. This class of network has been analyzed in previous theoretical works, and at various points, our calculations will parallel those in Xie et al. (2002); Itskov et al. (2011); Burak and Fiete (2012); Kilpatrick and Ermentrout (2013); Widloski (2015); Seeholzer et al. (2019); Mosheiff and Burak (2019).

There are two neurons at each position $i = 0, \ldots, N-1$ with population indexed by $\alpha \in \{\mathrm{L}, \mathrm{R}\}$ (Fig. 1A). For convenient calculation, we unwrap the ring and connect copies end-to-end, forming a linear network with continuous positions $x \in (-\infty, \infty)$. Unless otherwise specified, integrals are performed over the entire range. To map back onto the finite-sized ring network, we enforce our results to have a periodicity λ that divides N. For example, $\lambda = N$ corresponds to a single-bump configuration. Integrals would then be performed over $[0, N)$, with positions outside this range corresponding to their equivalents within this range.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The network obeys the following dynamics for synaptic inputs g:

$$\tau \frac{\mathrm{d}g_\alpha(x,t)}{\mathrm{d}t} + g_\alpha(x,t) = \sum_\beta \int \mathrm{d}y\, W_\beta(x,y) s_\beta(y,t) + A \pm_\alpha \gamma b(t) + \zeta_\alpha(x,t), \tag{33}$$

where $\pm_\mathrm{L}$ means $-$ and $\pm_\mathrm{R}$ means $+$, and the opposite for $\mp_\alpha$. τ is the neural time constant, W is the synaptic connectivity, and A is the resting input. The nonlinear activation function ϕ converts synaptic inputs to firing rates:

$$s_\alpha(x,t) = \phi[g_\alpha(x,t)]. \tag{34}$$

Most of our results will apply to general ϕ, but we also consider a ReLU activation function specifically:

$$\phi[g] = \begin{cases} g & g > 0 \\ 0 & g \leq 0. \end{cases} \tag{35}$$

In this section, we will explicitly mention when we specifically consider the ReLU case, and we will always simplify the function away. Thus, if an expression contains the symbol ϕ, then it applies to general ϕ. In the Results section, formulas and scalings for D_input, D_spike, v_drive, and $v_\mathrm{conn}(\theta)$, as well as all simulation results invoke Eq. 35. We will use this form in the Bump shape g subsection b is the driving input, γ is its coupling strength, and ζ is the noise, which can take different forms. γb and ζ are small compared to the rest of the right-hand side of Eq. 33. For notational convenience, we will often suppress dependence on t.

$W_\beta(x,y)$ obeys a standard continuous attractor architecture based on a symmetric and translation invariant W:

$$W_\beta(x,y) = W(x - y \mp_\beta \xi) \quad \text{where} \quad W(-x) = W(x). \tag{36}$$

Each population β deviates from W by a small shift $\xi \ll N$ in synaptic outputs. Thus, the following approximation holds:

$$\sum_\beta W_\beta(x,y) \approx 2W(x-y). \tag{37}$$

We will consider the specific form of W (Fig. 1B):

$$W(x) = \begin{cases} w \cdot \dfrac{\cos \pi x/l - 1}{2} & |x| < 2l \\ 0 & |x| \geq 2l \end{cases} = \begin{cases} w \cdot \dfrac{\cos kx - 1}{2} & |x| < 2\pi/k \\ 0 & |x| \geq 2\pi/k, \end{cases} \tag{38}$$

where $k = \pi/l$. We will explicitly mention when we specifically consider this form; in fact, we only do so for Eqs. 46, 47, 59, and 60, as well as for our simulation results in the Results section. Otherwise, each expression applies to general W.

Baseline configuration without drive and noise

Linearized dynamics and bump distance λ

First, we consider the case of no drive $b = 0$ and no noise $\zeta = 0$. The dynamical equation Eq. 33 becomes

$$\tau \frac{\mathrm{d}g_\alpha(x)}{\mathrm{d}t} + g_\alpha(x) = \sum_\beta \int \mathrm{d}y\, W_\beta(x,y) \phi[g_\beta(y)] + A. \tag{39}$$

Since the right-hand side no longer depends on α, g must be the same for both populations, and we can use Eq. 37 to obtain

$$\tau \frac{\mathrm{d}g(x)}{\mathrm{d}t} + g(x) = 2 \int \mathrm{d}y\, W(x-y) \phi[g(y)] + A. \tag{40}$$

We analyze these dynamics using the Fourier transform $\mathcal{F}$. Our chosen convention, applied to the function

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

h, is

$$\tilde{h}(q) = \mathcal{F}[h](q) = \int \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}qx} h(x)$$
$$h(x) = \mathcal{F}^{-1}[\tilde{h}](x) = \int \frac{\mathrm{d}q}{2\pi}\, \mathrm{e}^{\mathrm{i}qx} \tilde{h}(q). \tag{41}$$

Fourier modes $\tilde{h}(q)$ represent sinusoids with wavenumber q and corresponding wavelength $2\pi/q$. Applying this transform to Eq. 40, we obtain

$$\tau \frac{\mathrm{d}\tilde{g}(q)}{\mathrm{d}t} + \tilde{g}(q) = 2\tilde{W}(q)\mathcal{F}\big[\phi[g]\big](q) + 2\pi A\delta(q). \tag{42}$$

In this subsection, we consider the case of small deviations, such that $g(x) \approx g_0$ and $\phi[g(x)] \approx \phi[g_0] + \phi'[g_0](g(x) - g_0)$. Then, Eq. 42 becomes

$$\tau \frac{\mathrm{d}\tilde{g}(q)}{\mathrm{d}t} + \tilde{g}(q) = 2\phi'[g_0]\tilde{W}(q)\tilde{g}(q) + 2\pi A_0\delta(q), \tag{43}$$

where $A_0 = A + 2\tilde{W}(0)(\phi[g_0] - \phi'[g_0]g_0)$. The solution to this linearized equation for $q \neq 0$ is

$$\tilde{g}(q,t) = \tilde{g}(q,0)\mathrm{e}^{r(q)t}. \tag{44}$$

Each mode grows exponentially with rate

$$r(q) = \frac{2\phi'[g_0]\tilde{W}(q) - 1}{\tau}, \tag{45}$$

so the fastest-growing component of g is the one that maximizes $\tilde{W}(q)$, as stated in Eq. 5 of the Results section. The wavelength $2\pi/q$ of that component predicts the bump distance λ.

For the specific W in Eq. 38, its Fourier transform is

$$\tilde{W}(q) = -w\frac{k^2 \sin\frac{2\pi q}{k}}{k^2 q - q^3}, \tag{46}$$

so

$$\lambda = \frac{2\pi}{\underset{q}{\operatorname{argmin}} \dfrac{k^2 \sin\frac{2\pi q}{k}}{k^2 q - q^3}} = \frac{2l}{\underset{\psi}{\operatorname{argmin}} \dfrac{\sin 2\pi\psi}{\psi - \psi^3}} \approx 2.28l. \tag{47}$$

λ is proportional to l, as also noted by Fuhs and Touretzky (2006); Burak and Fiete (2009); Kang and Balasubramanian (2019); Khona et al. (2022).

Bump shape g

We call the steady-state synaptic inputs g without drive and noise the baseline configuration. To calculate its shape, we must account for the nonlinearity of the activation function ϕ and return to Eq. 42. We invoke our particular form of ϕ in Eq. 35 to calculate $\mathcal{F}\big[\phi[g]\big](q)$. g must be periodic, and its periodicity is the bump distance λ with wavenumber $\kappa = 2\pi/\lambda$. Without loss of generality, we take g to have a bump centered at 0. Since W is symmetric, g is an even function. We define z as the position where g crosses 0:

$$g(z) = 0. \tag{48}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

If g is approximately sinusoidal, then $g(x) > 0$ wherever $n\lambda - z < x < n\lambda + z$ for any integer n. The ReLU activation function in Eq. 35 implies

$$\phi[g(x)] = g(x)\Phi(x) \quad \text{where} \quad \Phi(x) = \sum_{n=-\infty}^{\infty} \Theta[x - n\lambda + z]\Theta[-(x - n\lambda - z)]. \tag{49}$$

Θ is the Heaviside step function. The Fourier transform for Φ is

$$\tilde{\Phi}(q) = 2\frac{\sin qz}{q}\sum_{n=-\infty}^{\infty} \mathrm{e}^{-2\pi i n q/\kappa} = 2\frac{\sin qz}{q}\sum_{n=-\infty}^{\infty}\delta\left(n - \frac{q}{\kappa}\right) = 2\kappa\frac{\sin qz}{q}\sum_{n=-\infty}^{\infty}\delta(q - n\kappa), \tag{50}$$

where the second equality comes from the Fourier series for a Dirac comb. Therefore,

$$\mathcal{F}\big[\phi[g]\big](q) = \frac{1}{2\pi}\int \mathrm{d}q'\,\tilde{\Phi}(q - q')\tilde{g}(q') = \frac{1}{\pi}\sum_{n=-\infty}^{\infty}\frac{\sin n\kappa z}{n}\tilde{g}(q - n\kappa), \tag{51}$$

so Eq. 42 becomes

$$\tau\frac{\mathrm{d}\tilde{g}(q)}{\mathrm{d}t} + \tilde{g}(q) = \frac{2}{\pi}\tilde{W}(q)\sum_{n=-\infty}^{\infty}\frac{\sin n\kappa z}{n}\tilde{g}(q - n\kappa) + 2\pi A\delta(q). \tag{52}$$

This equation describes the full dynamics of g with a ReLU activation function. It contains couplings between all modes q that are multiples of the wavenumber κ, which corresponds to the bump distance.

To find the baseline g, we set $\mathrm{d}\tilde{g}/\mathrm{d}t = 0$. We also simplify $\tilde{g}(q)$ by only considering the lowest modes that couple to each other: $q = 0, \pm\kappa$. Due to symmetry, $\tilde{W}(-q) = \tilde{W}(q)$ and $\tilde{g}(-q) = \tilde{g}(q)$. Eq. 52 gives

$$\begin{aligned}
\tilde{g}(0) &= \frac{2}{\pi}\tilde{W}(0)\Big[\kappa z\,\tilde{g}(0) + 2\sin(\kappa z)\,\tilde{g}(\kappa)\Big] + 2\pi A\delta(0)\\
\tilde{g}(\kappa) &= \frac{2}{\pi}\tilde{W}(\kappa)\left[\sin(\kappa z)\,\tilde{g}(0) + \left(\kappa z + \frac{\sin 2\kappa z}{2}\right)\tilde{g}(\kappa)\right].
\end{aligned} \tag{53}$$

Now we need to impose Eq. 48: $g(z) = 0$. To do so, we note that $\tilde{g}(0)$ and $\tilde{g}(\kappa)$ are both proportional to $\delta(0)$ according to Eq. 53. That means $\tilde{g}(q)$ has the form

$$\tilde{g}(q) = G_0\delta(q) + G\delta(q - \kappa) + G\delta(q + \kappa), \tag{54}$$

where G_0 and G are the Fourier modes with delta functions separated. This implies

$$g(x) = \frac{G_0}{2\pi} + \frac{G}{\pi}\cos\kappa x, \tag{55}$$

and $g(z) = 0$ implies

$$G_0 = -2\cos(\kappa z)\,G. \tag{56}$$

Substituting Eqs. 54 and 56 into Eq. 53, we obtain

$$\begin{aligned}
\frac{G}{\pi} &= -\frac{\pi A}{2\tilde{W}(0)\big(\sin\kappa z - \kappa z\cos\kappa z\big) + \pi\cos\kappa z}\\
\kappa z - \cos\kappa z\sin\kappa z &= \frac{\pi}{2\tilde{W}(\kappa)}.
\end{aligned} \tag{57}$$

We can solve the second equation of Eq. 57 for κz and then substitute it into the first equation to obtain G. This then gives us $g(x)$, which becomes through Eqs. 55 and 56

$$g(x) = \frac{G}{\pi}(\cos\kappa x - \cos\kappa z). \tag{58}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In particular, let's use the W defined by Eq. 38 with Fourier transform Eq. 46. Then,

$$\tilde{W}(0) = -\frac{2\pi w}{k} \qquad \text{and} \qquad \tilde{W}(\kappa) = -\frac{w}{\kappa}\frac{k^2}{k^2-\kappa^2}\sin\frac{2\pi\kappa}{k}. \tag{59}$$

Thus,

$$\begin{aligned} g(x) &= \frac{G}{\pi}(\cos\kappa x - \cos\kappa z) \\ \kappa z - \cos\kappa z \sin\kappa z &= -\pi \Bigg/ \left[\frac{2w}{\kappa}\frac{k^2}{k^2-\kappa^2}\sin\frac{2\pi\kappa}{k}\right] \\ \frac{G}{\pi} &= A \Bigg/ \left[\frac{4w}{k}\big(\sin\kappa z - \kappa z\cos\kappa z\big) - \cos\kappa z\right]. \end{aligned} \tag{60}$$

This provides expressions for a and d in Eq. 4 of the Results section, where $a = G/\pi$ and $d = -(G/\pi)\cos\kappa z$.

Lyapunov functional and bump distance λ

The dynamical equation in Eq. 40 admits a Lyapunov functional. In analogy to the continuous Hopfield model (Hopfield, 1984), we can define a Lyapunov functional in terms of $s(x) = \phi[g(x)]$:

$$L = -\iint \mathrm{d}x\,\mathrm{d}y\, W(x-y)s(x)s(y) + \int \mathrm{d}x \int_0^{s(x)} \mathrm{d}\rho\, \phi^{-1}[\rho] - A\int \mathrm{d}x\, s(x). \tag{61}$$

The nonlinearity ϕ must be invertible in the range $(0, s)$ for any possible firing rate s. For L to be bounded from below for a network of any size N, we need

1. $W(x)$ to be negative definite, and
2. $\int_0^s \mathrm{d}\rho\, \phi^{-1}[\rho] - As$ to be bounded from below for any possible firing rate s.

We can check that these hold for our particular functions. Equation 38 immediately shows that the first condition is met. Equation 35 states that $\phi^{-1}[\rho] = \rho$ when $\rho > 0$, so $\int_0^s \mathrm{d}\rho\, \phi^{-1}[\rho] - As = \frac{1}{2}s^2 - As$, which satisfies the second condition.

Now we take the time derivative and use Eq. 40:

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}t} &= -\int \mathrm{d}x \left\{2\int \mathrm{d}y\, W(x-y)s(y) - \phi^{-1}[s(x)] + A\right\}\frac{\mathrm{d}s(x)}{\mathrm{d}t} \\ &= -\tau \int \mathrm{d}x\, \frac{\mathrm{d}g(x)}{\mathrm{d}t}\frac{\mathrm{d}s(x)}{\mathrm{d}t} \\ &= -\tau \int \mathrm{d}x\, \phi'[g(x)]\left(\frac{\mathrm{d}g(x)}{\mathrm{d}t}\right)^2. \end{aligned} \tag{62}$$

As long as ϕ is a monotonically nondecreasing function, $\mathrm{d}L/\mathrm{d}t \le 0$. Thus, L is a Lyapunov functional.

Now we seek to simplify Eq. 61. Suppose we are very close to a steady-state solution, so $\mathrm{d}g/\mathrm{d}t \approx 0$. We substitute Eq. 40 into Eq. 61 to obtain

$$\begin{aligned} L &= -\frac{1}{2}\int \mathrm{d}x\, \big[g(x) - A\big]s(x) + \int \mathrm{d}x \int_0^{s(x)} \mathrm{d}\rho\, \phi^{-1}[\rho] - A\int \mathrm{d}x\, s(x) \\ &= -\frac{1}{2}\int \mathrm{d}x\, g(x)s(x) + \int \mathrm{d}x \int_0^{s(x)} \mathrm{d}\rho\, \phi^{-1}[\rho] - \frac{A}{2}\int \mathrm{d}x\, s(x). \end{aligned} \tag{63}$$

Now we invoke our ReLU ϕ from Eq. 35 to obtain

$$L = -\frac{1}{2}\int \mathrm{d}x\, \big[g(x) - s(x)\big]s(x) - \frac{A}{2}\int \mathrm{d}x\, s(x) = -\frac{A}{2}\int \mathrm{d}x\, s(x). \tag{64}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The last equality was obtained by noticing that for any x, either $s(x) = 0$ or $g(x) - s(x) = 0$ with our ϕ. Therefore, the stable solution that minimizes L is the one that maximizes the total firing rate.

We can apply our sinusoidal g in Eq. 58 to perform the integral:

$$L = -\frac{NAG}{2\pi^2}(\sin\kappa z - \kappa z\cos\kappa z), \tag{65}$$

where N is the network size. So L depends on G and the quantity κz, which we will rewrite as ψ. We now simplify Eq. 65 using Eq. 57:

$$L = -\frac{NA^2\big(\sin\psi - \psi\cos\psi\big)}{4\tilde{W}(0)\big(\sin\psi - \psi\cos\psi\big) - 2\pi\cos\psi} = -\frac{NA^2}{4\tilde{W}(0) + \frac{2\pi}{\psi - \tan\psi}}. \tag{66}$$

Note that $1/(\tan\psi - \psi)$ is a monotonically increasing function of ψ in the range $[0, \pi]$, so to minimize L, we need to minimize ψ. Meanwhile, Eq. 57 now reads $\psi - \cos\psi\sin\psi = \pi/2\tilde{W}(\kappa)$. The left-hand side is also a monotonically increasing function of ψ in the range $[0, \pi]$, so to minimize ψ, we need to maximize $\tilde{W}(\kappa)$. Thus, the Lyapunov stable wavelength $\lambda = 2\pi/\kappa$ is the one that maximizes $\tilde{W}(\kappa)$. This is the same mode that grows the fastest for the linearized dynamics in Eq. 45.

Bump motion under drive and noise

Dynamics along the attractor manifold

Now that we have determined the baseline configuration g, including the bump shape and bump distance, we investigate its motion under drive b and noise ζ. We introduce θ to label the position of the configuration. It can be defined as the center of mass or the point of maximum activity of one of the bumps. We expand the full time-dependent configuration with respect to the baseline configuration located at θ:

$$g_\alpha(x, t) = g(x - \theta) + \delta g_\alpha(x, t). \tag{67}$$

$g(x - \theta)$ solves Eq. 40 with $\mathrm{d}g/\mathrm{d}t = 0$; to facilitate calculations below, we will write the baseline equation in this form:

$$g(x - \theta) = \sum_\beta \int \mathrm{d}y\, W_\beta(x, y)\phi[g(y - \theta)] + A. \tag{68}$$

Substituting Eq. 67 into Eq. 33 and invoking Eq. 68, we obtain the following linearized dynamics for δg:

$$\tau\frac{\mathrm{d}\delta g_\alpha(x, t)}{\mathrm{d}t} + \delta g_\alpha(x, t) = \sum_\beta \int \mathrm{d}y\, W_\beta(x, y)\phi'[g(y - \theta)]\delta g_\beta(y, t) \pm_\alpha \gamma b(t) + \zeta_\alpha(x, t). \tag{69}$$

We can rewrite this as

$$\tau\frac{\mathrm{d}\delta g_\alpha(x, t)}{\mathrm{d}t} = \sum_\beta \int \mathrm{d}y\, K_{\alpha\beta}(x, y; \theta)\delta g_\beta(y, t) \pm_\alpha \gamma b(t) + \zeta_\alpha(x, t), \tag{70}$$

where

$$K_{\alpha\beta}(x, y; \theta) = W_\beta(x, y)\phi'[g(y - \theta)] - \delta_{\alpha\beta}\delta(x - y). \tag{71}$$

We will often suppress the argument of derivatives of g. If we consider a configuration located at θ, $\mathrm{d}g/\mathrm{d}x$ implies $\mathrm{d}g(x - \theta)/\mathrm{d}x$. We make the argument explicit when necessary.

If we differentiate Eq. 68 by θ, we obtain

$$\begin{aligned} \frac{\mathrm{d}g}{\mathrm{d}x} &= \sum_\beta \int \mathrm{d}y\, W_\beta(x, y)\phi'[g(y - \theta)]\frac{\mathrm{d}g}{\mathrm{d}y} \\ 0 &= \sum_\beta \int \mathrm{d}y\, K_{\alpha\beta}(x, y; \theta)\frac{\mathrm{d}g}{\mathrm{d}y}, \end{aligned} \tag{72}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which indicates that $\mathrm{d}g/\mathrm{d}x$ is a right eigenvector of K with eigenvalue 0. To be explicit about this, we recover the discrete case by converting continuous functions to vectors and matrices:

$$g_i = g(i-\theta), \qquad \Delta g_i = \left.\frac{\mathrm{d}g(x-\theta)}{\mathrm{d}x}\right|_{x=i}, \qquad K_{\alpha\beta ij} = K_{\alpha\beta}(i,j;\theta). \tag{73}$$

If we concatenate matrices and vectors across populations as

$$\mathsf{J} = \begin{pmatrix} \mathsf{K}_{\mathrm{LL}} & \mathsf{K}_{\mathrm{LR}} \\ \mathsf{K}_{\mathrm{RL}} & \mathsf{K}_{\mathrm{RR}} \end{pmatrix}, \qquad \mathbf{e} = \begin{pmatrix} \Delta\mathbf{g} \\ \Delta\mathbf{g} \end{pmatrix}, \tag{74}$$

$\mathbf{e}$ is the right null eigenvector of J: $0 = \sum_j J_{ij} e_j$.

Since K is not symmetric, its left and right eigenvectors may be different. To find the left null eigenvector, we again differentiate Eq. 68 with respect to θ, but this time interchanging variables x and y:

$$\begin{aligned} \frac{\mathrm{d}g}{\mathrm{d}y} &= \sum_\beta \int \mathrm{d}x\, W_\beta(y,x)\phi'[g(x-\theta)]\frac{\mathrm{d}g}{\mathrm{d}x} \\ &\approx 2\int \mathrm{d}x\, W(x-y)\phi'[g(x-\theta)]\frac{\mathrm{d}g}{\mathrm{d}x}. \end{aligned} \tag{75}$$

The second equality is obtained from Eqs. 36 and 37. Replacing the position y by $y \pm_\beta \xi$, where ξ is the connectivity shift, we get

$$\frac{\mathrm{d}g(y-\theta\pm_\beta\xi)}{\mathrm{d}y} \approx 2\int \mathrm{d}x\, W(x-y\mp_\beta\xi)\phi'[g(x-\theta)]\frac{\mathrm{d}g(x-\theta)}{\mathrm{d}x}, \tag{76}$$

where we have made the arguments of g explicit. Let's define shifted versions of the baseline g for each population α:

$$\bar{g}_\alpha(x) = g(x\pm_\alpha\xi). \tag{77}$$

Since ξ is small,

$$\sum_\alpha \bar{g}_\alpha(x) \approx 2g(x). \tag{78}$$

Applying these expressions to Eq. 76 and recalling Eq. 36,

$$\begin{aligned} \frac{\mathrm{d}\bar{g}_\beta}{\mathrm{d}y} &\approx 2\int \mathrm{d}x\, W_\beta(x,y)\phi'[g(x-\theta)]\frac{\mathrm{d}g}{\mathrm{d}x} \\ &\approx \sum_\alpha \int \mathrm{d}x\, W_\beta(x,y)\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}. \end{aligned} \tag{79}$$

Finally, we multiply both sides of the equation by $\phi'[g(y-\theta)]$ to obtain

$$\begin{aligned} \phi'[g(y-\theta)]\frac{\mathrm{d}\bar{g}_\beta}{\mathrm{d}y} &\approx \sum_\alpha \int \mathrm{d}x\, W_\beta(x,y)\phi'[g(y-\theta)]\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x} \\ 0 &= \sum_\alpha \int \mathrm{d}x\, K_{\alpha\beta}(x,y;\theta)\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}. \end{aligned} \tag{80}$$

Thus $\phi'[g(x-\theta)]\,\mathrm{d}\bar{g}_\alpha/\mathrm{d}x$ is the left null eigenvector for $K_{\alpha\beta}$. Again, to be explicit, the discrete equivalent is

$$\mathsf{J} = \begin{pmatrix} \mathsf{K}_{\mathrm{LL}} & \mathsf{K}_{\mathrm{LR}} \\ \mathsf{K}_{\mathrm{RL}} & \mathsf{K}_{\mathrm{RR}} \end{pmatrix}, \qquad \mathbf{f} = \begin{pmatrix} \phi'[\mathbf{g}]\odot\Delta\bar{\mathbf{g}}_{\mathrm{L}} \\ \phi'[\mathbf{g}]\odot\Delta\bar{\mathbf{g}}_{\mathrm{R}} \end{pmatrix}, \tag{81}$$

where $\odot$ represents element-wise (Hadamard) multiplication. Then, $\mathbf{f}$ is the left null eigenvector of J: $0 = \sum_i J_{ij} f_i$.

We now revisit Eq. 67 and assume that g changes such that the bumps slowly move along the attractor

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

manifold:

$$g_\alpha(x,t) \approx g(x-\theta(t)),$$
$$\frac{\mathrm{d}\delta g_\alpha(x,t)}{\mathrm{d}t} = \frac{\mathrm{d}g_\alpha(x,t)}{\mathrm{d}t} \approx -\frac{\mathrm{d}g(x-\theta(t))}{\mathrm{d}x}\frac{\mathrm{d}\theta}{\mathrm{d}t}. \tag{82}$$

Again for simplicity, we will often suppress arguments of derivatives of g and dependence on t. We return to Eq. 70, project it along the left null eigenvector, and apply Eq. 82 to obtain

$$-\tau\frac{\mathrm{d}\theta}{\mathrm{d}t}\sum_\alpha\int \mathrm{d}x\,\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}\frac{\mathrm{d}g}{\mathrm{d}x} = \gamma b\sum_\alpha\int \mathrm{d}x\,(\pm_\alpha 1)\cdot\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x} + \sum_\alpha\int \mathrm{d}x\,\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}\zeta_\alpha(x). \tag{83}$$

The velocity of bump motion is given by $\mathrm{d}\theta/\mathrm{d}t$. It is

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} \approx -\frac{\gamma b\sum_\alpha\int \mathrm{d}x\,(\pm_\alpha 1)\cdot\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha(x-\theta)}{\mathrm{d}x}}{2\tau\int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}g(x-\theta)}{\mathrm{d}x}\right)^2} - \frac{\sum_\alpha\int \mathrm{d}x\,\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha(x-\theta)}{\mathrm{d}x}\zeta_\alpha(x)}{2\tau\int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}g(x-\theta)}{\mathrm{d}x}\right)^2}, \tag{84}$$

where we have made the arguments of g explicit. This equation encapsulates all aspects of bump motion for our theoretical model. It includes dependence on both drive b and noise ζ, the latter of which is kept in a general form. We will proceed by considering specific cases of this equation.

Path integration velocity v_{drive} due to driving input b

The noiseless case of Eq. 84 with $\zeta_\alpha(x) = 0$ yields the bump velocity due to drive b, which is responsible for path integration:

$$v_{\mathrm{drive}} = -\frac{\gamma b\int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}\bar{g}_{\mathrm{R}}}{\mathrm{d}x} - \frac{\mathrm{d}\bar{g}_{\mathrm{L}}}{\mathrm{d}x}\right)}{2\tau\int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}. \tag{85}$$

Note that this expression is independent of the position θ. We can explicitly remove θ by shifting the dummy variable $x \to x+\theta$:

$$\begin{aligned} v_{\mathrm{drive}} &= -\frac{\gamma b\int \mathrm{d}x\,\phi'[g(x)]\left(\frac{\mathrm{d}g(x+\xi)}{\mathrm{d}x} - \frac{\mathrm{d}g(x-\xi)}{\mathrm{d}x}\right)}{2\tau\int \mathrm{d}x\,\phi'[g(x)]\left(\frac{\mathrm{d}g(x)}{\mathrm{d}x}\right)^2} \\ &\approx -\frac{\gamma b\xi\int \mathrm{d}x\,\phi'[g(x)]\frac{\mathrm{d}^2 g}{\mathrm{d}x^2}}{\tau\int \mathrm{d}x\,\phi'[g(x)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}. \end{aligned} \tag{86}$$

Now let's consider the specific ReLU activation function ϕ. Equation 35 implies

$$\phi'[g] = \begin{cases} 0 & g \le 0 \\ 1 & g > 0, \end{cases} \qquad \text{so} \quad \phi'[g]^2 = \phi'[g] \quad \text{and} \quad \phi'[g]\cdot\phi[g] = \phi[g]. \tag{87}$$

These identities, along with the definition for s (Eq. 34), give

$$\phi'[g(x)]\frac{\mathrm{d}^2 g}{\mathrm{d}x^2} = \frac{\mathrm{d}^2 s}{\mathrm{d}x^2}, \qquad \phi'[g(x)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2 = \left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2, \qquad \phi[g(x)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2 = s(x)\left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2. \tag{88}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Applying the first two equalities to Eq. 86 produces Eq. 8 of the Results section.

Now we reintroduce noise ζ and assume it is independent across neurons and timesteps, with mean $\langle\zeta\rangle$. If we average Eq. 84 over ζ, the numerator of the second term becomes

$$\sum_\alpha \int \mathrm{d}x\, \phi'[g(x-\theta)] \frac{\mathrm{d}\bar{g}_\alpha(x-\theta)}{\mathrm{d}x} \langle\zeta\rangle = 0. \tag{89}$$

The integral vanishes because g is even and $\sum_\alpha \mathrm{d}\bar{g}_\alpha/\mathrm{d}x$ is odd. Thus,

$$\left\langle \frac{\mathrm{d}\theta}{\mathrm{d}t} \right\rangle = v_{\mathrm{drive}}, \tag{90}$$

demonstrating that networks with independent noise still path integrate on average.

Diffusion D_{input} due to input noise

Independent noise ζ produces diffusion, a type of deviation in bump motion away from the average trajectory. It is quantified by the diffusion coefficient D:

$$\left\langle \left[\theta(t) - \langle\theta(t)\rangle\right]^2 \right\rangle = 2Dt. \tag{91}$$

In terms of derivatives of θ,

$$\left\langle \left[\theta(t) - \langle\theta(t)\rangle\right]^2 \right\rangle = \int_0^t \int_0^t \mathrm{d}t'\, \mathrm{d}t'' \left\langle \left(\frac{\mathrm{d}\theta}{\mathrm{d}t'} - \left\langle \frac{\mathrm{d}\theta}{\mathrm{d}t'} \right\rangle \right) \left(\frac{\mathrm{d}\theta}{\mathrm{d}t''} - \left\langle \frac{\mathrm{d}\theta}{\mathrm{d}t''} \right\rangle \right) \right\rangle. \tag{92}$$

Equations 84 and 90 imply

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} - \left\langle \frac{\mathrm{d}\theta}{\mathrm{d}t} \right\rangle = - \frac{\sum_\alpha \int \mathrm{d}x\, \phi'[g(x-\theta)] \frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x} \zeta_\alpha(x)}{2\tau \int \mathrm{d}x\, \phi'[g(x-\theta)] \left(\frac{\mathrm{d}g}{\mathrm{d}x} \right)^2}. \tag{93}$$

We then shift the dummy variable $x \to x + \theta(t)$ and reintroduce explicit dependence on t to obtain

$$\left\langle \left[\theta(t) - \langle\theta(t)\rangle\right]^2 \right\rangle = \int_0^t \int_0^t \mathrm{d}t'\, \mathrm{d}t'' \frac{\sum_{\alpha\beta} \iint \mathrm{d}x\, \mathrm{d}y\, \phi'[g(x)]\phi'[g(y)] \frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x} \frac{\mathrm{d}\bar{g}_\beta}{\mathrm{d}y} \left\langle \zeta_\alpha\big(x + \theta(t'), t'\big) \zeta_\beta\big(y + \theta(t''), t''\big) \right\rangle}{4\tau^2 \left[\int \mathrm{d}x\, \phi'[g(x)] \left(\frac{\mathrm{d}g}{\mathrm{d}x} \right)^2 \right]^2}. \tag{94}$$

One class of independent ζ is Gaussian noise added to the total synaptic input, which represents neural fluctuations at short timescales. We assume it is independent across neurons and timesteps with zero mean and fixed variance σ^2:

$$\langle \zeta_\alpha(x,t) \rangle = 0, \qquad \langle \zeta_\alpha(x,t) \zeta_\beta(y,t') \rangle = \sigma^2 \Delta t\, \delta(t-t') \delta_{\alpha\beta} \delta(x-y). \tag{95}$$

Δt is the simulation timestep, which defines the rate at which the random noise variable is resampled.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Equation 94 then becomes, with the help of Eq. 78,

$$\left\langle \left[\theta(t) - \langle\theta(t)\rangle\right]^2 \right\rangle = \int_0^t \mathrm{d}t' \frac{\sigma^2 \Delta t \sum_\alpha \int \mathrm{d}x\, \phi'[g(x)]^2 \left(\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}\right)^2}{4\tau^2 \left[\int \mathrm{d}x\, \phi'[g(x)] \left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2\right]^2}$$

$$\approx \frac{\sigma^2 \Delta t \int \mathrm{d}x\, \phi'[g(x)]^2 \left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}{2\tau^2 \left[\int \mathrm{d}x\, \phi'[g(x)] \left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2\right]^2} \cdot t. \tag{96}$$

Reconciling this with the definition of the diffusion coefficient D in Eq. 91 yields

$$D_{\text{input}} = \frac{\sigma^2 \Delta t \int \mathrm{d}x\, \phi'[g(x)]^2 \left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}{4\tau^2 \left[\int \mathrm{d}x\, \phi'[g(x)] \left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2\right]^2}. \tag{97}$$

Applying Eq. 88 for a ReLU ϕ gives Eq. 10 of the Results section.

Diffusion D_{spike} due to spiking noise

Instead of input noise, we consider independent noise arising from spiking neurons. In this case, the stochastic firing rate s is no longer the deterministic expression in Eq. 34. Instead,

$$s_\alpha(x,t) = \frac{c_\alpha(x,t)}{\Delta t}, \tag{98}$$

where c is the number of spikes emitted in a simulation timestep of length Δt. We model each $c_\alpha(x,t)$ as an independent Poisson-like random variable driven by the deterministic firing rate $\phi[g_\alpha(x,t)]$ with Fano factor F. It has mean $\phi[g_\alpha(x,t)]\Delta t$ and variance $F\phi[g_\alpha(x,t)]\Delta t$. Therefore,

$$s_\alpha(x,t) = \phi[g_\alpha(x,t)] + \sqrt{\frac{F\phi[g_\alpha(x,t)]}{\Delta t}}\eta_\alpha(x,t), \tag{99}$$

where each $\eta_\alpha(x,t)$ is an independent random variable with zero mean and unit variance:

$$\langle \eta_\alpha(x,t)\rangle = 0, \qquad \langle \eta_\alpha(x,t)\eta_\beta(y,t')\rangle = \Delta t\, \delta(t-t')\delta_{\alpha\beta}\delta(x-y). \tag{100}$$

As in Eq. 95, the simulation timestep Δt defines the rate at which η is resampled. By substituting Eq. 99 into Eq. 33, we see that spiking neurons can be described by deterministic firing rate dynamics with the stochastic noise term

$$\zeta_\alpha(x,t) = \sum_\beta \int \mathrm{d}y\, W_\beta(x,y) \sqrt{\frac{F\phi[g_\beta(y,t)]}{\Delta t}}\eta_\beta(y,t). \tag{101}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Now we calculate the diffusion coefficient produced by this noise. Equation 93 becomes

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} - \left\langle \frac{\mathrm{d}\theta}{\mathrm{d}t} \right\rangle = -\frac{\sum_{\alpha\beta} \iint \mathrm{d}x\,\mathrm{d}y\, W_\beta(x,y)\phi'[g(x-\theta)]\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}\sqrt{\frac{F\phi[g(y-\theta)]}{\Delta t}}\eta_\beta(y)}{2\tau\int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}$$
$$= -\frac{\sum_{\beta} \int \mathrm{d}y\,\frac{\mathrm{d}\bar{g}_\beta}{\mathrm{d}y}\sqrt{\frac{F\phi[g(y-\theta)]}{\Delta t}}\eta_\beta(y)}{2\tau\int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}. \tag{102}$$

We used Eq. 79 to obtain the second equality. We then proceed as for input noise to calculate

$$\left\langle \left[\theta(t) - \langle\theta(t)\rangle\right]^2 \right\rangle = \int_0^t\int_0^t \mathrm{d}t'\,\mathrm{d}t'' \frac{\frac{F}{\Delta t}\sum_{\alpha\beta}\iint \mathrm{d}x\,\mathrm{d}y\,\sqrt{\phi[g(x)]\phi[g(y)]}\frac{\mathrm{d}\bar{g}_\alpha}{\mathrm{d}x}\frac{\mathrm{d}\bar{g}_\beta}{\mathrm{d}y}\left\langle \eta_\alpha\big(x+\theta(t'),t'\big)\eta_\beta\big(y+\theta(t''),t''\big)\right\rangle}{4\tau^2\left[\int \mathrm{d}x\,\phi'[g(x)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2\right]^2}, \tag{103}$$

which yields the diffusion coefficient

$$D_{\text{spike}} = \frac{F\int \mathrm{d}x\,\phi[g(x)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2}{4\tau^2\left[\int \mathrm{d}x\,\phi'[g(x)]\left(\frac{\mathrm{d}g}{\mathrm{d}x}\right)^2\right]^2}. \tag{104}$$

After applying Eq. 88 for a ReLU ϕ and setting $F = 1$ for Poisson spiking, we obtain Eq. 20 of the Results section.

Drift velocity $v_{\text{conn}}(\theta)$ due to quenched connectivity noise

Suppose that we perturb the symmetric, translation-invariant W by a small component V representing deviations away from an ideal attractor architecture:

$$W_\beta(x,y) \to W_\beta(x,y) + V_{\alpha\beta}(x,y). \tag{105}$$

By Eq. 33, this produces the noise term

$$\zeta_\alpha(x,t) = \sum_\beta \int \mathrm{d}y\, V_{\alpha\beta}(x,y)\phi[g_\beta(y,t)]. \tag{106}$$

In contrast to input and spiking noise, this noise is correlated across neurons and time, so it cannot be averaged away as in Eqs. 89 and 90. Substituting Eq. 106 into Eq. 84, we obtain

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = v_{\text{drive}} + v_{\text{conn}}(\theta), \tag{107}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where the drift velocity is

$$v_{\text{conn}}(\theta) = -\frac{\sum_{\alpha\beta} \iint \mathrm{d}x\,\mathrm{d}y\, V_{\alpha\beta}(x,y)\phi'[g(x-\theta)]\frac{\mathrm{d}g(x-\theta)}{\mathrm{d}x}\phi[g(y-\theta)]}{2\tau \int \mathrm{d}x\,\phi'[g(x-\theta)]\left(\frac{\mathrm{d}g(x-\theta)}{\mathrm{d}x}\right)^2}. \tag{108}$$

We have made the dependence on bump position θ explicit to illustrate how it influences $v_{\text{conn}}(\theta)$. After applying Eq. 88 for a ReLU ϕ, we obtain Eq. 24 of the Results section.

We now make scaling arguments for speed difference (Eq. 30), speed variability (Eq. 31), and escape drive b_0 (Eq. 26). To do so, we impose a ReLU ϕ and return to discrete variables to be explicit:

$$v_{\text{conn};\theta} = -\frac{\sum_{\alpha\beta}\sum_{ij} V_{\alpha\beta ij}\cdot \Delta s_{i-\theta}\cdot s_{j-\theta}}{2\tau\sum_i (\Delta s_{i-\theta})^2}. \tag{109}$$

We need to understand how the numerator scales with M and N. It is a weighted sum of $4N^2$ independent Gaussian random variables $V_{\alpha\beta ij}$ and is thus a Gaussian random variable itself. It has zero mean, but its variance is proportional to $N^2\cdot M^2/N^2$. The N^2 comes from the number of terms in the sum and the M^2/N^2 comes from the scaling of $\mathrm{d}s/\mathrm{d}x$ (Eq. 11). In combination with the scaling of the denominator, we conclude that $v_{\text{conn};\theta}$ is a Gaussian random variable with

$$\mathrm{E}[v_{\text{conn};\theta}] = 0, \qquad \mathrm{Var}[v_{\text{conn};\theta}] \propto \frac{N^2}{M^2}. \tag{110}$$

Equation 109 implies that $v_{\text{conn};\theta}$ is correlated over θ. The weights for the sum over $V_{\alpha\beta ij}$ are the firing rates and their derivatives for a bump centered at θ. If θ is slightly changed, almost the same entries of V will be summed over with similar weights. The amount of correlation across θ is determined by the degree of overlap in weights, and therefore, by the width and number of bumps. Let's consider the effects of changing N and M on the covariance matrix $\mathrm{Cov}[v_{\text{conn};\theta}, v_{\text{conn};\theta'}]$. A larger N increases the bump width and the correlation length proportionally, so values of the main diagonal decay proportionally more slowly into the off diagonals. A larger M redistributes values among the diagonals by decreasing the bump width and adding more bumps, but it does not change the total amount of correlation. Thus,

$$\sum_{\theta,\theta'} \mathrm{Cov}[v_{\text{conn};\theta}, v_{\text{conn};\theta'}] \propto N^2\cdot \mathrm{Var}[v_{\text{conn};\theta}]. \tag{111}$$

This allows us to evaluate

$$\mathrm{Var}\Big[\operatorname*{mean}_{\theta} v_{\text{conn};\theta}\Big] = \mathrm{Var}\left[\frac{1}{N}\sum_{\theta} v_{\text{conn};\theta}\right] = \frac{1}{N^2}\sum_{\theta,\theta'}\mathrm{Cov}[v_{\text{conn};\theta}, v_{\text{conn};\theta'}] \propto \frac{N^2}{M^2}. \tag{112}$$

As a sum of zero-mean Gaussian random variables, $\mathrm{mean}_\theta\, v_{\text{conn};\theta}$ is also a zero-mean Gaussian random variable. That means $|\mathrm{mean}_\theta\, v_{\text{conn};\theta}|$ follows a folded normal distribution, which obeys

$$\mathrm{E}\left[\left|\operatorname*{mean}_{\theta} v_{\text{conn};\theta}\right|\right] = \sqrt{\frac{2}{\pi}\mathrm{Var}\Big[\operatorname*{mean}_{\theta} v_{\text{conn};\theta}\Big]} \propto \frac{N}{M}. \tag{113}$$

Combining this with Eqs. 12 and 14 produces the scalings for speed difference in Eq. 32.

We now study speed variability, which involves the expression

$$\operatorname*{std}_{\theta} v_{\text{conn};\theta} = \sqrt{\frac{1}{N}\sum_{\theta} v_{\text{conn};\theta}^2}. \tag{114}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Since each $v_{\text{conn};\theta}$ is Gaussian, the sum of their squares follows a generalized chi-square distribution. Its mean is the trace of the covariance matrix $\text{Cov}[v_{\text{conn};\theta}, v_{\text{conn};\theta'}]$, which is equal to N times the variance. Thus, by Eq. 110,

$$\text{E}\left[\frac{1}{N}\sum_{\theta} v_{\text{conn};\theta}^2\right] = \frac{1}{N}\cdot N \cdot \text{Var}[v_{\text{conn};\theta}] \propto \frac{N^2}{M^2}. \tag{115}$$

We are interested in the square root of the random variable on the left-hand side, and we anticipate its expected value to scale as the square root of the right-hand side. We can make this argument precise. Suppose H is a random variable with a probability distribution function $p(h)$ that scales with a power of the parameter B. We can write

$$p(h) = B^n\, P(B^m h) \tag{116}$$

for exponents n and m, where the rescaled probability distribution function P does not scale with B. Conservation of total probability implies

$$B^n \int \text{d}h\, P(B^m h) = B^n B^{-m} \int \text{d}h'\, P(h') = 1. \tag{117}$$

Thus, $m = n$. Next, suppose we know that $\text{E}[H] \propto B^o$:

$$\text{E}[H] = B^n \int \text{d}h\, h\, P(B^n h) = B^{-n} \int \text{d}h'\, h'\, P(h') \propto B^o. \tag{118}$$

Thus, $n = -o$. We can now conclude that $\text{E}[\sqrt{H}] \propto \sqrt{\text{E}[H]}$:

$$\text{E}[\sqrt{H}] = B^{-o} \int \text{d}h\, \sqrt{h}\, P(B^{-o} h) = B^{o/2} \int \text{d}h'\, \sqrt{h'}\, P(h') \propto B^{o/2}. \tag{119}$$

Applying this result to Eq. 115, we obtain

$$\text{E}\Big[\operatorname*{std}_{\theta} v_{\text{conn};\theta}\Big] = \text{E}\left[\sqrt{\frac{1}{N}\sum_{\theta} v_{\text{conn};\theta}^2}\right] \propto \sqrt{\text{E}\left[\frac{1}{N}\sum_{\theta} v_{\text{conn};\theta}^2\right]} \propto \frac{N}{M}. \tag{120}$$

Combining this with Eqs. 12 and 14 produces the scalings for speed variability in Eq. 32.

The escape drive b_0 involves the expression $\max_\theta |v_{\text{conn};\theta}|$. Extreme value statistics for correlated random variables is generally poorly understood. We follow Majumdar et al. (2020) and provide a heuristic argument for its scaling. We can partition $v_{\text{conn};\theta}$ across θ into groups that are largely independent from one another based on its correlation structure. As discussed above, $v_{\text{conn};\theta}$ is a weighted sum of independent Gaussian random variables $V_{\alpha\beta ij}$ (Eq. 109). The weights are products between the firing rates $s_{j-\theta}$ and their derivatives $\Delta s_{i-\theta}$ for a configuration centered at position θ. If we choose two θ's such that bumps do not overlap, the corresponding $v_{\text{conn};\theta}$'s will sum over different $V_{\alpha\beta ij}$'s and will be independent. Thus, λ/z roughly sets the number of independent components, where λ is the bump distance and z is the bump width. This ratio does not change with M or N in our networks (Fig. 2F), so the maximum function does not change the scaling of $|v_{\text{conn};\theta}|$:

$$\max_{\theta} |v_{\text{conn};\theta}| \propto |v_{\text{conn};\theta}|. \tag{121}$$

The scaling of $\text{E}[|v_{\text{conn};\theta}|]$ can be determined from $\text{Var}[v_{\text{conn};\theta}]$ through arguments similar to those made in Eqs. 116, 117, 118, and 119. Suppose we know that $\text{Var}[H] \propto B^o$ and $\text{E}[H] = 0$. Then,

$$\text{Var}[H] = B^n \int \text{d}h\, h^2\, P(B^n h) = B^{-2n} \int \text{d}h'\, (h')^2\, P(h') \propto B^o. \tag{122}$$

Thus, $n = -o/2$. We can now conclude that $\text{E}[|H|] \propto \sqrt{\text{Var}[H]}$:

$$\text{E}[|H|] = B^{-o/2} \int \text{d}h\, |h|\, P(B^{-o/2} h) = B^{o/2} \int \text{d}h'\, |h'|\, P(h') \propto B^{o/2}. \tag{123}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Applying this result to Eq. 121, we obtain

$$\mathrm{E}\Big[\max_{\theta} |v_{\mathrm{conn};\theta}|\Big] \propto \mathrm{E}\big[|v_{\mathrm{conn};\theta}|\big] \propto \sqrt{\mathrm{Var}\big[v_{\mathrm{conn};\theta}\big]} \propto \frac{N}{M}. \tag{124}$$

Combining this with Eqs. 12, 13, and 26 produces the scalings for the escape drive b_0 in Eq. 27.

Simulation methods

Dynamics and parameter values

To simulate the dynamics Eq. 33, we discretize the network by replacing neural position x with index i and propagate forward in time with the simple Euler method:

$$\tau \frac{g_{\alpha i}(t+\Delta t) - g_{\alpha i}(x,t)}{\Delta t} + g_{\alpha i}(x,t) = \sum_{\beta j} W_{\beta ij} s_{\beta j}(t) + A \pm_{\alpha} \gamma b(t) + \zeta_{\alpha i}(t). \tag{125}$$

We use $\tau = 10\,\mathrm{ms}$. We use $\Delta t = 0.5\,\mathrm{ms}$ and $A = 1$ for all simulations except those with spiking neurons. In the latter case, we use finer timesteps $\Delta t = 0.1\,\mathrm{ms}$ and set $A = 0.1\,\mathrm{ms}^{-1}$. Synaptic inputs g and resting inputs A can be dimensionless for rate-based simulations, but they must have units of rate for spiking simulations. We use $\gamma = 0.1$ for rate-based simulations and $\gamma = 0.01\,\mathrm{ms}^{-1}$ for spiking simulations. In all cases, we run the simulation for 1000 timesteps before recording any data to form the bumps. To achieve the relationship in Eq. 13 for circular mapping, we rescale γ with network size N and bump number M:

$$\gamma \to \gamma \cdot \frac{N}{600} \cdot \frac{3}{M}. \tag{126}$$

The connectivity W takes the form in Eq. 38. Unless otherwise specified, we use shift $\xi = 2$. To produce M bumps in a network of size N, we turn to Eq. 47 and set $l = 0.44N/M$. We use $w = 8M/N \approx 3.5/l$. For the case of $2l > N/2$, which corresponds to a one-bump network, the tails of the cosine function extend beyond the network size. Instead of truncating them, we wrap them around the ring:

$$W(x) \to W(x) + W(x-N) + W(x+N). \tag{127}$$

This procedure, along with the scaling of w with N and M, accomplishes Eq. 7 and keeps the total connectivity strength per neuron $\sum_i W_i$ constant across all N and M, where W_i is the discrete form of $W(x)$.

To generate the Poisson-like spike counts $c_{\alpha i}(t)$ in Eq. 98, we rescale Poisson random variables:

$$c_{\alpha i}(t) = F \cdot C_{\alpha i}(t), \qquad C_{\alpha i}(t) \sim \mathrm{Pois}\big[\phi[g_{\alpha i}(t)]\Delta t/F\big]. \tag{128}$$

These counts will be multiples of the Fano factor F. To produce a $c_{\alpha i}(t)$ whose domain is the natural numbers, one can follow Burak and Fiete (2009), who take multiple samples of $C_{\alpha i}(t)$ during each timestep.

To obtain theoretical values in Figs. 3, 5, 7, and 8, we need to substitute the baseline inputs g_i into the appropriate equations. We use noiseless and driveless simulations to generate g_i instead of using Eq. 4.

Bump position

We track the position θ of each bump using the firing rate summed across both populations $S_i(t) = \sum_{\alpha} \phi[g_{\alpha i}(t)]$. We first estimate the positions of all the bumps by partitioning the network into segments of length $\lfloor N/M \rfloor$. If N/M is not an integer, we skip one neuron between some segments to have them distributed as evenly as possible throughout the network. We sum $S_i(t)$ across all the segments and find the position i_0 with maximum value. We perform a circular shift of the original $S_i(t)$ such that i_0 is shifted to the middle of the first segment $\lfloor N/2M \rfloor$. The purpose of this process is to approximately center each bump within a segment so that $S_i(t)$ drops to 0 before reaching segment boundaries. We then calculate the center of mass of $S_i(t)$ within each segment. After reversing the circular shift, these centers of masses are taken to be the bump positions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As an alternative, we can obtain a bump position between 0 and N/M by simply computing the circular mean of $S_i(t)$ with periodicity N/M. However, this method does not track the position of each bump, so we do not use it.

Path integration velocity and diffusion

To obtain our results in Figs. 3 and 5, we run each simulation for $T = 5\,\mathrm{s}$. To extract the bump velocity v produced by a constant drive b, we calculate the mean displacement Θ as a function of time offset u:

$$\Theta(u) = \frac{\Delta t}{T-u} \sum_t \big[\theta(t+u) - \theta(t)\big]. \tag{129}$$

θ is the bump position. This equation averages over fiducial starting times t, which ranges from 0 to $T-u-\Delta t$ in increments of Δt. We vary u between 0 and $T/2$ in increments of Δt; the maximum is $T/2$ to ensure enough t's for accurate averaging. We then fit $\Theta(u)$ to a line through the origin to obtain the velocity:

$$\Theta(u) \approx vu. \tag{130}$$

We calculate the diffusion coefficient D based on an ensemble of replicate simulations. In this section, angle brackets will indicate averaging over this ensemble. Following the definition of D in Eq. 92, we calculate each bump's position relative to the mean motion of the ensemble:

$$\omega(t) = \theta(t) - \langle\theta(t)\rangle \tag{131}$$

We compute squared displacements and then average over fiducial starting times to obtain a mean squared displacement for each bump as a function of time offset u:

$$\Omega(u) = \frac{\Delta t}{T-u} \sum_t \big[\omega(t+u) - \omega(t)\big]^2. \tag{132}$$

t and u span the same time ranges as they did for Θ. We average $\Omega(u)$ over the ensemble and fit it to a line through the origin to obtain the diffusion coefficient:

$$\big\langle\Omega(u)\big\rangle \approx 2Du. \tag{133}$$

For simulations with M bumps, we arbitrarily assign identity numbers $1, \ldots, M$ to bumps in each simulation. We perform ensemble averaging over bumps with the same identity numbers; that is, we only average over one bump per simulation. This way, we obtain separate values for each bump in Fig. 3E–H; nevertheless, these values lie on top of each other. In Fig. 3B, C, each point represents v averaged across bumps. To calculate the mean velocity $\langle v\rangle$ in Fig. 3E, F, we fit $\langle\Theta(u)\rangle$ to a line through the origin. To estimate standard deviations for Fig. 3E–H and Fig. 5, we create 48 bootstrapped ensembles, each of which contains 48 replicate simulations sampled with replacement from the original ensemble. We calculate $\langle v\rangle$ or D for each bootstrapped ensemble and record the resulting standard deviation. In Fig. 5, each point represents D and its estimated standard deviation averaged across bumps.

Trapping and position-dependent velocity

For simulations with connectivity noise, we determine the escape drive b_0 (Fig. 7), the smallest drive that allows the bumps to travel through the entire network, by a binary search over b. We perform 8 rounds of search between the limits 0 and 1.28 and another 8 rounds between 0 and -1.28 to obtain b_0 within an accuracy of 0.01. In each round, we run a simulation with the test b and see whether the bumps travel through the network or get trapped. Traveling through the network means that every position (rounded to the nearest integer) has been visited by a bump, and trapping means that the motion of at least one bump slows below a threshold for a length of time.

To obtain the position-dependent bump velocity $v(\theta)$ produced by connectivity noise when $|b| > b_0$, we run a simulation until the bumps have traveled through the network. At each timestep, we record the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

positions of the bumps (binned to the nearest integer) and their instantaneous velocities with respect to the previous timestep. We smooth the velocities in time with a Gaussian kernel of width 10 ms, which is the neural time constant τ. We calculate the mean and standard deviation of these smoothed velocities for each position bin.

Mutual information

For simulations with input noise, we explore the mutual information between encoded coordinate and single-neuron activity (Fig. 6). To do so, we must generate data from which we can calculate $p(s|u)$ in Eq. 22, for coordinate $u \in \mathcal{U}$ and activity $s \in \mathcal{S}$. We have chosen one set of conditions for performing this analysis, which we detail below.

We first choose to represent either a linear or circular coordinate, which we take to be position or orientation, respectively. We then choose to represent a narrow or wide coordinate range $u_{\max}$, which is 20 cm or 200 cm for position and 36° or 360° for orientation. We divide the range into 20 equally spaced coordinates such that $\mathcal{U} = \{u_{\max}/20, \ldots, u_{\max}\}$. We convert these coordinates to network positions according to the mappings in Fig. 4. For each coordinate value u, we initialize 96 replicate simulations at the corresponding network position by applying additional synaptic input to the desired bump positions during bump formation. We run the simulations for 5 s, record the final firing rates, and bin them using 6 equally spaced bins from 0 to the 99th percentile across all neurons. All rates above the 99th percentile are also added to the 6th bin. These bins define the discrete $\mathcal{S}$, and normalizing the bin counts produces $p(s|u)$. We marginalize over u to obtain $p(s)$, and $p(u)$ is uniform. We can then use Eq. 22 to calculate the mutual information.

The 4 local cues in Fig. 6F–H correspond to 4 activity states $\mathcal{S}_{\text{cue}}$ separate from the 6 activity bins of the CAN neurons, $\mathcal{S}_{\text{neuron}}$. The joint sample space of a single neuron with cues is thus $\mathcal{S} = \mathcal{S}_{\text{neuron}} \times \mathcal{S}_{\text{cue}}$ with $6 \times 4 = 24$ total states. We bin neural activity across these more numerous states, using the coordinate value u to determine the cue state value, to again calculate $p(s|u)$ and then the mutual information.

We choose to calculate mutual information with single-neuron activities binned into 6 discrete states due to computational tractability. A better indication of encoding quality for the entire network would involve using the joint activity of multiple neurons. However, assuming the same binning process, that would involve estimating probability distributions over 6^n states for n neurons, which would require exponentially more replicate simulations per coordinate value than the 96 we use. Alternatively, one could reduce the dimensionality of the network activity by projecting it onto various attractor configurations, as done by Roudi and Treves (2008).

Appendix

In this Appendix, we revisit many major results for input, spiking, and connectivity noise, but for either a different activation function ϕ (Fig. 9) or for connectivity strengths W that do not scale with bump number and network size (Fig. 10). To calculate theoretical predictions for each set of results, we need to substitute the baseline synaptic inputs g into the appropriate equations. They are obtained by running simulations without noise and drive. Notably, the theory still demonstrates close agreement with simulation results under these new conditions.

In Fig. 9, we use a logistic sigmoid activation function ϕ to convert synaptic inputs g to firing rates s:

$$\phi[g] = \frac{1}{1 + \mathrm{e}^{-g}}. \tag{134}$$

All results with this ϕ are qualitatively identical to those obtained with a ReLU ϕ in the Results section. To calculate theoretical values, we can no longer use equations from the Results, which are simplified for a ReLU ϕ. To calculate D_{input}, D_{spike}, v_{drive}, and $v_{\text{conn}}(\theta)$, we use Eqs. 97, 104, 86, and 108 from the Theoretical methods section instead.

In the Results section, we assumed that the connectivity strengths W obey Eq. 7 to maintain the same scaled bump shape across bump numbers M and network sizes N. In addition to the theoretical advantages of obtaining simple scaling relationships, this choice can be loosely biologically motivated. Consider the tuning curves of grid cells, which are thought to function as CANs. Their scaled shapes are roughly similar

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

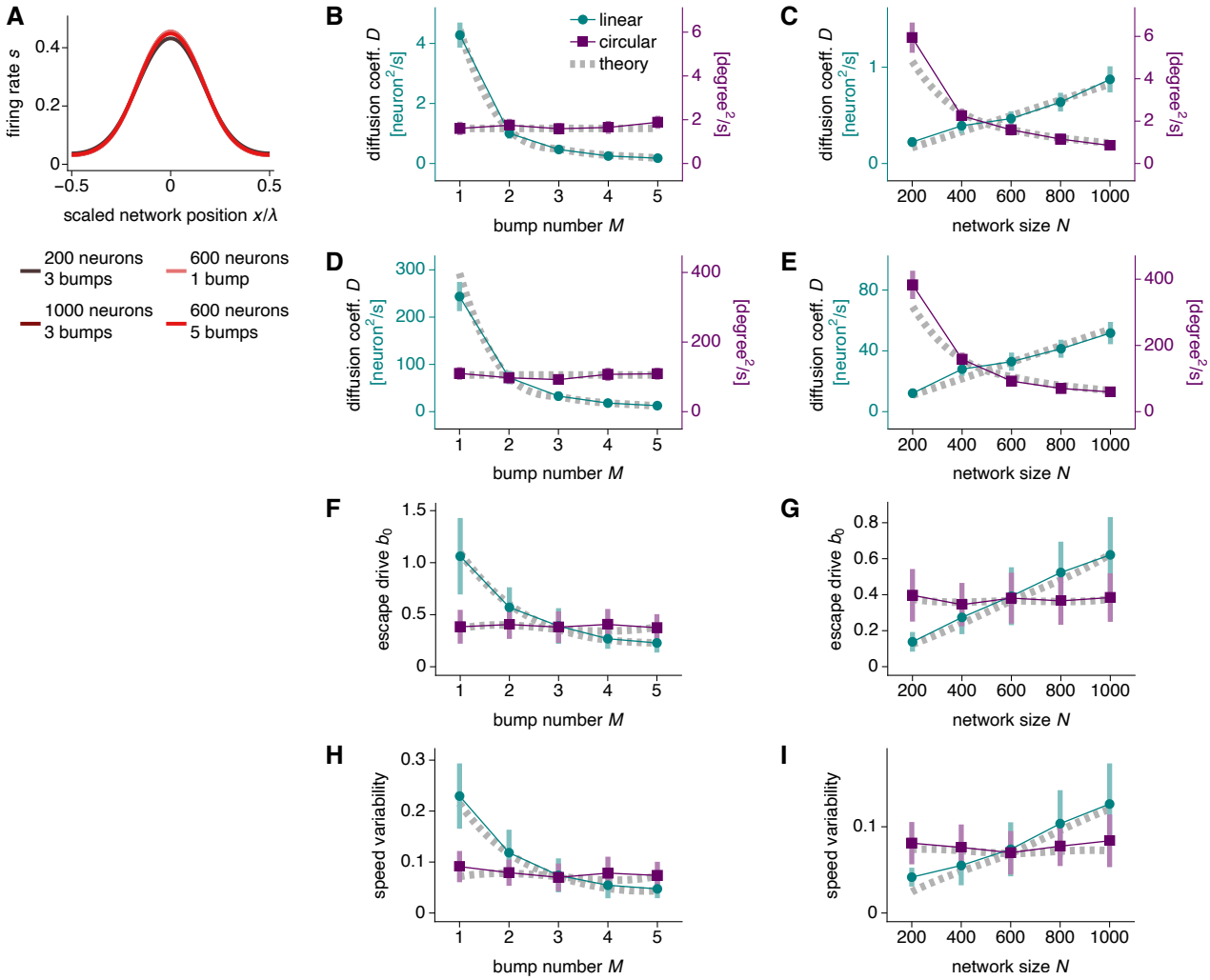

Figure 9: Main results repeated for networks with a logistic activation function. (**A**) The scaled bump shape remains invariant across network sizes and bump numbers, accomplished by rescaling connectivity strengths according to Eq. 7. Curves for different parameters lie over one another. (**B, C**) Networks with synaptic input noise. Bump diffusion follows the same qualitative behavior as in Fig. 5A, B. (**D, E**) Networks with Poisson spiking noise. Bump diffusion follows the same qualitative behavior as in Fig. 5C, D. (**F–I**) Networks with connectivity noise. (**F, G**) Escape drive follows the same qualitative behavior as in Fig. 7D, E. (**F, G**) Bump speed variability follows the same qualitative behavior as in Fig. 8D, E. The activation function ϕ takes the form in Eq. 134. In **F–I**, we use connectivity noise of magnitude 0.003. In **H**, **I**, we use drive $b = 2.5$. The rest of the parameters are identical in value to those used in the main text.

across modules (Stensola et al., 2012), which may differ in bump number (Gu et al., 2018; Kang and Balasubramanian, 2019; Khona et al., 2022), and across mammalian taxa from rodents to primates (Killian et al., 2012; Jacobs et al., 2013), whose brains certainly differ in neuron number. This crude observation supports the choice to maintain a fixed scaled bump shape across M and N. Nonetheless, in Fig. 10, we do not assume Eq. 7 and bump shape invariance. Instead, we fix $w = 0.04$ in Eq. 38, which fixes the maximum synaptic strength across all networks. This change produces qualitative differences only for circular mapping. Here, under circular mapping, networks with fewer bumps are more robust to all three forms of noise, and larger networks are more robust to connectivity noise. For the corresponding simulations in the Results section, no major dependence on bump number was observed.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

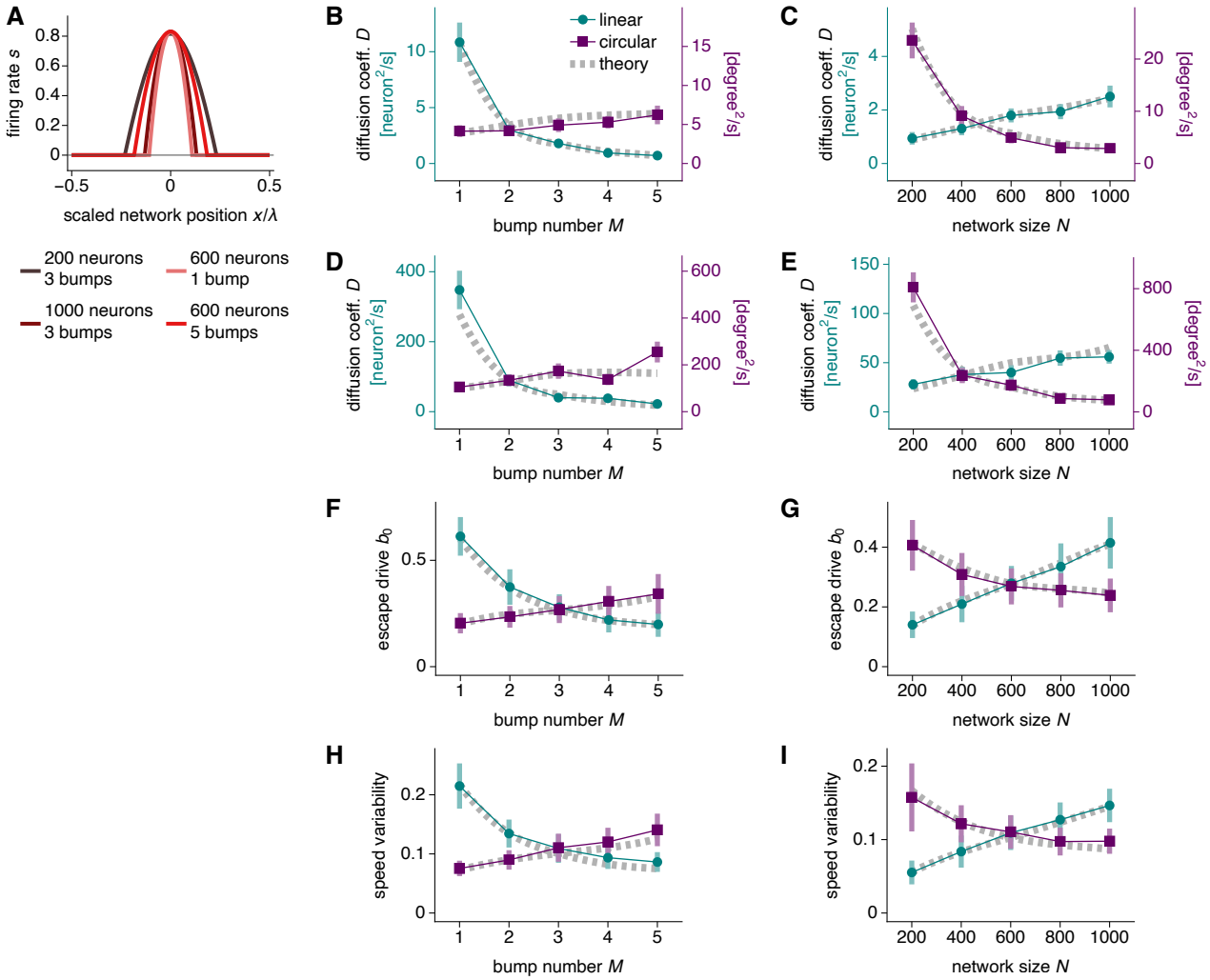

Figure 10: Main results repeated for networks without rescaling of connectivity strengths according to Eq. 7. (**A**) The scaled bump shape no longer remains invariant across network sizes and bump numbers. (**B, C**) Networks with synaptic input noise. Bump diffusion follows the same qualitative behavior as in Fig. 5A, B, except that here it slightly increases with bump number under circular mapping. (**D, E**) Networks with Poisson spiking noise. Bump diffusion follows the same qualitative behavior as in Fig. 5C, D, except that here it slightly increases with bump number under circular mapping. (**F–I**) Networks with connectivity noise. (**F, G**) Escape drive follows the same qualitative behavior as in Fig. 7D, E under linear mapping. It slightly increases with bump number and decreases with network size under circular mapping. (**H, I**) Bump speed variability follows the same qualitative behavior as in Fig. 8D, E under linear mapping. It slightly increases with bump number and decreases with network size under circular mapping. The connectivity W still takes the form in Eq. 38, except that here we fix $w = 0.04$ across all bump numbers and network sizes. In **H**, **I**, we use drive $b = 1.0$. The rest of the parameters are identical in value to those used in the main text.

Acknowledgments

We are grateful to Steven Lee for sharing his code and to John Widloski for a careful reading of this manuscript.

References

R. Almeida, J. Barbosa, and A. Compte. Neural circuit basis of visuo-spatial working memory precision: a computational and behavioral study. *J. Neurophysiol.*, 114(3):1806–1818, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S.-i. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.*, 27(2): 77–87, 1977.

R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 92(9):3844–3848, 1995.

F. Bouchacourt and T. J. Buschman. A flexible model of working memory. *Neuron*, 103(1):147–160.e8, 2019.

C. D. Brody, R. Romo, and A. Kepecs. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr. Opin. Neurobiol.*, 13(2):204–211, 2003.

Y. Burak and I. R. Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLOS Comput. Biol.*, 5(2):e1000291, 2009.

Y. Burak and I. R. Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl. Acad. Sci. U.S.A.*, 109(43):17645–17650, 2012.

D. Bush and N. Burgess. A hybrid oscillatory interference/continuous attractor network model of grid cell firing. *J. Neurosci.*, 34(14):5065–5079, 2014.

T. Can and K. Krishnamurthy. Emergence of memory manifolds. *arXiv* 2109.03879, 2021.

S. C. Cannon, D. A. Robinson, and S. Shamma. A proposed neural network for the integrator of the oculomotor system. *Biol. Cybern.*, 49(2):127–136, 1983.

A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex*, 10(9):910–923, 2000.

C. Constantinidis and X.-J. Wang. A neural circuit basis for spatial working memory. *Neuroscientist*, 10(6): 553–565, 2004.

F. Edin, T. Klingberg, P. Johansson, F. McNab, J. Tegnér, and A. Compte. Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences*, 106(16):6802–6807, 2009.

G. B. Ermentrout and J. D. Cowan. A mathematical theory of visual hallucination patterns. *Biol. Cybern.*, 34(3):137–150, 1979.

I. R. Fiete, Y. Burak, and T. Brookings. What grid cells convey about rat location. *J. Neurosci.*, 28(27): 6858–6871, 2008.

M. C. Fuhs and D. S. Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.*, 26(16):4266–4276, 2006.

C. C. A. Fung, K. Y. M. Wong, and S. Wu. A moving bump in a continuous manifold: a comprehensive study of the tracking dynamics of continuous attractor neural networks. *Neural Comput.*, 22(3):752–792, 2010.

R. J. Gardner, L. Lu, T. Wernle, M.-B. Moser, and E. I. Moser. Correlation structure of grid cells is preserved during sleep. *Nat. Neurosci.*, 22(4):598–608, 2019.

R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, 2022.

J. Goldstone. Field theories with Superconductor solutions. *Nuovo Cimento*, 19(1):154–164, 1961.

J. P. Goodridge, P. A. Dudchenko, K. A. Worboys, E. J. Golob, and J. S. Taube. Cue control and head direction cells. *Behav. Neurosci.*, 112(4):749–761, 1998.

J. Green, A. Adachi, K. K. Shah, J. D. Hirokawa, P. S. Magani, and G. Maimon. A neural circuit architecture for angular integration in Drosophila. *Nature*, 546(7656):101–106, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Y. Gu, S. Lewallen, A. A. Kinkhabwala, C. Domnisoru, K. Yoon, J. L. Gauthier, I. R. Fiete, and D. W. Tank. A map-like micro-organization of grid cells in the medial entorhinal cortex. *Cell*, 175(3):736–750, 2018.

A. Guanella, D. Kiper, and P. Verschure. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.*, 17(04):231–240, 2007.

T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

K. Hardcastle, S. Ganguli, and L. M. Giocomo. Environmental boundaries as an error correction mechanism for grid cells. *Neuron*, 86(3):827–839, 2015.

J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U.S.A.*, 81(10):3088–3092, 1984.

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

J. Jacobs, C. T. Weidemann, J. F. Miller, A. Solway, J. F. Burke, X.-X. Wei, N. Suthana, M. R. Sperling, A. D. Sharan, I. Fried, and M. J. Kahana. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat. Neurosci.*, 16(9):1188 – 1190, 2013.

L. Kang and V. Balasubramanian. A geometric attractor mechanism for self-organization of entorhinal grid modules. *eLife*, 8:e46687, 2019.

L. Kang and M. R. DeWeese. Replay as wavefronts and theta sequences as bump oscillations in a grid cell attractor network. *eLife*, 8:e46351, 2019.

M. Khona, S. Chandra, and I. R. Fiete. From smooth cortical gradients to discrete modules: spontaneous and topologically robust emergence of modularity in grid cells. *bioRxiv* 2021.10.28.466284, 2022.

N. J. Killian, M. J. Jutras, and E. A. Buffalo. A map of visual space in the primate entorhinal cortex. *Nature*, 425(7426):184, 2012.

Z. P. Kilpatrick and B. Ermentrout. Wandering bumps in stochastic neural fields. *SIAM J. Appl. Dyn. Syst.*, 12(1):61–94, 2013.

S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman. Ring attractor dynamics in the Drosophila central brain. *Science*, 356(6340):849–853, 2017.

N. Krishnan, D. B. Poll, and Z. P. Kilpatrick. Synaptic efficacy shapes resource limitations in working memory. *J. Comput. Neurosci.*, 44(3):273–295, 2018.

J. Krupic, M. Bauza, S. Burton, C. Lever, and J. O'Keefe. How environment geometry affects grid cell symmetry and what we can learn from it. *Philos. Trans. R. Soc. B*, 369(1635):20130188, 2014.

C. K. Machens, R. Romo, and C. D. Brody. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science*, 307(5712):1121–1124, 2005.

S. N. Majumdar, A. Pal, and G. Schehr. Extreme value statistics of correlated random variables: A pedagogical review. *Phys. Rep.*, 840:1–32, 2020.

A. Mathis, A. V. M. Herz, and M. Stemmler. Optimal population codes for space: grid cells outperform place cells. *Neural Comput.*, 24(9):2280–2317, 2012.

B. L. McNaughton, L. L. Chen, and E. J. Markus. Dead reckoning, landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. *J. Cognit. Neurosci.*, 3(2):190–202, 1991.

B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser. Path integration and the neural basis of the 'cognitive map'. *Nat. Rev. Neurosci.*, 7(8):663–678, 2006.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

J. Milnor. On the concept of attractor. *Commun. Math. Phys.*, 99(2):177–195, 1985.

N. Mosheiff and Y. Burak. Velocity coupling of grid cell modules enables stable embedding of a low dimensional variable in a high dimensional neural attractor. *eLife*, 8:e48494, 2019.

Y. Nambu. Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.*, 117(3): 648–663, 1960.

Z. Navratilova, L. M. Giocomo, J.-M. Fellous, M. E. Hasselmo, and B. L. McNaughton. Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics. *Hippocampus*, 22(4):772–789, 2012.

J. O'Keefe and N. Burgess. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus*, 15(7):853–866, 2005.

A. Renart, P. Song, and X.-J. Wang. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38(3):473–485, 2003.

Y. Roudi and A. Treves. Representing where along with what information in a model of a cortical patch. *PLOS Comput. Biol.*, 4(3):e1000012, 2008.

A. Samsonovich and B. L. McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.*, 17(15):5900–5920, 1997.

A. Seeholzer, M. Deger, and W. Gerstner. Stability of working memory in continuous attractor networks under the control of short-term plasticity. *PLOS Comput. Biol.*, 15(4):e1006928, 2019.

J. D. Seelig and V. Jayaraman. Neural dynamics for landmark orientation and angular path integration. *Nature*, 521(7551):186–191, 2015.

H. S. Seung. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U.S.A.*, 93(23):13339–13344, 1996.

W. E. Skaggs, J. J. Knierim, H. S. Kudrimoti, and B. L. McNaughton. A model of the neural basis of the rat's sense of direction. *Adv. Neural Inf. Process. Syst.*, 7:173–180, 1995.

D. C. Somers, S. B. Nelson, and M. Sur. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, 15(8):5448–5465, 1995.

B. Sorscher, G. Mel, S. Ganguli, and S. Ocko. A unified theory for the origin of grid cells through the lens of pattern formation. *Adv. Neural Inf. Process. Syst.*, 32:10003–10013, 2019.

S. Sreenivasan and I. Fiete. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat. Neurosci.*, 14(10):1330–1337, 2011.

M. Stemmler, A. Mathis, and A. V. M. Herz. Connecting multiple spatial scales to decode the population activity of grid cells. *Sci. Adv.*, 1(11):e1500816–e1500816, 2015.

H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.-B. Moser, and E. I. Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72–78, 2012.

S. M. Stringer, T. P. Trappenberg, E. T. Rolls, and I. E. T. d. Araujo. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Netw. Comput. Neural Syst.*, 13(2):217–242, 2002.

S. M. Stringer, E. T. Rolls, and T. P. Trappenberg. Self-organising continuous attractor networks with multiple activity packets, and the representation of space. *Neural Networks*, 17(1):5–27, 2004.

S. M. Stringer, E. T. Rolls, and T. P. Trappenberg. Self-organizing continuous attractor network models of hippocampal spatial view cells. *Neurobiol. Learn. Mem.*, 83(1):79–92, 2005.

J. S. Taube. The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.*, 30(1): 181–207, 2007.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.22.481545; this version posted June 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

J. S. Taube, R. U. Muller, and J. B. Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.*, 10(2):420–435, 1990.

K. Thurley, C. Leibold, A. Gundlfinger, D. Schmitz, and R. Kempter. Phase precession through synaptic facilitation. *Neural Comput.*, 20(5):1285–1324, 2008.

M. Tsodyks and T. Sejnowski. Associative memory and hippocampal place cells. *Int. J. Neural Syst.*, 6: S81–S86, 1995.

D. Turner-Evans, S. Wegener, H. Rouault, R. Franconville, T. Wolff, J. D. Seelig, S. Druckmann, and V. Jayaraman. Angular velocity integration in a fly heading circuit. *eLife*, 6:e04577, 2017.

X.-X. Wei, J. Prentice, and V. Balasubramanian. A principle of economy predicts the functional architecture of grid cells. *eLife*, 4:e08362, 2015.

Z. Wei, X. J. Wang, and D. H. Wang. From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *J. Neurosci.*, 32(33):11228 – 11240, 2012.

J. Widloski. Grid cell attractor networks: development and implications. *Doctoral dissertation, University of Texas at Austin*, 2015.

J. Widloski, M. P. Marder, and I. R. Fiete. Inferring circuit mechanisms from sparse neural recording and global perturbation in grid cells. *eLife*, 7:e33503, 2018.

K. Wimmer, D. Q. Nykamp, C. Constantinidis, and A. Compte. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.*, 17(3):431–439, 2014.

S. Wu, S.-i. Amari, and H. Nakahara. Population coding and decoding in a neural field: a computational study. *Neural Comput.*, 14(5):999–1026, 2002.

S. Wu, K. Hamaguchi, and S.-i. Amari. Dynamics and computation of continuous attractors. *Neural Comput.*, 20(4):994–1025, 2008.

X. Xie, R. H. R. Hahnloser, and H. S. Seung. Double-ring network model of the head-direction system. *Phys. Rev. E*, 66(4):041902, 2002.

K. Yoon, M. A. Buice, C. Barry, R. Hayman, N. Burgess, and I. R. Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.*, 16(8):1077–1084, 2013.

K. Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *J. Neurosci.*, 16(6):2112–2126, 1996.