Fundamental limits on persistent activity in networks of noisy neurons

Yoram Burak*bc and Ila R. Fiete*dc

*Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem 91904, Israel; bRacah Institute of Physics, Hebrew University, Jerusalem 91904, Israel; cCenter for Brain Science, Harvard University, Cambridge, MA 02138; and dCenter for Learning and Memory, University of Texas at Austin, Austin, TX 78712

Edited by Terrence J. Sejnowski, Salk Institute for Biological Studies, La Jolla, CA, and approved September 14, 2012 (received for review October 23, 2011)

Neural noise limits the fidelity of representations in the brain. This limitation has been extensively analyzed for sensory coding. However, in short-term memory and integrator networks, where noise accumulates and can play an even more prominent role, much less is known about how neural noise interacts with neural and network parameters to determine the accuracy of the computation. Here we analytically derive how the stored memory in continuous attractor networks of probabilistically spiking neurons will degrade over time through diffusion. By combining statistical and dynamical approaches, we establish a fundamental limit on the accuracy of the computation. This result takes the form of an information-diffusion inequality. We derive some unexpected consequences: Despite the persistence of memory, it does not pay to accumulate spikes for longer than the cellular time-constant to read out their contents. For certain neural transfer functions, the conditions for optimal sensory coding coincide with those for optimal storage, implying that short-term memory may be co-localized with sensory representation.

Short-term memory is essential for survival in a world where data sources may disappear before consequent decisions or actions are taken. In many cases, the object of the memory is a continuous variable, for example, the direction of heading during navigation, the location of the eyes, or the orientation of a visual stimulus. Short-term memory is associated with patterns of neural firing that correlate with the stored quantity (1, 2) and persist during the storage period (3–5). When the stored variable is continuous, the persistent activity can take a continuum of states (1, 2, 5, 6).

Such persistent activity is thought, in many cases, to possess an underlying continuous attractor structure (7–10): the instantaneous value of a continuous variable is represented by setting the system to a point on a continuous manifold of stable fixed points. The stability of the fixed points generates a long time-constant for memory, even when cellular time-constants are short.

Moreover, continuous attractors provide some degree of resistance to ongoing noise: Any perturbation of the state of the system away from the attractor manifold is corrected because the state quickly decays back to the attractor. However, the component of noise parallel to the attractor manifold will remain uncorrected, causing loss of information about the stored memory (9, 11). Such noise-induced drift has been observed in numerical simulations of continuous attractor networks (10, 12, 13) and has been studied analytically in a specific continuous attractor network (14, 15). However, a general understanding of how neural and network properties affect the rate of drift and short-term memory performance has been missing.

In this work, we consider the dynamics of neural networks with a continuous attractor and otherwise arbitrary connectivity, weights, and neural nonlinearity. First, we use a theory of weak fluctuations in large networks to derive analytically how, in the absence of any external anchoring input, stochastic neural activity causes the represented variable to drift. We show that this drift is diffusive. We derive an expression quantifying the rate of memory loss—which for symmetric networks takes a particularly simple form—in terms of the number of neurons, the range of the stored variable, the neural noise, the neural nonlinearity, and the shape of the tuning curves within the network.

Second, we establish a fundamental relationship between a dynamical property, the rate of drift of the memory state due to internal noise in the network; and an abstract statistical limit on estimation, which bounds the information an external observer can gain about the underlying state of the stochastic network from a finite sample of its spikes. We show that the statistical limit on estimation enforces a finite minimum rate of diffusion in the network dynamics.

Third, we derive an unexpected result on the time required to read out the contents of continuous attractor memory networks: Despite the stochastic nature of spiking and the persistence of activity in the network, there is no advantage for an ideal observer in accumulating spikes beyond the biophysical time-constant of single neurons. Fourth, our results imply that in certain cases, the conditions for good sensory representation coincide with the conditions for good short-term memory.

It has long been realized that in physical systems, a deep relationship exists between thermodynamics and memory (16–19). More recently, such principles have been explored in man-made computing devices at the nanoscale, where fluctuations become prominent (19, 20). It is of interest to understand whether similar general principles relate internal noise in nonequilibrium biological systems to the dynamics of information and memory. Our results, relating stochastic dynamics in neural networks to memory degradation, establish that a statistical measure of internal noise constrains the dissipation of memory.

Results

We consider a network of Poisson spiking neurons: Each neuron $i$ generates Poisson spikes based on its time-varying firing rate $r_i(t) = \phi \left\{ \sum W_{ij} s_j(t) + b_i \right\}$, where $\sum W_{ij} s_j$ is the summed recurrent input, $b_i$ is a bias, and $\phi$ is the nonlinear neural transfer function. Spikes in neuron $j$ drive synaptic activation $s_j(t)$ (Eq. 7) which, weighted by the synaptic strength $W_{ji}$, is the input to postsynaptic neuron $i$. The Poisson model is motivated by the Poisson-like statistics of spike trains in cortical areas (21, 22) and is a simple way of generating the spiking statistics that would be obtained under certain conditions from conductance-based neurons receiving noisy inputs (23, 24).

Author contributions: Y.B. and I.R.F. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence may be addressed. E-mail: ila.fiete@mail.clm.utexas.edu or yoram.burak@elsc.huji.ac.il.

This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073pnas.1117386109126Supplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1117386109

PNAS Early Edition | 1 of 6
results, we assume that each neuron receives a large number of spikes within its cellular time-constant, so that network dynamics can be reduced to rate equations, Eq. 10, with weak fluctuations arising from the variability of the Poisson process (referred to as the weak fluctuation limit, Materials and Methods).

We are interested in networks whose rate dynamics admit a continuous manifold of stable steady states (attractors), \( \vec{r}(\theta) \), parametrized by a continuous variable \( \theta \). For simplicity of presentation, we describe below our results for a one-dimensional manifold, but also show (SI Appendix) that they readily generalize to higher dimensions. At each attractor state, the synaptic activation obeys

\[
\delta_i(\theta) = \tau \phi \left( \sum_j W_{ij} \delta_j(\theta) + b_i \right) = \tau \vec{r}_i(\theta),
\]

where \( \tau \) is the synaptic time-constant.

In what follows, we consider the network evolving under its own noisy dynamics, Eqs. 7–9 in Materials and Methods. The instantaneous network state \( s(t) \) remains close to the attractor manifold, but is not precisely on it because of continuously generated neural noise (Fig. 1A). Nevertheless, the instantaneous state can be mapped to a single point \( \theta(t) \) on the low-dimensional attractor manifold, defined as the point to which the state would flow under noise-free rate dynamics. We refer to this point as the instantaneous attractor state.

**Diffusive Drift due to Ongoing Neural Noise.** We seek to derive how noisy neural spiking drives the instantaneous attractor state to drift. The effect of noise is to perturb the network away from the attractor manifold. Most components of this perturbation are rapidly erased by flow to the attractor manifold, and are thus irrelevant, but the component parallel to the attractor manifold is not and is responsible for gradual drift of the instantaneous state along the attractor. In general, noise can introduce two types of drift along the attractor: The first is random, and the second is systematic. We assume below that the system is well-tuned to eliminate systematic drift (for discussion of this assumption, see SI Appendix, section II) and focus on the random component because it is responsible for memory degradation and information loss. We use the neural network dynamics [Eqs. 7–9] to obtain that at short times, the variance of the attractor state, relative to its initial value, grows linearly with the elapsed time:

\[
\langle \theta(t) - \theta(0) \rangle^2 = 2 \Delta t / D(\theta, D).
\]

A priori, internal FI appears unconnected to the network’s diffusive dynamics, but as we will show in the next section, it is in fact intimately related to the diffusion coefficient.

The expression for internal FI appears to have the same mathematical form as the Fisher information derived for sensory coding (27, 28), but only if \( \vec{r}_i \) represents the stimulus-evoked response of the network and the variable \( \theta \) is the value of a present

---

*The result for symmetric networks is of special interest because many continuous attractor network models involve symmetric or nearly symmetric weights (7, 8, 11).*
Fig. 2. Numerical validation of the diffusivity result in different attractor networks. (A) Ring network: Schematic of a memory network representing a periodic variable such as orientation. Every neuron $i$ has an associated angle $\theta_i$, and the synaptic weight between any pair of neurons $i, j$ depends only on $\theta_i - \theta_j$. Attractor states are labeled by an angular variable $\theta$. (B) Simulation of a 1,024 neuron network, with exponential neural transfer function. Inset: snapshots of population activity at two times separated by 10 s. Horizontal axis: neural indices along the ring. Solid traces: firing rate $r$ (peak firing rates are approximately 85 Hz). Vertical lines: spikes emitted during a 10 ms interval. Main plot: random drift of the attractor state. (C, D) Mean squared displacement of the attractor state over elapsed time intervals $\Delta t$ (variance is computed by taking differences in $\theta$ modulo $2\pi$ and confining them to the range $(-\pi, \pi]$). Circles: numerical simulations with exponential (gray) or sigmoidal (black) neural transfer function (see Inset). Solid lines: theoretical prediction of Eq. 2. Dashed lines: bound on $\sqrt{2}$ from the information-diffusion inequality for sigmoidal nonlinearity. Error bars in C and D are s.e.m. over 100 trials (s.e.m. smaller than symbols in C). (E) Mutual-inhibition network. Schematic of a memory network representing a nonperiodic 1-d variable. Two neurons or two populations, inhibit each other with equal weights. The difference of the two firing rates represents the stored variable. (F) Random drift of the firing rates (in 1/ms) in the mutual-inhibition network. (G) Mean squared displacement of the attractor state over an elapsed time $\Delta t$. Circles: simulation of mutual-inhibition network, with linear transfer function. Solid line: theoretical prediction of Eq. 2. See SI Appendix, section III for a model network with a state-dependent diffusion coefficient.

**Information-Diffusion Inequality.** Equipped with the expressions for the diffusion coefficient (Eq. 2, and SI Appendix, Eq. S15) and the internal FI (Eq. 3), we obtain the second main result of this paper, the information-diffusion inequality (proof in SI Appendix, section V for arbitrary connectivity and neural nonlinearity):

$$D \geq \frac{1}{2\pi^2} J^{-1}. \tag{4}$$

In words, the internal FI sets a lower bound on the diffusion of the attractor state.

The information-diffusion inequality can be interpreted as follows. In maintaining its persistent activity, the network effectively acts as an estimator of its state along the attractor, based on the spikes emitted within a time $\tau$. Under this interpretation, Eq. 4 follows from a pair of inequalities: First, an ideal observer estimating the instantaneous attractor state from spike-counts has a squared error that equals or exceeds the inverse of the internal FI, according to the well-known Cramér–Rao bound. Second, the accuracy of the neural network in estimating its own instantaneous attractor state can be no better than, and may be significantly worse than, the ideal estimator. These combined limits on the network’s estimation of its instantaneous attractor state constrain its dynamics to be diffusive with a lower bound on the diffusion coefficient provided by the internal FI. The inequality contains an overall factor of $\tau^2$ because over the characteristic time-scale $\tau$, the internal FI is proportional to $\tau$, as is the squared displacement from diffusion.

**Generalizations.** The information-diffusion inequality holds under a number of generalizations. First, for multidimensional attractors, the inequality retains the same form as in Eq. 4, with $\mathcal{D}$ referring to the diffusion coefficient tensor and $J^{-1}$ representing the inverse of the internal FI matrix. The inequality signifies that the difference between the two sides is a positive-semidefinite matrix (SI Appendix, section V). Second, when the neural dynamics involve more than one time-scale, for instance a synaptic time-constant and a membrane time-constant, the same information-diffusion inequality [Eq. 4] holds. However, $\tau$ is replaced by the sum of the two time-scales. Thus, the larger of the two time-scales dominates in setting the limit on diffusivity (SI Appendix, section VII). Third, when the Poisson-like noise in the dynamics is replaced by additive Gaussian noise (of constant variance), then the conditions under which the information-diffusion inequality is saturated change (see below), but Eq. 4 remains the same (SI Appendix, section VII).

**Saturation of the information-diffusion inequality.** Here we consider whether the information-diffusion inequality is tight, i.e., whether the bound on diffusivity given by the inequality is ever saturated. The nested inequalities contained within our interpretation of the information-diffusion inequality, above, must become equalities for the bound to be saturated. In the weak fluctuation limit—where many spikes are emitted within the cellular time-constant—it is guaranteed that the Cramér–Rao bound can be saturated by an ideal estimator (27, 28). However, it is much less clear whether any continuous attractor network could match the performance of an ideal estimator in tracking its own instantaneous attractor state. It has been shown previously that continuous attractor networks, when evolving deterministically (noise free), can estimate the location of a bump input from a sample of spikes in a Bayes-optimal way (30). The memory network faces a more difficult task because it needs to continuously estimate a state that is drifting, and its dynamics as a readout network are noisy.

Surprisingly, a large class of networks can match an ideal observer in estimating their own evolving attractor state. Using Eqs. 3 and 4 it is straightforward to verify that in symmetric networks, if the neural transfer function is exponential, $\phi(x) \propto \exp(ax)$, then diffusion equals the inverse internal FI over the time-scale $\tau$, and the information-diffusion inequality is saturated. Thus, an exponential transfer function enables a symmetric network to optimally estimate its own state. In the remainder of this subsection we comment on networks with symmetric connectivity.

What aspect of the network dynamics dictates that an exponential nonlinearity is optimal for self-state estimation? If the Poison noise model for neural response is replaced by white Gaussian noise with fixed variance, the expressions for $\mathcal{D}$ and $J$ change, but the information-diffusion inequality remains the same (SI Appendix, section VII). However, the inequality is saturated with a linear rather than exponential neural transfer.

---

1Because the internal FI is defined as the rate of Fisher information over an infinitesimally short time interval, there is no diffusion over such an interval. Thus, $\mathcal{D}$ does not contribute to $J$ (but $J$ bounds $\mathcal{D}$).

2The converse is also true: if the information-diffusion inequality is saturated, the transfer function is exponential (SI Appendix, section V).
function (SI Appendix, section VII). Therefore, the requirement of an exponential transfer function for saturation of the inequality is related to the Poisson variability in neural spiking.

We note that saturation of the information-diffusion inequality is not equivalent to minimizing diffusivity. As just shown, saturating the information-diffusion inequality constrains the neural nonlinearity but not the network weights, whereas minimizing the diffusivity could involve optimizing the weights as well.

Nevertheless, in networks with exponential neural nonlinearity, diffusivity can be minimized by maximizing internal FI. Two interesting consequences follow from the fact that internal FI has the same mathematical dependence on the tuning curves as the Fisher information for sensory representation when the internal attractor state variable \( \Theta(t) \) is replaced by the stimulus value \( \Theta \).

First, the extensive literature on optimal tuning curves for sensory representation (28, 31) can be directly applied to identify which network architectures minimize diffusivity. For example, we can conclude from the sensory coding literature (31) that narrower tuning curves should improve memory persistence in a 1-d attractor, but broader tuning curves should improve persistence in 3-d. Second, if the sensory response of the network resembles its input-free tuning curves, as expected for a continuous attractor network (7), it follows that the same network architecture that optimizes sensory representation is also optimal for retaining short-term memories.

**Optimal Observation Time for Memory Readout.** Suppose that a memory network’s initial state was set to \( \Theta \) on the attractor by a stimulus that was then removed at \( t = 0 \). At a later time \( T \), a cue is presented signaling that the stimulus variable has to be recalled. Over what interval \( \Delta T \) following the cue should a decoder collect spikes from the memory network to obtain an accurate estimate of \( \Theta \)?

In sensory representation, where the sensory stimulus remains present and the aim is to decode \( \Theta \) from the spikes of the encoding neurons, the variance of the ideal estimate drops linearly with the duration over which spikes are collected. Thus, the longer the time over which spikes are collected, the better the estimate of the encoded variable.

In memory networks, the stored variable diffuses not only over the delay period but also over the recall interval, so there is a cost associated with waiting to collect more spikes: The longer the time over which spikes are collected, the more statistics are gained about the recent state of the network, but the memory of the input has also further dissipated. An ideal estimator (SI Appendix, section VI) of \( \Theta \) can continue to gain some information along the attractor is independent of the attractor state (i.e., if \( \Delta T \) is fixed, performance degrades due to diffusion of the memory network starting at time \( T \)). The first term in the sum on the right-hand side is due to diffusion, and the second arises from the memory state along the attractor, with the diffusion coefficient given by Eq. 5. Dashed trace: Performance of a naive estimator (SI Appendix, Eq. S40), which assumes that the attractor state is fixed in time, is similar to that of the ideal estimator for small \( \Delta T \). For \( \Delta T > \Delta T_{\text{naive}} \), performance degrades due to diffusion of the attractor state. The optimal readout time of this naive estimator is \( \Delta T_{\text{naive}} = \sqrt{3\Delta T} \) (dotted vertical line). For both estimators, the traces show the variance rescaled by \( \sqrt{2D/\tau} \).

**Minimum Variance of Stimulus Estimation.** Our result on the diffusion of a memory state, Eq. 2, allows us to quantify the expected squared error in recalling, at time \( T \), the initial state \( \Theta \) of the network, set by a stimulus that was removed at \( t = 0 \). For large intervals \( T \), this squared error is dominated by diffusion of the memory state along the attractor, with the diffusion coefficient given by Eq. 2. If the represented variable is unbounded and if diffusion along the attractor is independent of the attractor state (i.e., if the diffusion coefficient is independent of \( \Theta \)), the expected squared error will be at least

\[
\langle |\hat{\Theta}(T) - \Theta|^2 \rangle \geq 2\Theta(T + \tau),
\]

where \( \hat{\Theta}(T) \) is the estimate of \( \Theta \) based on all spikes emitted by the memory network starting at time \( T \). The first term in the sum on the right-hand side is due to diffusion, and the second arises from the uncertainty in estimating the instantaneous attractor state from spikes by an ideal observer (SI Appendix, section VI), as analyzed in the previous section. For bounded variables and for state-dependent diffusion coefficients, the growth of diffusive variance can be similarly obtained, by integrating the diffusion equation (SI Appendix, section III) with the derived value of \( \Delta T_{\text{naive}} \) (Eq. 2 or SI Appendix, Eq. S15). Thus, our analysis bounds how information about a stored variable dissipates over time due to neural noise.

**Discussion**

In summary, we have derived an expression in terms of neural and network parameters for how information in continuous attractor memory networks dissipates over time due to ongoing neural noise and proved that this diffusive information loss is fundamentally lower bounded by a statistical limit on estimation of the instantaneous attractor state by an ideal observer. Our results,

![Fig. 3.](image-url)

*Fig. 3.* Estimation error of stored variable as a function of recall time interval. Solid trace: Variance of an ideal estimator improves monotonically with the recall time (SI Appendix, Eq. S36). The estimator attempts to infer the attractor state at time \( T \) based on all neural spikes in the memory network emitted between time \( T \) and \( T + \Delta T \). Due to diffusion, the estimate does not improve significantly for \( \Delta T > \Delta T_{\text{naive}} \), where \( \Delta T \ll \tau \) [Eq. 5]. Dashed trace: Performance of a naive estimator (SI Appendix, Eq. S40), which assumes that the attractor state is fixed in time, is similar to that of the ideal estimator for small \( \Delta T \). For \( \Delta T > \Delta T_{\text{naive}} \) performance degrades due to diffusion of the attractor state. The optimal readout time of this naive estimator is \( \Delta T_{\text{naive}} = \sqrt{3\Delta T} \) (dotted vertical line). For both estimators, the traces show the variance rescaled by \( \sqrt{2D/\tau} \).*
derived under quite general conditions, are valid for highly non-linear systems, in contrast to existing results on information loss through neural noise in linear memory networks (29, 33). Our results also contrast with existing analytical results on network models of integration and short-term memory, which are predominantly deterministic or focus on memory degradation due to systematic parameter mistuning (8, 9, 25, 26, 34–36).

Neural coding is typically analyzed from the perspective of a downstream recipient of spikes. Our results highlight that the way a variable is coded in a memory network has consequences not only for representational capacity, as viewed from outside the network, but also for the network's own internal dynamics.

Our spiking model for single neurons is quite simple, and more detailed models of neural spiking will doubtless produce somewhat different results for the diffusivity and internal FI. Already, in two extensions of the basic Poisson neuron model (a modified noise model and a model of neurons with two time-constants), the diffusion coefficient and the internal FI had modified forms (SI Appendix, section VII). Interestingly, however, in each case the information-diffusion inequality remained valid (SI Appendix, section VII). The very general nature of the principle underlying the information-diffusion inequality suggests that it may hold for neural models that contain more biophysical detail (24) and include more dynamical time-scales (37); for networks where spike–spike correlations are important and modify the Fisher information (38, 39) and diffusivity; as well as for different kinds of biological systems like gene–protein regulatory networks where stochasticity in the copy number of transcriptional regulators plays the role of Poisson noise (40). Deterministic spiking networks can have chaotic dynamics and produce Poisson-like spike trains (41, 42). If our assumption of Poisson independence holds true in those networks, then our results may also apply to such systems. These conjectures remain to be tested.

Our results suggest that in continuous attractor neural integrators (8–10, 25, 26, 34), neural noise has a similar effect as noise in the input variable being integrated, justifying the phenomenological approach taken by many (25, 26, 43), who add a Gaussian noise term to the input of an otherwise perfect integrator. When this added term represents noise in only the stimulus [e.g., in random-dot motion stimuli (26)], its statistics are known and can be used to set the noise variance. However, in general the variance of added noise should also include the effects of stochasticity in the neural integrator, which is the diffusivity we have computed here. The variance of this term depends on the network structure and the nonlinearity of the integrator neurons, as given in Eq. 2 or SI Appendix, Eq. S15. This intrinsic neural noise reduces the accuracy of the integrator but does not shorten the integration time-constant.

Our results have implications for empirical studies of memory. The result that memory degradation due to neural noise is diffusive, implies that the variance in the recall of analog variables (44, 45), if stored directly in a continuous attractor network, is predicted by the diffusion equation (SI Appendix, Eq. S26), using the diffusion coefficient $\mathcal{D}$. The latter could be derived from our results, Eq. 2 or SI Appendix, Eq. S15, and in addition (or instead) be measured experimentally—by observing the instantaneous growth of recall variance starting from all possible states of the variable. At short times in the tuned system, the variance of recall should increase linearly, with slope $\mathcal{D}$.

The quantitative expression for the diffusion rate derived here can be used to probe to what extent short-term memory errors are due to fundamental neural fluctuations and, thus, whether the system performance is otherwise optimized. For example, it may be possible to estimate how much of the drift during oculomotor fixations is due to neural noise in the oculomotor integrator (46). The diffusion formula of Eq. 2 cannot be directly applied in this case because oculomotor spikes are significantly sub-Poisson. However, the information-diffusion inequality suggests that the internal FI can provide a good estimate of neural noise-driven diffusivity. In turn, at least in the goldfish oculomotor integrator with its well-characterized tuning curves and parameters, it may be possible to compute the internal FI (2, 47). Thus, this work provides a theoretical framework for assessing whether drift during fixations is dominated by integrator noise rather than other sources of variability, such as motor noise.

Using the information-diffusion inequality, we found that estimation of the memory network's state does not significantly improve by waiting to collect spikes over longer intervals than approximately $\tau$. This result solves what would otherwise be a conundrum: If it were profitable to integrate spikes for longer than a single-neuron time-constant, then a memory network would be required to read out the variable stored in another memory network. Instead, a network with no persistence can be near optimal in this task. We predict that our readout performance curves (Fig. 3) should match psychometric curves on recall performance as a function of the length of the postcue recall interval, with $\tau$ likely corresponding to biophysical single-neuron time-scales. Further, whether the measured psychometric curve is found to be monotonically increasing or peaked at the single-neuron time-constant may reveal whether memory readout is based on an ideal estimator or a simple but near-optimal estimator that assumes a static memory state.

We conclude with an intriguing implication of the information-diffusion inequality: With an exponential neural nonlinearity, minimizing diffusivity in symmetric networks is equivalent to maximizing internal FI, which implies that the same conditions optimize both memory performance and sensory representation.5 This suggests that a single brain area could perform both sensory representation and maintain a short-term memory, and thus provides mathematical support for the hypothesis from recent empirical findings (4, 5, 48), that sensory brain areas may be involved in storing the variables they represent.

Materials and Methods

Neural Network. We consider a network of $N$ linear-nonlinear Poisson (LNP) neurons with the following dynamics: The synaptic activation $s_i$ due to spikes produced by neuron $i$ is obtained by convolving the spike train with an exponential function:

$$s_i(t) = \sum_{t} \delta(t - t_i^s),$$  \hspace{1cm} [7]

where $\tau$ is the synaptic time-constant, taken for simplicity to be identical in all synapses, and $t_i^s$ are the spike times. The total synaptic input to neuron $i$ is given by a weighted sum over the synaptic activations of its afferents

$$g_i = \sum_{j} W_{ij} s_j^s + b_i,$$  \hspace{1cm} [8]

where $W_{ij}$ is the weight matrix. The neuron emits spikes as an inhomogeneous Poisson process with rate

$$r_i = \lambda(g_i).$$  \hspace{1cm} [9]

Weak Fluctuations. In the limit in which each neuron is driven by many spikes, Eq. 7 can be replaced by (see SI Appendix, section I)

$$\dot{s}_i = s_i \frac{g_i}{\tau} = r_i + \xi_i,$$  \hspace{1cm} [10]

where $\xi_i$ is a Gaussian process with covariance

$$\langle \xi_i(t) \xi_j(t') \rangle = r_i \delta_{ij} \delta(t - t').$$  \hspace{1cm} [11]

This approximation is justified when the appropriate corresponding fluctuations in $g$, are small compared to the mean. It may seem more justified to introduce the small noise term in $g$ rather than $s$, but formally the two formulations are equivalent (SI Appendix, section I) because $g$ and $s$ are

5The conclusion supposes that internal FI equals the Fisher information in sensory representation, or in other words, that tuning curves with and without external inputs are similar, which is typically true in continuous attractor networks (7).
Finally, the network weights are assumed to be tuned to support a continuous attractor manifold in the noise-free dynamics.¹

Network Simulations. The ring-attractor network of Fig. 2 A–C consisted of N = 1,024 neurons with rotationally invariant inhibitory weights \( W_J = w(\theta_0 - \theta_i) \), where \( \theta_0 = 2\pi/n \) and \( W_J = -A \exp[\lambda \cos(\theta_0 - \theta_i)] - A \exp[\lambda \cos(\theta_0 + \theta_i)] \), with parameters: \( A = 1, x_1 = 1 \) and \( k_x = 0.3 \). The neural transfer functions used in the rate model of Fig. 2A were computed as \( \tanh(\Delta t \tau_0 \theta + \theta^b) \) (shown in Fig. 2C, Inset); \( \theta^b = -2 \) and the synaptic time-constant was \( \tau = 10 \) ms. Synaptic activations \( \dot{\theta} \) were initialized as uniformly distributed random numbers between 0 and 0.01, and the dynamics were simulated for 100 trials lasting 1,000 s each, using a stochastic Euler integrator with time-step \( dt = 0.1 \) ms. At each time-step per trial a spike in neuron \( i \) was generated with probability \( r_i(t) dt \). To produce Fig. 2C, the instantaneous attractor state \( \theta_i(t) \) was measured at each time-step by finding the angle \( \theta_i \) at which \( r_i(t) \) is maximal. Subsequently, \( \dot{\theta}_i(t) = (\dot{\theta}_i(t) - \dot{\theta}_i(t-1)) \) was averaged over the whole simulation, except for the first 10 ms. Here \( [\dot{\theta}_i(t) + \dot{\theta}_i(t-1) / 2] \) defined to be in the range \([-\pi, \pi]\). Values of \( \tau_0 \) and \( J \) were obtained from Eqs. 2, 3 and the steady state numerical solution for \( \theta \) in the noise-free limit. The black trace in Fig. 2D was obtained using the theoretical prediction for \( \theta \) and the analytic expression for the variance of a periodic variable, SI Appendix, Eq. E52.


The two-neuron network of Fig. 2 E–G consisted of \( N = 2 \) neurons coupled with weight matrix: \( W_{11} = W_{22} = 0; W_{12} = W_{21} = -1 \), and common feedforward input \( b = 5,000 \). The synaptic time-constant was \( \tau = 100 \) ms, and synaptic dynamics, given by Eq. 7, were integrated using the Euler method with time-step \( dt = 0.4 \) ms. Neural spiking was generated by an inhomogeneous Poisson process with rate given by \( r(t) dt \), where \( r(t) = \phi(W_1(t) + b) \). The motion transfer functions of the linear, \( \tau \phi(x) = x \). Dynamics were simulated for \( T_{max} = 200 \) s. The two neurons may equivalently be thought of as two neuron groups of \( M \) neurons each, with a neuron from one group connecting to all neurons in the other group through uniform weights \( W/M \). In this case, the firing rate \( r_i(t) \) of the \( i \)th group is simply the sum of the firing rates of all its individual cells. Fig. 2F shows the firing rates \( r_i(t) = g_i(t), r_2(t) = g_2(t) \) of the two neuron groups. The numerical variance \( \sigma^2(T) \) of the firing rates was computed in the noise-free limit. The expectation was taken over the times \( t \in [0, T_{max}] \) in the simulation. The value of \( \sigma^2 \) was obtained from Eq. 2, to be \( \sigma^2 = b \), by inserting the analytical solution for the steady states of neural activity in this two-neuron network: \( 2\tau r = \left( \theta_1(t) - \theta_2(t) \right) \).

ACKNOWLEDGMENTS. We thank Sophie Denève, Bard Ermentrout, and Haim Sompolinsky for helpful conversations and Ishay Mor for comments on the manuscript. Ila Fiete is a Sloan Foundation Fellow, a Searle Scholar, and a McKnight Scholar, and acknowledges funding from the National Science Foundation through NSF-EAGER 1148973.

¹Small tuning correction, of higher order in \( 1/N \), may be needed to eliminate all systematics in the noise system, see SI Appendix, section II.