

Stochastic neural fields as gradient dynamical systems

Paul C. Bressloff and Samuel R. Carroll

Department of Mathematics, University of Utah, Salt Lake City, Utah 84112, USA

(Received 20 April 2019; published 3 July 2019)

Continuous attractor neural networks are used extensively to model a variety of experimentally observed coherent brain states, ranging from cortical waves of activity to stationary activity bumps. The latter are thought to play an important role in various forms of neural information processing, including population coding in primary visual cortex (V1) and working memory in prefrontal cortex. However, one limitation of continuous attractor networks is that the location of the peak of an activity bump (or wave) can diffuse due to intrinsic network noise. This reflects marginal stability of bump solutions with respect to the action of an underlying continuous symmetry group. Previous studies have used perturbation theory to derive an approximate stochastic differential equation for the location of the peak (phase) of the bump. Although this method captures the diffusive wandering of a bump solution, it ignores fluctuations in the amplitude of the bump. In this paper, we show how amplitude fluctuations can be analyzed by reducing the underlying stochastic neural field equation to a finite-dimensional stochastic gradient dynamical system that tracks the stochastic motion of both the amplitude and phase of bump solutions. This allows us to derive exact expressions for the steady-state probability density and its moments, which are then used to investigate two major issues: (i) the input-dependent suppression of neural variability and (ii) noise-induced transitions to bump extinction. We develop the theory by considering the particular example of a ring attractor network with $SO(2)$ symmetry, which is the most common architecture used in attractor models of working memory and population tuning in V1. However, we also extend the analysis to a higher-dimensional spherical attractor network with $SO(3)$ symmetry which has previously been proposed as a model of orientation and spatial frequency tuning in V1. We thus establish how a combination of stochastic analysis and group theoretic methods provides a powerful tool for investigating the effects of noise in continuous attractor networks.

DOI: [10.1103/PhysRevE.100.012402](https://doi.org/10.1103/PhysRevE.100.012402)**I. INTRODUCTION**

One of the fundamental abilities of the brain is to hold information in short term memory. Experimental evidence suggests that one mechanism for short term memory is persistent neural activity that is sustained for a period of time after a stimulus has been removed [1,2]. This persistent state is often represented as an activity bump, which is a localized state of activity in a spatially structured neural network. For example, certain neurons respond to local regions of luminance in the visual field; the stimulus region that elicits the maximum average firing rate is known as the neuron's receptive field. Thus, at a population level, if a stimulus is at location x in the visual field, then cells with receptive fields containing x will have the highest firing rate, whereas the firing rate of cells with receptive fields away from x will decrease as the distance from x increases. In networks responsible for working memory, this localized state persists after the stimulus is removed and the location of the peak of the bump indicates the remembered location [3–7].

Mathematically speaking, short term memory is typically modeled using a so-called bump attractor network, which simultaneously supports a stable uniform rest state and a continuous family of stable bump solutions in the absence of an external input or stimulus. The existence of a family of bump solutions reflects a continuous symmetry of the underlying neural field equations, and allows the network to encode a continuous variable such as spatial location [8].

When an input is presented, it induces a transition from the rest state to one of the bump solutions, which persists when the input is removed, thus maintaining a memory of the input.

Another major application of bump attractor networks is to modeling the formation of population tuning curves in primary visual cortex (V1), which encode stimulus features such as orientation [9–11], spatial frequency [12,13], and texture [14]. The underlying picture is that recurrent excitatory connections amplify weakly biased feed-forward inputs from the thalamus, which are subsequently sculpted by lateral inhibitory connections. The resulting tuning width and other aspects of cortical responses are thus primarily determined by intracortical connections rather than thalamic inputs. The output activity bump is said to amplify the input bias, and provides a network-based encoding of the stimulus that can be processed by downstream networks.

However, one major limitation of continuous attractor networks is that the location of the peak of a bump can diffuse due to intrinsic network noise, resulting in degradation of a working memory trace or a population code in V1 [3,5,15]. This is a natural consequence of the marginal stability of bump solutions with respect to the action of the underlying symmetry group in the absence of an input. One method for analyzing the effects of noise in attractor networks is to use the theory of stochastic neural fields. The latter has received growing attention recently, not only within the context of working memory [16–19], but also with regard to traveling

waves [20–23], binocular rivalry [24], and stimulus-dependent neural variability [25]. In these studies, noise is typically assumed to be weak. Translation symmetry is exploited to write the stochastic neural field, to leading order, as a shifted deterministic bump solution where the shift, or phase, is a stochastic variable. An approximate stochastic differential equation (SDE) for the phase is then derived using perturbation methods. Although such an approach captures the diffusive wandering of a bump solution, it ignores fluctuations in the amplitude of the bump. It is possible to take into account the coupling between amplitude and phase fluctuations by considering higher-order terms in the perturbation expansion, however, the analysis is rather involved and is restricted to weak noise.

In this paper, we follow a different approach by imposing certain constraints on the synaptic weight function and intrinsic noise so that a reduced, but exact, SDE can be derived that tracks the stochastic motion of both the amplitude and phase of bump solutions for arbitrary levels of noise. In particular, the resulting SDE can be expressed as a stochastic gradient dynamical system, which allows us to derive an exact expression for the steady-state probability density and its moments. We develop the theory by considering the particular example of a ring attractor network with $SO(2)$ symmetry, which is the most common architecture used in attractor models of working memory and population tuning in V1. However, we also extend the analysis to a higher-dimensional spherical attractor network with $SO(3)$ symmetry, which has previously been proposed as a model of orientation and spatial frequency tuning in V1 [12,13].

We address two major issues within the context of ring attractor networks. First, we determine how fluctuations in the amplitude and phase of a bump depend on the input. This is directly related to a topic of considerable current interest, namely, understanding the neural mechanisms underlying the suppression of cortical variability following the onset of a stimulus [26–34]. Since trial-by-trial variability and noise correlations are known to affect the information capacity of neurons, such suppression could improve the accuracy of population codes. We show that increasing the input greatly suppresses steady-state phase fluctuations, consistent with a recent study based on perturbation methods [25]. Moreover, depending on parameter values, the input can either have a negligible or a noticeable effect on amplitude fluctuations, but the reduction in variability is much less than that for the phase. Thus, amplitude fluctuations could be significant outside the weak noise regime. Second, in the case of a bistable network, we calculate the mean time for bump extinction due to a noise-induced transition to a zero state or a low amplitude bump. These transitions contribute to the steady-state variance in the amplitude, but would be rare events in the case of weak noise.

The paper is organized as follows. In Sec. II we show how to reduce the deterministic neural field equation for a ring attractor network to a planar gradient dynamical system, which tracks the amplitude and phase of bump solutions. We then use symmetry arguments to investigate the existence and stability of bumps. In Sec. III, we turn to a stochastic version of the neural field equation and the corresponding stochastic gradient system. The latter allows us to compute a steady-

state probability density, which determines the distribution of the phase and amplitude of a stochastically wandering bump solution. We use this to investigate how inputs can suppress amplitude and phase fluctuations. We also compare our exact results to previous results obtained using perturbation theory in the weak noise limit. In Sec. IV we analyze the escape time problem for bump extinction in the presence of noise. Finally, in Sec. V we extend our theory to spherical attractor networks, thus establishing more generally how a combination of stochastic analysis and group theoretic methods provides a powerful tool for investigating the effects of noise in continuous attractor networks.

II. RING ATTRACTOR NETWORK AS A GRADIENT SYSTEM

Consider the deterministic neural field equation

$$\frac{\partial u}{\partial t} = -u + \int_{-\pi}^{\pi} K(\theta - \theta') f(u(\theta', t)) d\theta' + I(\theta), \quad (2.1)$$

where $u(\theta, t)$ represents the activity of neurons with stimulus preference $\theta \in [-\pi, \pi)$, $f(u)$ is a sigmoidal firing rate function

$$f(u) = \frac{1}{1 + e^{-\eta(u-\kappa)}},$$

with gain η and threshold κ , K describes the synaptic strength between neurons with stimulus preferences θ and θ' , and $I(\theta)$ is the external input to the population. Here, θ could represent the direction preference of neurons in area-middle temporal cortex (MT) [31], the orientation preference of V1 neurons, after rescaling $\theta \rightarrow \theta/2$ [9–11], or a coordinate in spatial working memory [3,15,16]. Following Ref. [16], suppose that the 2π -periodic function $K(\theta)$ is even and has the finite cosine series expansion

$$K(\theta) = \sum_{n=0}^N w_n \cos(n\theta).$$

Since u and I are also periodic in θ , they have the corresponding Fourier series expansions

$$u(\theta, t) = \sum_{n=0}^{\infty} [a_n(t) \cos(n\theta) + b_n(t) \sin(n\theta)],$$

$$I(\theta) = \sum_{n=0}^{\infty} [I_n^C \cos(n\theta) + I_n^S \sin(n\theta)].$$

For notational convenience, we will rewrite these equations as

$$u(\theta, t) = \sum_{n=0}^{\infty} \mathbf{x}_n(t) \cdot \mathbf{e}(n\theta), \quad I(\theta) = \sum_{n=0}^{\infty} \mathbf{I}_n \cdot \mathbf{e}(n\theta), \quad (2.2)$$

where

$$\mathbf{x}_n(t) = \begin{bmatrix} a_n(t) \\ b_n(t) \end{bmatrix}, \quad \mathbf{I}_n = \begin{bmatrix} I_n^C \\ I_n^S \end{bmatrix}, \quad \text{and} \quad \mathbf{e}(\theta) = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}.$$

Substituting Eqs. (2.2) into (2.1) and taking inner products with respect to $\cos(n\theta)$ and $\sin(n\theta)$ generates a system of ordinary differential equations (ODEs) for a_n, b_n :

$$\frac{d\mathbf{x}_n(t)}{dt} = -\mathbf{x}_n(t) + \int_{-\pi}^{\pi} w_n f\left(\sum_{m=0}^{\infty} \mathbf{x}_m \cdot \mathbf{e}(m\theta)\right) \mathbf{e}(n\theta) d\theta + \mathbf{I}_n, \quad n \leq N \quad (2.3a)$$

$$\frac{d\mathbf{x}_n(t)}{dt} = -\mathbf{x}_n(t) + \mathbf{I}_n, \quad n > N. \quad (2.3b)$$

Thus, all modes outside the range of the integral operator exponentially approach \mathbf{I}_n . Therefore, for simplicity, we will take initial conditions $\mathbf{x}_n(0) = \mathbf{I}_n$ for $n \geq N + 1$ and focus on the corresponding finite-dimensional space

$$W_N^R = \left\{ \sum_{n=1}^N [a_n \cos(n\theta) + b_n \sin(n\theta)] \right\} \cong \mathbb{R}^{2N}. \quad (2.4)$$

Now, define the scalar function

$$V = \sum_{n=0}^N \frac{1}{w_n} \left[\frac{1}{2} \|\mathbf{x}_n\|^2 - \mathbf{x}_n \cdot \mathbf{I}_n \right] - \int_{-\pi}^{\pi} F\left(\sum_{n=0}^N \mathbf{x}_n \cdot \mathbf{e}(n\theta') + \sum_{n=N+1}^{\infty} \mathbf{I}_n \cdot \mathbf{e}(n\theta')\right) d\theta', \quad (2.5)$$

where

$$F(u) = \int_0^u f(u') du'. \quad (2.6)$$

We can then rewrite Eq. (2.3) as the scaled gradient system

$$\frac{d\mathbf{x}_n}{dt} = -w_n \nabla_{\mathbf{x}_n} V, \quad (2.7)$$

where

$$\nabla_{\mathbf{x}_n} V = \begin{bmatrix} \partial_{a_n} V \\ \partial_{b_n} V \end{bmatrix}.$$

Almost all studies of ring attractor networks take the dominant mode of $K(\theta)$ to be $\cos \theta$ [4,9,11,16]. First, this represents a version of the so-called Mexican hat weight distribution, in which neurons with similar stimulus preferences excite each other, and those with dissimilar stimulus preferences inhibit each other. Second, in an appropriate parameter regime, the resulting neural field equation supports the existence of a stable stationary bump solution, whose dominant mode is of the form $\cos(\theta - \theta_0)$, where $\theta_0 \in [-\pi, \pi)$ is the location of the peak of the bump. In the absence of a stimulus, θ_0 is arbitrary. On the other hand, a stimulus of the form $\bar{I} \cos(\theta - \bar{\Delta})$ can fix the location of the bump such that $\theta_0 = \bar{\Delta}$.

In light of these observations, it is reasonable to take [16]

$$K(\theta) = w \cos(\theta),$$

$$I(\theta) = I^C \cos(\theta) + I^S \sin(\theta) = \bar{I} \cos(\theta - \bar{\Delta}).$$

Within our particular formulation, we then have a traditional planar gradient system

$$\frac{d\mathbf{x}}{dt} = -\nabla V_0(\mathbf{x}), \quad (2.8)$$

where $\mathbf{x} = (a_1, b_1) = (a, b)$,

$$V_0(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|^2 - \mathbf{x} \cdot \mathbf{I} - G(\mathbf{x}), \quad (2.9)$$

with $\mathbf{I} = (I^C, I^S) = \bar{I} \mathbf{e}(\bar{\Delta})$, and

$$G(\mathbf{x}) = w \int_{-\pi}^{\pi} F(\mathbf{x} \cdot \mathbf{e}(\theta)) d\theta. \quad (2.10)$$

It immediately follows that time-dependent solutions are of the form of a time-dependent bump

$$u(\theta, t) = a(t) \cos(\theta) + b(t) \sin(\theta) = A(t) \cos[\theta - \Delta(t)],$$

where we have introduced the polar coordinates $\mathbf{x}(t) = A(t) \mathbf{e}(\Delta(t))$, that is,

$$a(t) = A(t) \cos \Delta(t), \quad b(t) = A(t) \sin \Delta(t), \quad (2.11)$$

with $A(t)$ the amplitude of the bump and $\Delta(t)$ the phase. At the deterministic level, reexpressing the bump dynamics in terms of a gradient system does not yield any major new results regarding the existence and stability of equilibrium solutions. The advantages will become apparent when we look at the stochastic model in Sec. III. However, in order to provide the necessary background for the stochastic analysis, it is useful to rederive previous results concerning existence and stability [16] using the alternative gradient formulation.

A. Homogeneous network

When $I(\theta) = 0$ in the original neural field equation (2.1), the system is $\text{SO}(2)$ equivariant with respect to the action $\varphi u(\theta) = u(\theta - \varphi)$. Since $u(\theta) = \mathbf{x} \cdot \mathbf{e}(\theta)$, it follows that

$$u(\theta - \varphi) = \mathbf{x} \cdot \mathbf{e}(\theta - \varphi) = \mathbf{x} \cdot [R(-\varphi) \mathbf{e}(\theta)] = R(\varphi) \mathbf{x} \cdot \mathbf{e}(\theta),$$

where $R(\varphi)$ is the 2×2 rotation matrix by angle φ , and therefore the induced action on $\mathbf{x} \in \mathbb{R}^2$ is the standard action $\varphi \cdot \mathbf{x} = R(\varphi) \mathbf{x}$. Hence, $V_0(\mathbf{x})$ is $\text{SO}(2)$ invariant when $\mathbf{I} = \mathbf{0}$, so that $\nabla V_0(\mathbf{x})$ commutes with rotations. Consequently the gradient system (2.8) is $\text{SO}(2)$ equivariant with respect to the standard action on \mathbb{R}^2 . Moreover, since equilibria are determined by extremal values of the potential $V_0(\mathbf{x})$ and $V_0(\mathbf{x})$ has radial symmetry, there is a continuum of equilibria constituting a circle. This symmetry implies that we can, without loss of generality, take stationary solutions $\mathbf{x}_0^* = (a^*, 0)$ with $a^* > 0$ satisfying

$$a^* = \mathcal{F}(a^*) \equiv w \int_{-\pi}^{\pi} f(a^* \cos(\theta)) \cos(\theta) d\theta. \quad (2.12)$$

Hence, existence of a stationary bump solution reduces to finding a nontrivial solution to Eq. (2.12). Once the existence of an amplitude a^* has been established, we obtain group orbits of solutions as

$$\mathbf{x}^* = R(\psi) \mathbf{x}_0^* = a^* (\cos(\psi), \sin(\psi)), \quad \forall \psi \in [-\pi, \pi).$$

(Note that we can take $a^* > 0$ since its negative counterpart is obtained by applying a rotation by π .)

Linearizing about some equilibrium solution $\mathbf{x}^* = a^* \mathbf{e}(\psi)$ yields the eigenvalue problem

$$\lambda \mathbf{y} = H \mathbf{y}, \quad \mathbf{y} \in \mathbb{R}^2 \quad (2.13)$$

where H is the Hessian matrix for $V_0(\mathbf{x})$:

$$H_{ij} = -\left. \frac{\partial^2 V_0}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{x}^*} = -\delta_{i,j} + w \int_{-\pi}^{\pi} f'(\mathbf{x}^* \cdot \mathbf{e}(\theta)) e_i(\theta) e_j(\theta) d\theta. \quad (2.14)$$

If $\mathbf{x}^* \neq 0$, then $\text{SO}(2)$ equivariance implies that H always has a zero eigenvalue with corresponding eigenvector $\mathbf{x}^{*\perp}$, the vector orthogonal to the equilibrium \mathbf{x}^* . In order to show this, we first note that if \mathbf{x}^* is an equilibrium, then

$$\nabla V_0(\mathbf{x}^*) = 0 \Rightarrow \nabla V_0(R(s)\mathbf{x}^*) = 0$$

for all s . Differentiating with respect to s and evaluating at $s = 0$ gives

$$0 = -\left. \frac{d}{ds} \nabla V_0(R(s)\mathbf{x}^*) \right|_{s=0} = H R'(0)\mathbf{x}^*.$$

Thus,

$$R'(0)\mathbf{x}^* = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{x}^* \equiv \mathbf{x}^{*\perp} = a^* (-\sin(\psi), \cos(\psi))$$

spans $\text{null}(H)$. Equation (2.14) then shows that

$$-a^* \sin \psi + w \int_{-\pi}^{\pi} f'(a^* \cos(\theta - \psi)) \cos \theta \sin(\theta - \psi) d\theta = 0,$$

which, after exploiting the $\text{SO}(2)$ invariance of the integral, yields the identity

$$w \int_{-\pi}^{\pi} f'(a^* \cos(\theta)) \sin^2(\theta) d\theta = 1. \quad (2.15)$$

The analogous result in the original neural field variable is the following: if $u(\theta)$ is a nonconstant solution, then $u'(\theta)$ is in the null space of the linearization. This reflects the translation symmetry of the system, i.e., $u(\theta)$ is a solution if and only if $u(\theta - \psi)$ is a solution for any $\psi \in [-\pi, \pi)$. Thus, perturbations $u = u_0(\theta) + \varepsilon u'_0(\theta)$ correspond to infinitesimally shifting the phase of $u_0(\theta)$ while perturbations $\mathbf{x} = \mathbf{x}^* + \varepsilon \mathbf{x}^{*\perp}$ correspond to infinitesimally rotating \mathbf{x}^* along a circle.

With this in mind, we can split \mathbb{R}^2 into two eigenspaces

$$\mathbb{R}^2 = \text{span}\{\mathbf{x}^*\} \oplus \text{span}\{\mathbf{x}^{*\perp}\}.$$

The restriction of H to \mathbf{x}^* yields

$$H\mathbf{x}^* = -\mathbf{x}^* + w \int_{-\pi}^{\pi} f'(\mathbf{x}^* \cdot \mathbf{e}(\theta)) \mathbf{e}(\theta) [\mathbf{x}^* \cdot \mathbf{e}(\theta)] d\theta.$$

Hence, taking the dot product of both sides of $\lambda \mathbf{x}^* = H\mathbf{x}^*$ with \mathbf{x}^* and dividing by $\|\mathbf{x}^*\|^2$ yields

$$\begin{aligned} \lambda &= -1 + w \int_{-\pi}^{\pi} f'(\mathbf{x}^* \cdot \mathbf{e}(\theta)) \cos^2(\theta - \psi) d\theta \\ &= -1 + w \int_{-\pi}^{\pi} f'(a^* \cos(\theta)) \cos^2(\theta) d\theta \\ &= -1 + \mathcal{F}(a^*), \end{aligned} \quad (2.16)$$

where the second equality follows from $\text{SO}(2)$ invariance of the integral, and \mathcal{F} is defined in Eq. (2.12). Perturbations along \mathbf{x}^* corresponding to perturbations in the bump

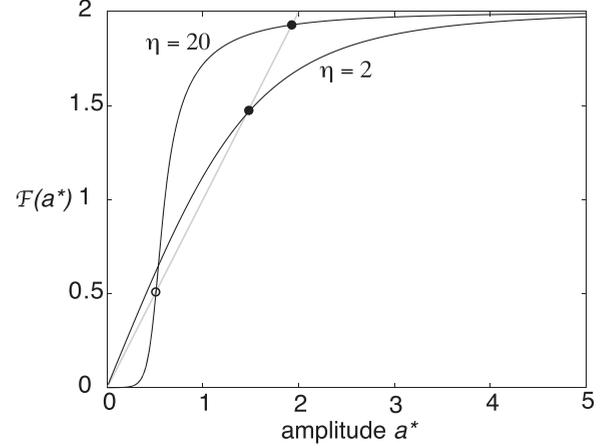


FIG. 1. Plot of nontrivial fixed points of $\mathcal{F}(a^*)$ for both high gain ($\eta = 20$) and low gain ($\eta = 2$). The black curves are the graphs for $y = \mathcal{F}(a^*)$ and the gray curve is the graph for $y = a^*$. Filled circles denote stable equilibria while open circles denote unstable equilibria. High gain supports a bistable system, whereas low gain supports only a single nonzero stable state. Other parameters used are $w = 1$ and $\kappa = 0.5$.

amplitude. Therefore, the equilibrium bump solution will be marginally stable if and only if

$$\mathcal{F}'(a^*) < 1. \quad (2.17)$$

In particular, the stability condition for the trivial solution $a^* = 0$ is $\pi w f'(0) < 1$. Note that if the trivial solution is unstable, $\mathcal{F}'(0) > 1$, then there exists at least one stable nontrivial bump solution. This follows from the observation that $\mathcal{F}(a)$ is continuous and $\lim_{a \rightarrow \infty} \mathcal{F}(a) = 2w$. Therefore, there must exist some $a_0 > 0$ such that $a_0 = \mathcal{F}(a_0)$. If there is more than one solution, let a_0 denote the smallest one. Then, $f'(a_0) < 1$ since $f(a) > a$ for $0 < a < a_0$ and $f(a) < a$ for $a_0 < a$ sufficiently close. Hence, the solution with the smallest magnitude is stable. On the other hand, stability of the trivial solution does not imply anything about existence of nontrivial solutions. It will depend on how large w is. If w is too small, then the trivial solution will be the only equilibrium. On the other hand, if w is sufficiently large, then there will be two equilibria corresponding to bump states, with the larger amplitude being stable. Hence, the system is bistable with a stable rest state and a stable bump state. See Fig. 1 for an illustration of these results.

B. Inhomogeneous network

When $I(\theta) \neq 0$, the system (2.8) is no longer $\text{SO}(2)$ equivariant, but it has other symmetry properties. In particular, if \mathbf{x}^* is a stationary solution when $\mathbf{I} = \mathbf{I}_0$, then $R(\psi)\mathbf{x}^*$ is a stationary solution when $\mathbf{I} = R(\psi)\mathbf{I}_0$. Moreover, we can set $\nabla V_0(\mathbf{x}) = \nabla U_0(\mathbf{x}) - \mathbf{I}$, where

$$U_0(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|^2 - G(\mathbf{x}), \quad (2.18)$$

which is $\text{SO}(2)$ invariant. Thus, stationary solutions satisfy $\nabla U_0(\mathbf{x}^*) = \mathbf{I}$. Radial symmetry of $U_0(\mathbf{x})$ implies that, for each \mathbf{x} , $\nabla U_0(\mathbf{x})$ is collinear with \mathbf{x} and therefore stationary solutions must be collinear with \mathbf{I} . We can therefore, without loss of generality, take $\mathbf{I}_0 = (\bar{I}, 0)^T$ with $\bar{I} > 0$ and, since \mathbf{x}^*

must be collinear, we can take $\mathbf{x}^* = (a^*, 0)^T$. We thus obtain the fixed point equation

$$a^* - \bar{I} = \mathcal{F}(a^*). \quad (2.19)$$

In contrast to the solutions of Eq. (2.12), negative and positive solutions for a^* will no longer be equivalent up to a rotation. A positive solution to Eq. (2.19), $a^* > 0$, will correspond to a bump that is in phase with the input (peak at $\Delta = \bar{\Delta}$), whereas a negative solution, $a^* < 0$, will correspond to an antiphase bump (peak at $\Delta = \bar{\Delta} + \pi$).

The eigenvalue equation is identical to the homogeneous case since \mathbf{I} vanishes after differentiation. The main difference is that, since the inhomogeneous system is not equivariant, the Hessian H no longer trivially has zero eigenvalues. However, we can still make use of the symmetry property that if \mathbf{x}^* is an equilibrium solution when $\mathbf{I} = \mathbf{I}_0$, then $R(\psi)\mathbf{x}^*$ is a solution when $\mathbf{I} = R(\psi)\mathbf{I}_0$. This means that

$$0 = -\nabla U_0(\mathbf{x}^*) + \mathbf{I}_0 \Rightarrow 0 = -\nabla U_0(R(\psi)\mathbf{x}^*) + R(\psi)\mathbf{I}_0$$

for all rotation matrices $R(\psi)$ with $\psi \in [-\pi, \pi)$. Differentiating with respect to ψ and setting $\psi = 0$ then yields

$$H\mathbf{x}^{*\perp} = -\mathbf{I}_0^\perp,$$

where

$$\mathbf{x}^{*\perp} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{x}^* \quad \text{and} \quad \mathbf{I}_0^\perp = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{I}_0.$$

Since \mathbf{x}^* is collinear with \mathbf{I}_0 ,

$$\mathbf{I}_0 = \bar{I}\mathbf{e}(\bar{\Delta}) \Rightarrow \mathbf{x}^* = a^*\mathbf{e}(\bar{\Delta}), \quad \bar{\Delta} \in [-\pi, \pi),$$

with $\bar{I} > 0$ and a^* satisfying Eq. (2.19). Therefore,

$$H\mathbf{x}^{*\perp} = -\frac{I_0}{a^*}\mathbf{x}^{*\perp},$$

showing that $\mathbf{x}^{*\perp}$ is an eigenvector with corresponding eigenvalue

$$\lambda_o = -\frac{\bar{I}}{a^*}.$$

Hence, in-phase bump solutions ($a^* > 0$) will always be stable with respect to phase changing perturbations while antiphase bump solutions ($a^* < 0$) will always be unstable. (In the absence of an input, $\bar{I} = 0$, we recover marginal stability with respect to phase shifts.) As far as we are aware, no one has explicitly calculated λ_o before, with the exception of some specific cases (for example, using a Heaviside firing rate [16]). Interestingly, the general expression does not depend on the form of firing rate function.

The eigenvalue corresponding to the eigenvector \mathbf{x}^* is identical to Eq. (2.16). Since negative solutions $a < 0$ are unconditionally unstable, we focus on stability of $a > 0$ and specifically consider the bistable regime. Let $0 < a_r < a_u < a_s$ denote the three positive equilibria. Note that if a_s exists for $\bar{I} = 0$, then it will exist for all $\bar{I} > 0$. When $\bar{I} = 0$, a_r and a_s are stable while a_u is unstable. Increasing \bar{I} increases the value of a_r and a_s , while it decreases the value of a_u . Since $\mathcal{F}(a)$ saturates, then $\mathcal{F}'(a_s)$ decreases as a_s increases [provided that $\mathcal{F}'(a_s) < 1$ for $\bar{I} = 0$]. Therefore, increasing \bar{I} decreases the amplitude changing eigenvalue and, hence,

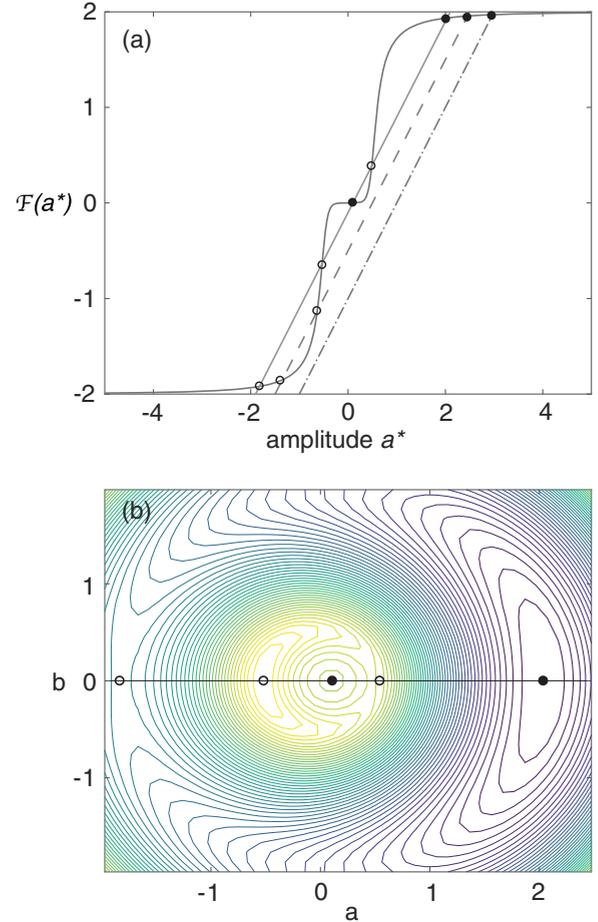


FIG. 2. Equilibrium solutions for the inhomogeneous system. (a) The black curve is the graph for $y = \mathcal{F}(a^*)$ and the gray lines are for $y = a^* - \bar{I}$ and various input amplitudes \bar{I} . The solid gray line is for $\bar{I} = 0.1$, the dashed line for $\bar{I} = 0.5$, and the dashed-dotted line for $\bar{I} = 1$. Filled circles denote stable equilibria while open circles denote unstable equilibria. (b) Contour plot of the potential $V_0(\mathbf{x})$ for $\bar{I} = 0.1$ and $\mathbf{x} = (a, b)$. Other parameters used are $\eta = 20$, $w = 1$, and $\kappa = 0.5$.

a_s remains stable. Similarly, for small enough \bar{I} , a_r remains stable and a_u remains unstable.

At some critical value of \bar{I} , the system loses a_r and a_u (a saddle-node bifurcation occurs). Finally, as \bar{I} increases further, only the stable in-phase bump state a_s remains. See Fig. 2 for an illustration of the effects of input. The change in the number of fixed points works similarly for the case when the homogeneous system only has one stable fixed point (modulo symmetry).

III. STOCHASTIC RING MODEL

Now, consider the stochastic ring model

$$du = \left[-u + \int_{-\pi}^{\pi} K(\theta - \theta') f(u(\theta', t)) d\theta' + I(\theta) \right] dt + dW(\theta, t), \quad (3.1)$$

where dW is a Q -Wiener process (colored in space, white in time) with $\mathbb{E}[dW(\theta, t)] = 0$ and

$$\mathbb{E}[dW(\theta, t)dW(\theta', t')] = C(\theta - \theta')\delta(t - t')dt dt',$$

where $C(\theta)$ is taken to be an even periodic function of the form

$$C(\theta) = \sum_{n=1}^{\infty} \sigma_n \cos(n\theta), \quad \sigma_n > 0.$$

It is common to assume that the correlation function is $SO(2)$ invariant. Moreover, the restriction that C is even and $\sigma_n > 0$ follows from requiring that the corresponding integral operator $\int_{-\pi}^{\pi} C(\theta - \theta')u(\theta')d\theta'$ is symmetric and positive definite on W_N^R . For simplicity, we assume that $\sigma_n = 0$ for $n > M$ for some $M \geq N$. As in the deterministic case, solutions will have the form

$$u(\theta, t) = \sum_{n=0}^M a_n(t) \cos(n\theta) + b_n(t) \sin(n\theta),$$

except that now $a_n(t), b_n(t)$ are stochastic variables. Define the two-dimensional (2D) Wiener processes

$$d\tilde{\mathbf{W}}_n(t) = \int_{-\pi}^{\pi} dW(\theta', t) \begin{bmatrix} \cos(n\theta') \\ \sin(n\theta') \end{bmatrix} d\theta',$$

and set $\mathbf{x}_n = (a_n, b_n)$ as before. The vector-valued Wiener process $d\tilde{\mathbf{W}} = (d\tilde{\mathbf{W}}_1, \dots, d\tilde{\mathbf{W}}_M)$ has the diagonal correlation matrix

$$\mathbb{E}[d\tilde{\mathbf{W}}_n(t)d\tilde{\mathbf{W}}_m(t')^T] = \delta_{n,m}\sigma_n\delta(t - t')dt dt'.$$

Thus, we can write

$$d\tilde{\mathbf{W}}_n = \sqrt{\sigma_n}d\mathbf{W}_n,$$

where each $d\mathbf{W}_n$ is a 2D uncorrelated Wiener process with unit variance. Substituting into Eq. (3.1) and taking inner products with $\cos(n\theta)$ and $\sin(n\theta)$ yields the SDE

$$d\mathbf{x}_n = \begin{cases} -w_n \nabla_{\mathbf{x}_n} V(\mathbf{x}_1, \dots, \mathbf{x}_M)dt + \sqrt{\sigma_n}d\mathbf{W}_n, & 1 \leq n \leq N \\ [-\mathbf{x}_n + \mathbf{I}_n]dt + \sqrt{\sigma_n}d\mathbf{W}_n, & N < n \leq M. \end{cases} \quad (3.2)$$

Thus, \mathbf{x}_n for $n \leq N$ follows a scaled stochastic gradient system while \mathbf{x}_n for $n > N$ describes an Ornstein-Uhlenbeck (OU) process.

A. Planar gradient system and stationary probability density

Let us consider the simplest case $N = M = 1$ and $\mathbf{I}_n = \mathbf{0}$ for $n \neq 1$, so that we have the planar stochastic gradient system

$$d\mathbf{x} = -\nabla V_0(\mathbf{x})dt + \sqrt{\sigma}d\mathbf{W}, \quad (3.3)$$

where we set $\sigma = \sigma_1$ throughout the rest of our analysis. The corresponding Fokker-Planck (FP) equation for the probability density $p(\mathbf{x}, t)$ takes the form

$$\frac{\partial p}{\partial t} = \nabla \cdot [\nabla V_0(\mathbf{x})p(\mathbf{x}, t)] + \frac{\sigma}{2} \nabla^2 p(\mathbf{x}, t), \quad (3.4)$$

with $\nabla^2 = \nabla \cdot \nabla$ and $V_0(\mathbf{x})$ given by Eq. (2.9). The advantage of a multivariate stochastic gradient system is that one can

explicitly write the stationary probability density

$$p(\mathbf{x}) = \frac{1}{Z} e^{-2V_0(\mathbf{x})/\sigma}, \quad (3.5)$$

which exists provided that the normalization factor

$$Z := \int_{\mathbb{R}^2} e^{-2V_0(\mathbf{x})/\sigma} d\mathbf{x} < \infty.$$

Observe that

$$|F(u)| = \left| \int_0^u f(u)du \right| \leq |u|$$

since $0 < f < 1$ is bounded, and therefore from Eq. (2.10)

$$\begin{aligned} |G(\mathbf{x})| &\leq \int_{-\pi}^{\pi} |F(\mathbf{x} \cdot \mathbf{e}(\theta))| d\theta \\ &\leq \|\mathbf{x}\| \int_{-\pi}^{\pi} |\cos[\theta - \arg(\mathbf{x})]| d\theta = 4\|\mathbf{x}\|. \end{aligned}$$

Hence,

$$V_0(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 - \mathbf{x} \cdot \mathbf{I} - G(\mathbf{x}) \sim \frac{1}{2}\|\mathbf{x}\|^2, \quad \text{as } \|\mathbf{x}\| \rightarrow \infty$$

and Z is indeed finite.

Since $V_0(\mathbf{x}) \rightarrow U_0(\mathbf{x})$ when $\mathbf{I} = \mathbf{0}$ and $U_0(\mathbf{x})$ is radially symmetric, Eq. (3.5) shows that the stochastic phase of the bump is uniformly distributed in the absence of an input, which is consistent with previous studies of wandering bumps [16]. The novelty here is that we can also determine the stationary density for the bump amplitude. When $\mathbf{I} \neq \mathbf{0}$, $V_0(\mathbf{x})$ is not radially symmetric and thus has a discrete set of local minima that lie along the line spanned by \mathbf{I} . It is useful to rewrite the stochastic dynamics in terms of the amplitude A and phase Δ of the bump by converting to polar coordinates $\mathbf{x} = A(\cos(\Delta), \sin(\Delta))$. To that end we set

$$p(\mathbf{x}, t)d\mathbf{x} = \hat{p}(A, \Delta, t)A dA d\Delta.$$

Transforming the time-dependent FP equation (3.4) to polar coordinates yields (after dropping the hat)

$$\begin{aligned} \frac{\partial p}{\partial t} &= \frac{1}{A} \frac{\partial}{\partial A} \left\{ A \left[\frac{\partial U_0(A)}{\partial A} - \bar{I} \cos(\Delta - \bar{\Delta}) \right] p \right\} \\ &\quad + \frac{1}{A} \frac{\partial}{\partial \Delta} [\bar{I} \sin(\Delta - \bar{\Delta}) p] \\ &\quad + \frac{\sigma}{2} \left(\frac{\partial^2}{\partial A^2} + \frac{1}{A} \frac{\partial}{\partial A} + \frac{1}{A^2} \frac{\partial^2}{\partial \Delta^2} \right) p. \end{aligned} \quad (3.6)$$

Multiplying both sides of (3.6) by A , we deduce that the corresponding SDE for the amplitude and phase takes the form

$$dA = \left[-\partial_A U_0(A) + \frac{\sigma}{2A} + \bar{I} \cos(\Delta - \bar{\Delta}) \right] dt + \sqrt{\sigma} dW_A, \quad (3.7a)$$

$$d\Delta = -\frac{\bar{I}}{A} \sin(\Delta - \bar{\Delta}) dt + \frac{\sqrt{\sigma}}{A} dW_{\Delta}, \quad (3.7b)$$

with $W_A(t)$ and $W_{\Delta}(t)$ independent Wiener processes with unit variance. Note that Eqs. (3.7) could also be derived by performing a change of variables in Eq. (3.3) and using Ito's lemma.

Using polar coordinates, we can derive a relationship between the stationary densities with and without an input (which we will call the homogeneous and inhomogeneous densities, respectively). The probability density in polar coordinates is given by

$$Ap(A, \Delta) = \frac{1}{Z} A e^{2\bar{I}A \cos(\Delta - \bar{\Delta})/\sigma} e^{-2U_0(A)/\sigma},$$

$$Z = \int_{-\pi}^{\pi} \int_0^{\infty} A e^{2\bar{I}A \cos(\Delta - \bar{\Delta})/\sigma} e^{-2U_0(A)/\sigma} dA d\Delta,$$

since $V_0(\mathbf{x}) = U_0(\mathbf{x}) - \mathbf{x} \cdot \mathbf{I}$, that is, $V_0(A, \Delta) = U_0(A) - A\bar{I} \cos(\Delta - \bar{\Delta})$ with

$$U_0(A) = \frac{1}{2}A^2 - w \int_{-\pi}^{\pi} F(A \cos \theta) d\theta. \quad (3.8)$$

When $\bar{I} = 0$ we have

$$Ap(A, \Delta) = \frac{1}{2\pi Z_h} A e^{-2U_0(A)/\sigma} \equiv \frac{1}{2\pi} p_A^h(A), \quad (3.9)$$

where

$$Z_h = \int_0^{\infty} A e^{-2U_0(A)/\sigma} dA,$$

and we define $p_A^h(A)$ as the homogeneous marginal density for the amplitude. Obviously, this shows that the variables A and Δ are independent, and the distribution for the phase is uniform over S^1 . An example of the amplitude density p_A^h is presented in Fig. 3(a). Note that, although the system is bistable in the chosen parameter regime, p_A^h has no peak at $A = 0$. This is due to the conversion to polar coordinates and the fact that the amplitude of the mean of \mathbf{x} is not the same as the mean of the amplitude of \mathbf{x} . The function $p(\mathbf{x}, t)$ has peaks at $\mathbf{x} = 0$ and $\|\mathbf{x}\| = a^*$ and thus $\hat{p}(A, \Delta, t)$ has peaks at $A = 0$ and $A = a^*$. However, changing variables also introduces a Jacobian factor of A , which contributes to the expression for p_A^h , and causes the peak at $A = 0$ to vanish.

Similarly, for $\bar{I} \neq 0$ we have the following relationship between the homogeneous marginal amplitude density and the full inhomogeneous density:

$$p_{A,\Delta}^{ih}(A, \Delta) \equiv Ap(A, \Delta) = \frac{Z_h}{Z} e^{2\bar{I}A \cos(\Delta - \bar{\Delta})/\sigma} p_A^h(A). \quad (3.10)$$

An example contour plot of p^{ih} in the (a, b) is shown in Fig. 3(b). It follows that the statistics in the inhomogeneous case can be calculated using the density from the homogeneous case. For example, the moments of A and circular moments of Δ with respect to the inhomogeneous density are equivalent to

$$\begin{aligned} \mathbb{E}[A^n e^{im\Delta}]_{ih} &= \frac{2\pi Z_h}{Z} \int_0^{\infty} A^n p_A^h(A) \\ &\times \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{im\Delta} e^{2\bar{I}A \cos(\Delta - \bar{\Delta})/\sigma} d\Delta \right] dA \\ &= \zeta \mathbb{E}[A^n I_m(\beta A)]_h e^{im\bar{\Delta}}, \end{aligned} \quad (3.11)$$

where we define $\beta = 2\bar{I}/\sigma$, $\zeta = 2\pi Z_h/Z$, and, after shifting $\Delta \rightarrow \Delta + \bar{\Delta}$ and factoring out $e^{i\bar{\Delta}}$, the integral in the square brackets is the definition of the m th order modified Bessel

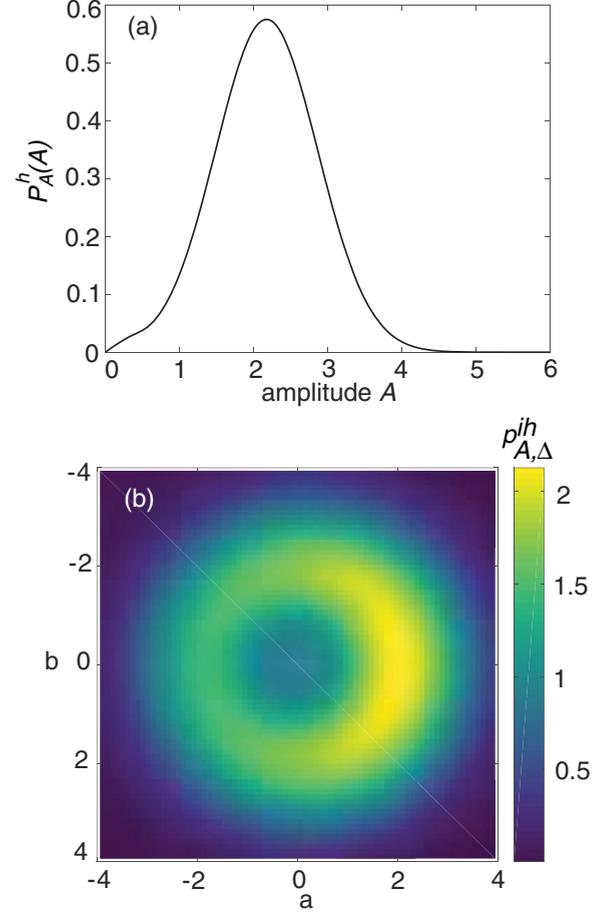


FIG. 3. Probability densities in the bistable regime for (a) the homogeneous ring model p_A^h , and (b) the inhomogeneous ring model $p_{A,\Delta}^{ih}$ (unnormalized), with $\mathbf{I} = (\mathbf{0.1}, \mathbf{0})$. Other parameter values are the same as Fig. 1.

function

$$I_m(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(x \cos \theta) \cos(m\theta) d\theta. \quad (3.12)$$

(The latter can also be related to circular moments of the so-called von Mises distribution, see Sec. III B.) In particular, the means are

$$\langle A \rangle = \zeta \mathbb{E}[AI_0(\beta A)]_h, \quad (3.13a)$$

$$\langle e^{i\Delta} \rangle = \zeta \mathbb{E}[I_1(\beta A)]_h e^{i\bar{\Delta}}, \quad (3.13b)$$

the variances are computed as

$$\text{Var}(A) = \zeta \mathbb{E}[A^2 I_0(\beta A)]_h - \zeta^2 \mathbb{E}[AI_0(\beta A)]_h^2, \quad (3.14a)$$

$$\text{Var}(e^{i\Delta}) = (\zeta \mathbb{E}[I_2(\beta A)]_h - \zeta^2 \mathbb{E}[I_1(\beta A)]_h^2) e^{i2\bar{\Delta}}, \quad (3.14b)$$

and the covariance is given by

$$\text{cov}(A, e^{i\Delta}) = (\zeta \mathbb{E}[AI_1(\beta A)] - \zeta^2 \mathbb{E}[I_0(\beta A)]_h \mathbb{E}[I_1(\beta A)]_h) e^{i\bar{\Delta}}. \quad (3.15)$$

Moreover, for $\bar{\Delta} = 0$,

$$\begin{aligned} \langle \cos(\Delta) \rangle &= \text{Re}[\langle e^{i\Delta} \rangle] \\ &= \zeta \mathbb{E}[I_1(\beta A)]_h, \end{aligned} \quad (3.16a)$$

$$\begin{aligned} \text{Var}[\cos(\Delta)] &= \frac{1}{2} [1 + \text{Re}\langle e^{2i\Delta} \rangle] - [\text{Re}\langle e^{i\Delta} \rangle]^2 \\ &= \frac{1}{2} [1 + \zeta \mathbb{E}[I_2(\beta A)]_h - 2\zeta^2 \mathbb{E}[I_1(\beta A)]_h^2], \end{aligned} \quad (3.16b)$$

$$\begin{aligned} \text{cov}(A, \cos(\Delta)) &= \text{Re}[\text{cov}(A, e^{i\Delta})] = \zeta \mathbb{E}[AI_1(\beta A)] \\ &\quad - \zeta^2 \mathbb{E}[I_0(\beta A)]_h \mathbb{E}[I_1(\beta A)]_h, \end{aligned} \quad (3.16c)$$

and the expressions are unchanged when using $\cos(\Delta - \bar{\Delta})$. Note that, in the absence of an input, setting $\beta = 0$ shows that there is no correlation between amplitude and phase and the presence of the input introduces correlations.

Finally, if we only want to analyze the amplitude and phase separately, we can compute the marginal densities as

$$p_A^h(A) = \zeta I_0(\beta A) p_A^h(A), \quad (3.17a)$$

$$p_\Delta^h(\Delta) = \frac{\zeta}{2\pi} \mathbb{E}[e^{2iA \cos(\Delta - \bar{\Delta})/\sigma}]_h. \quad (3.17b)$$

Thus, the inhomogeneous amplitude density is proportional to the homogeneous density modulated by the Bessel function and the inhomogeneous phase density is proportional to the expected value of a von Mises distribution with respect to the homogeneous amplitude density. It is interesting that the distributions for the inhomogeneous system can be completely determined by the probability distributions of the amplitude from the homogeneous system.

In Fig. 4 we show an example of the steady-state variances and covariance as a function of input strength \bar{I} for $\sigma = 1$ and two different values for the firing rate threshold: $\kappa = 0.5$ and 0.9 . (Since we wish to investigate the effects of amplitude fluctuations outside the weak noise regime, we consider relatively strong noise throughout.) In both cases, the variance of $\cos(\Delta)$ monotonically decreases toward zero as the input increases, while the covariance has a weak nonmonotonic dependence but still approaches zero for large input. On the other hand, the nature of the dependence of the amplitude variance on input strength depends on parameter values. For $\kappa = 0.5$ we see that the variance remains approximately unchanged, while for $\kappa = 0.9$ the presence of a sufficiently large input decreases amplitude variability. The difference in the input dependence of phase and amplitude fluctuations can be understood in terms of the linear stability analysis of Sec. II B. Recall that the discrete spectrum consists of two eigenvalues. The first is $\lambda_e = -1 + \mathcal{F}'(a^*)$, which corresponds to stability with respect to perturbations in amplitude, and the second is $\lambda_o = -\bar{I}/a^*$, corresponding to phase changing perturbations. Note that a^* depends implicitly on \bar{I} according to Eq. (2.19). In particular, $a^* = \mathcal{F}(a^*) + \bar{I}$ and thus the stable amplitude monotonically increases with \bar{I} [see Fig. 2(a)]. Moreover, $\mathcal{F}(a) \rightarrow 2$ for as $a \rightarrow \infty$ and thus asymptotically $a^* \sim \bar{I} + 2$ as $\bar{I} \rightarrow \infty$. Therefore, $\lambda_o = 0$ at $\bar{I} = 0$ and monotonically decreases toward $\lambda_o \rightarrow -1$ as $\bar{I} \rightarrow \infty$. Now, focus on λ_e . Since $\mathcal{F}'(a) > 0$ for all a and monotonically approaches 0 as $a \rightarrow \infty$, then $\lambda_e > -1$ for all \bar{I} and monotonically decreases toward -1 as $\bar{I} \rightarrow \infty$, which is the same asymptotic behavior

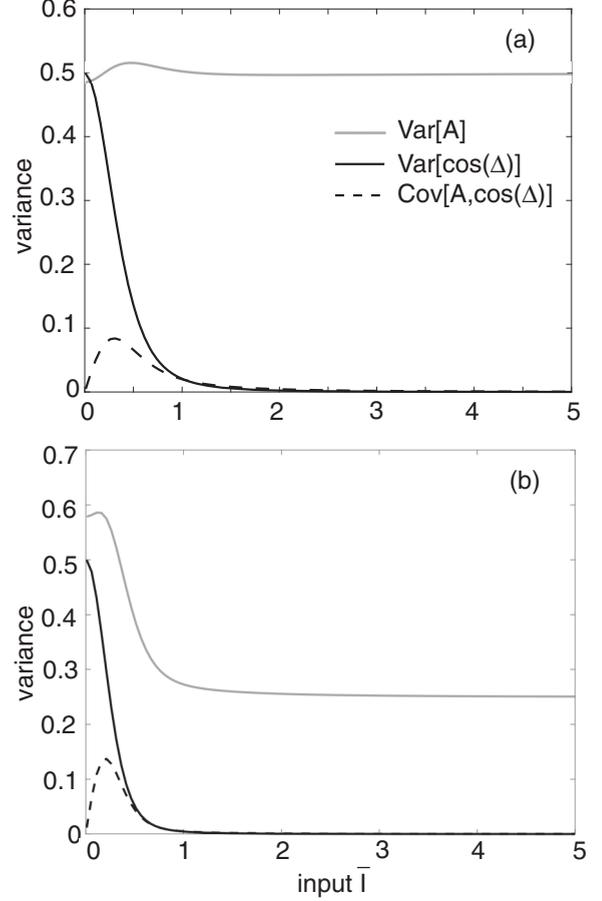


FIG. 4. Variances and covariance as a function of input strength \bar{I} for $\sigma = 1$ and (a) $\kappa = 0.5$, (b) $\kappa = 0.9$. All other parameter values are the same as in Fig. 2.

as λ_o . The main difference is the value at $\bar{I} = 0$, as illustrated in Fig. 5, which accounts for the fact that input suppression is stronger in the case of phase fluctuations. There are at least two factors contributing to the differences in noise

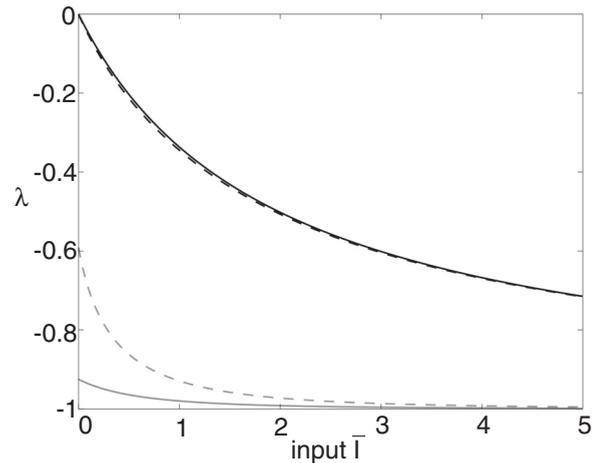


FIG. 5. Eigenvalues λ_e (gray lines) and λ_o (black lines) as a function of input \bar{I} for $\kappa = 0.5$ (solid lines) and $\kappa = 0.9$ (dashed lines).

suppression of the amplitude for $\kappa = 0.5$ and 0.9 . First, if $\kappa = 0.5$, then $\mathcal{F}'(a^*)$ is relatively small and there is a small change in λ_e as the input increases. On the other hand, if $\kappa = 0.9$, then $\mathcal{F}'(a^*)$ is larger and there is a more prominent change in λ_e (see Fig. 5). Second, when $\kappa = 0.5$, the homogeneous bistable system spends almost all of its time in the stable bump state so that noise-induced transitions to the stable zero state are rare events. On the other hand, when $\kappa = 0.9$, noise-induced transitions are more common, and can thus be suppressed by an external input (see Sec. IV). Such a mechanism for input suppression has previously been suggested by Litwin-Kumar and Doiron [30].

As highlighted elsewhere [25], the input-dependent reduction of phase fluctuations in ring attractor networks provides one possible mechanism for the experimentally observed suppression of cortical variability following the onset of a stimulus [26–29,31–34]. Our previous work showed that the suppression of phase fluctuations is able to account for a number of experimental results, provided that the noise is relatively weak ($\sigma \sim 0.1$ rather than $\sigma \sim 1$), so that amplitude fluctuations can be ignored to leading order [25]. This is a common assumption in perturbation theoretic approaches to the analysis of wandering bumps, which we turn to next. Our current analysis suggests that this particular noise-suppression mechanism could break down in the strong noise regime.

B. Weak noise limit and perturbation theory

Previous works on stochastic neural fields have utilized perturbation theory to derive an effective SDE for the stochastic motion of the phase of a bump (or traveling front) in the presence of weak noise [16,17,20,25]. Given a stable stationary solution $u = u_0(\theta)$ to the deterministic neural field, it is assumed that one can write the stochastic solution as

$$u(\theta, t) = u_0[\theta - \Delta(t)] + \epsilon^{1/2}v(\theta, t),$$

where $\epsilon \ll 1$ is the strength of the noise, $\Delta(t)$ is a stochastic variable tracking the phase displacement, and $v(\theta, t)$ is a correction term. This decomposition is motivated by the fact that the system is SO(2) equivariant when $\epsilon = 0$. Suppose that the weight kernel is $K(\theta) = w \cos(\theta)$, so that the stationary solution is $u_0 = a^* \cos(\theta)$, where a^* is a solution of Eq. (2.12). In the presence of a weak input of the form

$$I(\theta) = \sqrt{\epsilon} \bar{I} \cos(\theta - \bar{\Delta}),$$

perturbation theory leads, to leading order,

$$d\Delta = -\frac{\sqrt{\epsilon} \bar{I}}{a^*} \sin(\Delta - \bar{\Delta}) dt + \sqrt{\epsilon D} dW, \quad (3.18)$$

where

$$D = \frac{\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) \partial_{\theta'} f(u_0(\theta')) C(\theta - \theta') d\theta d\theta'}{\left[\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) u_0'(\theta) d\theta \right]^2} \quad (3.19)$$

is an effective diffusion coefficient. The phase equation (3.18) is identical in form to Eq. (3.7 b), except that the former effectively approximates the amplitude as a constant $A(t) = a^*$, and we have the rescalings $\bar{I} \rightarrow \sqrt{\epsilon} \bar{I}$, $\sqrt{\sigma} \rightarrow \sqrt{\epsilon D} a^*$.

It is important to note that the system (3.7) is exact, but requires the noise correlations to be of the specific form

$C(\theta - \theta') = \sigma \cos(\theta - \theta')$. On the other hand, the approximate phase equation (3.18) holds in the weak noise limit for an arbitrary noise correlation function. Expanding $C(\theta)$ as a Fourier series, we find that

$$D = \frac{\sum_{n=0}^{\infty} \sigma_n \left[\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) \sin(n\theta) d\theta \right]^2}{\left[\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) u_0'(\theta) d\theta \right]^2}$$

since $u_0(\theta)$ can be taken to be even, without loss of generality, and thus $\partial_{\theta} f(u_0(\theta))$ is orthogonal to $\cos(n\theta)$ for all n . In the specific case $C(\theta - \theta') = \sigma \cos(\theta - \theta')$, we have

$$\begin{aligned} & \int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) \partial_{\theta'} f(u_0(\theta')) C(\theta - \theta') d\theta d\theta' \\ &= \sigma \left[\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) \sin(\theta) d\theta \right]^2. \end{aligned}$$

The denominator in the expression for D can be evaluated as

$$\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) u_0'(\theta) d\theta = -a^* \int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) \sin(\theta) d\theta,$$

and thus $D = \sigma / a^{*2}$. Hence, we recover the noise amplitude of Eq. (3.7 b) for $A(t) = a^*$ and $\epsilon = 1$.

At the other extreme, when $C(\theta) = \sigma \delta(\theta)$, the diffusion coefficient is

$$\begin{aligned} D &= \sigma \frac{\int_{-\pi}^{\pi} [\partial_{\theta} f(u_0(\theta))]^2 d\theta}{\left[\int_{-\pi}^{\pi} \partial_{\theta} f(u_0(\theta)) u_0'(\theta) d\theta \right]^2} \\ &= \frac{\sigma}{a^{*2}} \frac{\int_{-\pi}^{\pi} [f'(u_0(\theta))]^2 \sin^2(\theta) d\theta}{\left[\int_{-\pi}^{\pi} f'(u_0(\theta)) \sin^2(\theta) d\theta \right]^2} \\ &= \frac{\sigma w^2}{a^{*2}} \int_{-\pi}^{\pi} [f'(u_0(\theta))]^2 \sin^2(\theta) d\theta, \end{aligned}$$

where the last equality follows from Eq. (2.15). Using the integral mean value theorem, there exists a ξ such that

$$D = \frac{\sigma w^2}{A^2} f'(u_0(\xi)) \int_{-\pi}^{\pi} \sin^2(\theta) d\theta = \frac{\sigma w}{a^{*2}} f'(u_0(\xi))$$

since $f'(u) > 0$ for all u . The noise amplitude is thus modified by the multiplicative factor $\sqrt{w f'(u_0(\xi))}$.

As highlighted in Ref. [25], Eq. (3.18) is known as a von Mises process, which can be regarded as a circular analog of the OU process on a line, and generates distributions that frequently arise in circular or directional statistics [35]. In particular, the phase has a stationary density given by the von Mises distribution

$$M(\Delta; \bar{\Delta}, \beta) = \frac{1}{2\pi I_0(\beta)} e^{\beta \cos(\Delta - \bar{\Delta})}, \quad \beta = \frac{2\bar{I}}{\sqrt{\epsilon} a^* D}. \quad (3.20)$$

Sample plots of the von Mises distribution are shown in Fig. 6 for $\bar{\Delta} = 0$. One finds that $M(\Delta; 0, \beta) \rightarrow 1/2\pi$ as $\beta \rightarrow 0$; since $\beta \sim \bar{I}$ this recovers the uniform distribution of pure Brownian motion on the circle. On the other hand, the von Mises distribution becomes sharply peaked as $\beta \rightarrow \infty$. More specifically, for large positive β ,

$$M(\Delta; 0, \beta) \approx \sqrt{\frac{\beta}{2\pi}} e^{-\beta \Delta^2/2}. \quad (3.21)$$

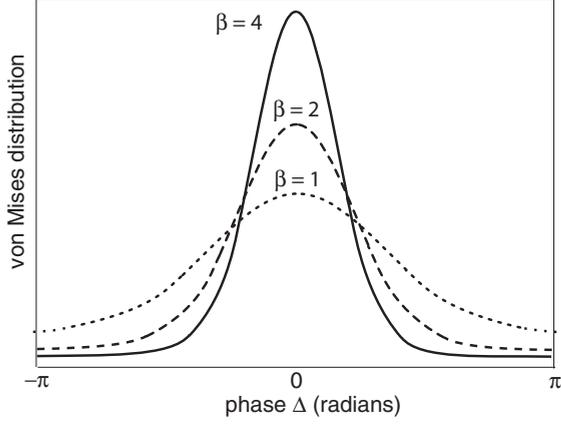


FIG. 6. Sample plots of the von Mises distribution $M(\Delta; 0, \beta)$ for various values of β .

Moments of the von Mises distribution are usually calculated in terms of the circular moments of the complex exponential $z = e^{i\beta}$ [see also Eq. (3.11)]. The n th circular moment is given by

$$\mu_n = \langle z^n \rangle_{\beta, \bar{\Delta}} := \int_{-\pi}^{\pi} z^n M(\Delta; \bar{\Delta}, \beta) d\Delta = \frac{I_n(\beta)}{I_0(\beta)} e^{in\bar{\Delta}}, \quad (3.22)$$

where again I_n is the modified Bessel function defined in Eq. (3.12). In particular, for $\bar{\Delta} = 0$, the first and second circular moments yield

$$\langle \cos \Delta \rangle_{\beta} = \frac{I_1(\beta)}{I_0(\beta)}, \quad \langle \sin \Delta \rangle_{\beta} = 0, \quad (3.23)$$

$$\langle \cos^2 \Delta \rangle_{\beta} = \frac{1}{2} [1 + \langle \cos(2\Delta) \rangle_{\beta}] = \frac{1}{2} \left[1 + \frac{I_2(\beta)}{I_0(\beta)} \right],$$

and

$$\langle \sin^2 \Delta \rangle_{\beta} = 1 - \langle \cos^2 \Delta \rangle_{\beta} = \frac{1}{2} \left[1 - \frac{I_2(\beta)}{I_0(\beta)} \right].$$

It follows that

$$\text{Var}[\cos(\Delta)] = \frac{1}{2} \left[1 + \frac{I_2(\beta)}{I_0(\beta)} - 2 \left(\frac{I_1(\beta)}{I_0(\beta)} \right)^2 \right], \quad (3.24a)$$

$$\text{Var}[\sin(\Delta)] = \frac{1}{2} \left[1 - \frac{I_2(\beta)}{I_0(\beta)} \right]. \quad (3.24b)$$

The best way to compare our gradient-based results to Eq. (3.20) is to compute the probability density for Δ conditioned on a fixed amplitude A . Using Eq. (3.10) for the density in polar coordinates and the marginal amplitude density in Eq. (3.17a) we obtain the conditional density

$$p_{\Delta|A}(\Delta|A) = \frac{P_{A,\Delta}^{jh}(A, \Delta)}{p_A(A)} = \frac{1}{2\pi I_0\left(\frac{2IA}{\sigma}\right)} e^{\frac{2IA}{\sigma} \cos(\Delta - \bar{\Delta})}, \quad (3.25)$$

which agrees exactly with Eq. (3.20) when $C(\theta) = \sigma \cos(\theta)$, $\epsilon = 1$, and $A = a^*$ since $D = \sigma/a^{*2}$.

C. Numerical results

We now simulate the time-dependent dynamics of two different systems: (i) the gradient system with first-order noise ($\sigma_n = \sigma \delta_{n,1}$) whose corresponding amplitude-phase SDE is given by Eq. (3.7) and (ii) the neural field Eq. (3.1) with higher-order noise. We compare these with the steady-state results from both the gradient system and approximate phase SDE in Eq. (3.18). For the sake of illustration, we take the correlation function of the noise in the neural field equation to be given by the truncated convergent series $C(\theta) = \sum_{n=1}^{20} \frac{1}{n^2} \cos(n\theta)$. We use the Euler-Maruyama method for time integration and the trapezoidal rule for integration in θ :

$$\begin{aligned} u(\theta_i, t_{j+1}) = & u(\theta_i, t_j) + dt [-u(\theta_i, t_j) + \bar{I} \cos(\theta_i) \\ & + \cos(\theta_i) \mathcal{I}_c(u(\theta, t_j)) + \sin(\theta_i) \mathcal{I}_s(u(\theta, t_j))] \\ & + \sqrt{dt} \sum_{n=1}^M \sqrt{\sigma_n} [\eta_n^c \cos(n\theta_i) + \eta_n^s \sin(n\theta_i)], \end{aligned}$$

where \mathcal{I}_c and \mathcal{I}_s are the trapezoidal rules for the integrals

$$\int_{-\pi}^{\pi} f(u(\theta, t)) \cos(\theta) d\theta, \quad \int_{-\pi}^{\pi} f(u(\theta, t)) \sin(\theta) d\theta,$$

respectively, and η_n^c, η_n^s are independent normally distributed random variables with unit variance. To obtain a more direct comparison between the gradient system and the neural field with higher-order noise, we use the same realization of noise for the first-order contribution. That is, we generate η_1^c, η_1^s from a normal distribution and use these same values in both equations. Finally, for each iteration we set the initial condition to

$$A(0) = a^*, \quad \Delta(0) = \theta_0, \quad u(\theta, 0) = a^* \cos(\theta - \theta_0),$$

where a^* is the stable bump amplitude and θ_0 is pulled from the uniform distribution on $[-\pi, \pi]$. We chose a random angle because it helped significantly to increase the convergence speed to steady state for the phase distribution, while the amplitude distribution converged relatively quickly regardless. To obtain the amplitude for the neural field solution with higher-order noise we compute $A = \max_{\theta \in [0, 2\pi]} u(\theta)$ and then we define the phase to be the value of θ that yields this max.

In Fig. 7 we plot the time-dependent mean of $\cos(\Delta)$ and A , for both the homogeneous and inhomogeneous systems. The solid black and gray lines denote results from simulations for the first-order noise and higher-order noise, respectively. The red solid lines indicate the theoretical steady-state means obtained from Eqs. (3.13a) and (3.16a), and the red dashed line in Fig. 7(b) corresponds to the approximate theoretical mean from Eq. (3.23). We see that there is a very close match between the first-order noise case and theory, which is to be expected. For the higher-order noise system, the qualitative dynamics are very similar, with an increase in the mean amplitude and a decrease in the mean of $\cos(\Delta)$. In Fig. 8 we show corresponding plots of the variance of the phase and amplitude. (The covariance is plotted in Fig. 9.) The theoretical variances are obtained from Eqs. (3.16b) and (3.14a), and the dashed red line in Fig. 8(b) denotes the theoretical variance

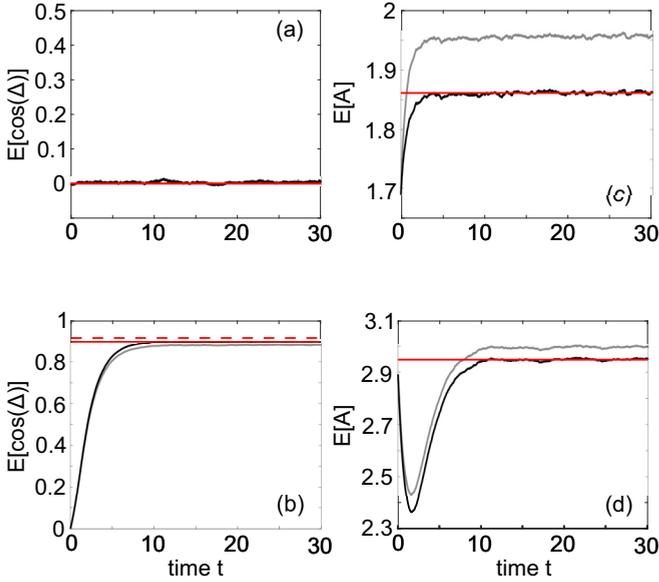


FIG. 7. Plots of the time-dependent mean of $\cos(\Delta(t))$ and the amplitude $A(t)$. The black solid lines correspond to simulations of the gradient system with first-order noise, evolving according to the amplitude-phase SDE (3.7), and the gray solid lines correspond to simulations of the full neural field equation (3.1) with higher-order noise. The solid straight (red) lines are the exact analytical results obtained from Eqs. (3.16a) and (3.13a). (a), (b) $\mathbb{E}[\cos(\Delta(t))]$ for a homogeneous and an inhomogeneous network, respectively. The dashed straight (red) line corresponds to the approximation from the perturbative approach in Eq. (3.23). (c), (d) $\mathbb{E}[A(t)]$ for a homogeneous and an inhomogeneous network, respectively. The input, noise strengths, and firing rate threshold are $\bar{I} = 0.5$, $\sigma = 1$, and $\kappa = 0.9$. We take the discretizations $\Delta t = 0.01$ and $\Delta \theta = 0.01$ and run 50 000 realizations. All other parameters are the same as in Fig. 2.

from the approximate phase SDE in Eq. (3.24a). Note that in the homogeneous system, the steady-state phase variance is the same for the gradient system and the phase SDE, and thus there is only one red line in Fig. 8(a). Figure 8(c) shows a very close match between the exact theory and simulations for the first-order noise. The main effect of higher-order noise is to change the overall variance. For the homogeneous system, variance is decreased when including higher-order modes whereas for the inhomogeneous system, the variances are approximately the same. We have no intuition for this, but the fact that the variance is decreased does not seem to be a general result. For other parameter values, the variances in the homogeneous system are approximately the same (results not shown). The error in the approximate variance obtained from the phase SDE is also quite small, particularly, given the size of the noise strength, and becomes very accurate for $\sigma \ll 1$. The results for the amplitude variance are similar, although there seems to be a negligible effect from higher-order noise.

Finally, in Figs. 10 and 11 we compare the marginal densities $p_\Delta(\Delta)$ and $p_A(A)$ with numerical simulations of the 2D gradient system as well as the full neural field equation with higher-order noise. The main difference with the higher-order noise is that the amplitude distribution is shifted slightly to

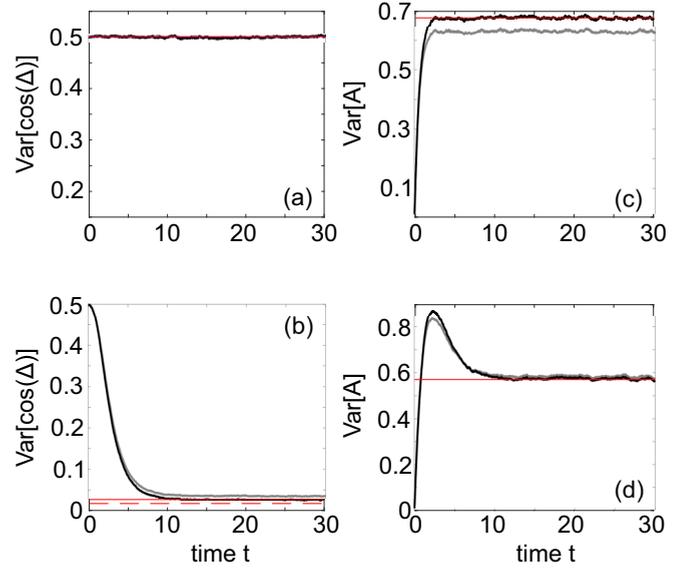


FIG. 8. Plots of the time-dependent variance of $\cos(\Delta(t))$ and the amplitude $A(t)$. The black solid lines correspond to simulations of the gradient system with first-order noise, evolving according to the amplitude-phase SDE (3.7), and the gray solid lines correspond to simulations of the full neural field equation (3.1) with higher-order noise. The solid straight (red) lines are the exact analytical results obtained from Eqs. (3.16b) and (3.14a). (a), (b) $\text{Var}[\cos(\Delta(t))]$ for a homogeneous and an inhomogeneous network, respectively. The dashed straight (red) line corresponds to the approximation from the perturbative approach in Eq. (3.23). (c), (d) $\text{Var}[A(t)]$ for a homogeneous and an inhomogeneous network, respectively. All other parameters are the same as Fig. 7.

the right. This is consistent with the fact that the higher-order noise increased the mean amplitude as seen in Fig. 8. It can be seen that there is hardly a difference between the two

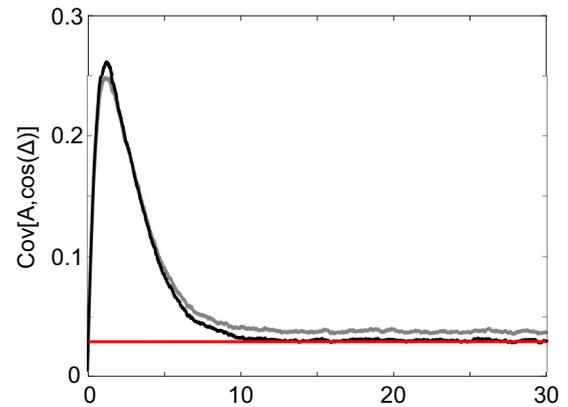


FIG. 9. Plots of the time-dependent covariance between A and $\cos(\Delta)$ for the inhomogeneous network. Note that, as we show above, the steady-state covariance is zero for the homogeneous network so we exclude this case here. The black solid line corresponds to simulations of the gradient system with first-order noise, and the gray solid line corresponds to simulations of the full neural field equation with higher-order noise. The straight (red) line corresponds to the exact steady-state variance from Eq. (3.16c). Parameter values are as in Fig. 7.

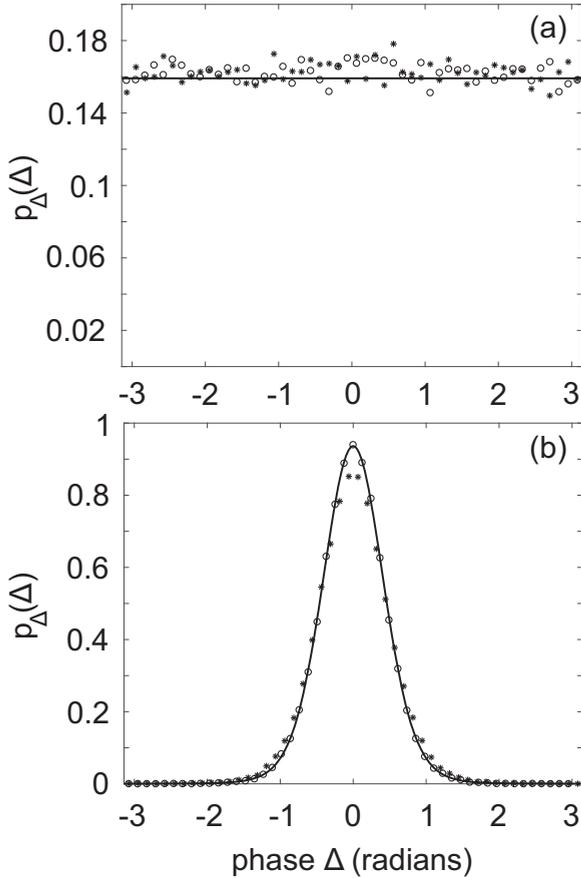


FIG. 10. Simulation of steady-state distribution for Δ in the (a) homogeneous system and (b) inhomogeneous system. Solid line indicates the theoretical distribution from Eq. (3.17b). Circles correspond to results from the full 2D stochastic gradient system. Stars correspond to results from the full neural field model with higher-order noise included. Parameters are the same as in Fig. 7.

types of noise and both show good agreement with Eq. (3.17), with the minor exception of the homogeneous phase, where the convergence to steady state is very slow. We also find that $p_{\Delta}(\Delta)$ is very well approximated by the von Mises distribution [which is equivalent to the conditional density $p(\Delta|a^*)$] for the parameter values used in the simulations (not shown).

IV. ESCAPE TIME AND BUMP EXTINCTION

Recall that one application of a bistable ring attractor network is to working memory. In the deterministic case, a transient input can switch the network from a uniform rest state to a stable bump state that persists when the input is removed. As we have already discussed, noise can induce the location of the peak to wander from its original position due to phase diffusion. However, it is also possible for the bump to disappear completely due to large fluctuations in the amplitude of the bump in the absence of an input. The “extinction” of the bump indicates a loss of memory of the stimulus. This then raises the important issue of how long, on average, it takes the bump to go extinct. We will assume that the network operates in a bistable regime with the stable rest

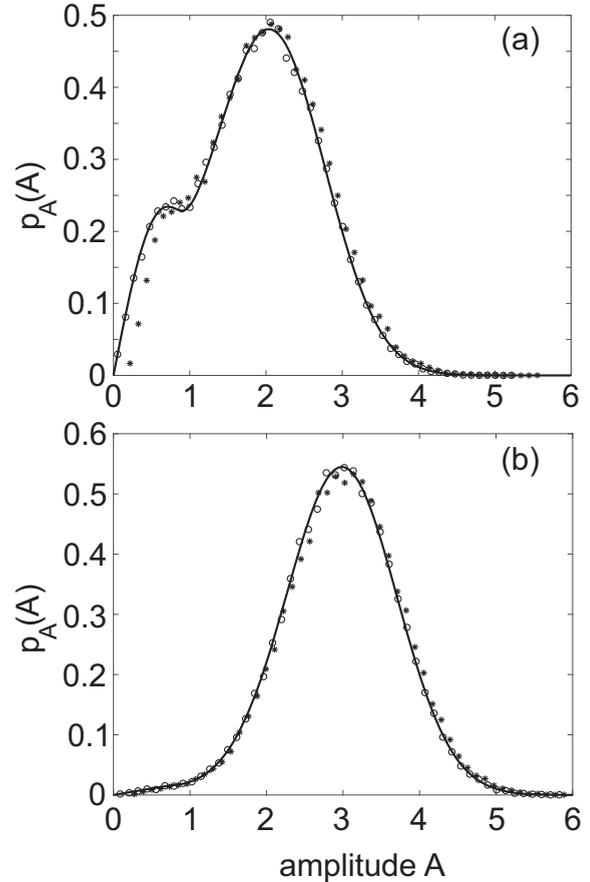


FIG. 11. Simulation of steady-state distribution for A in the (a) homogeneous system and (b) inhomogeneous system. Solid line indicates the theoretical distribution from Eq. (3.17). Circles correspond to results from the 2D stochastic gradient system. Stars correspond to results from the full neural field model with higher-order noise included. Parameters are the same as in Fig. 7.

state and stable bump state with amplitude a^* separated by an unstable bump state with amplitude a_0 , $0 < a_0 < a^*$. In other words, the radially symmetric potential $U_0(A)$ of Eq. (2.18) has minima at $A = 0$ and a^* and a maximum at $A = a_0$. (The analysis in the case of a nonzero input is considerably more involved since one has to deal with the full planar dynamical system. Now, there is a noise-induced transition from a large amplitude stable bump to a low amplitude stable bump rather than a zero state.)

Let $T(r)$, $r > a_0$, denote the mean first passage time (MFPT) that the system starting at r reaches a_0 . From radial symmetry it is well known that $T(r)$ satisfies the boundary value problem [36]

$$-U'_0(r)T'(r) + \frac{\sigma}{2r} \frac{d}{dr}(rT'(r)) = -1, \quad (4.1)$$

together with the absorbing boundary condition $T(a_0) = 0$. Define the modified potential

$$U(r) = U_0(r) - \frac{\sigma}{2} \ln(r),$$

and rewrite Eq. (4.1) as

$$-\frac{2}{\sigma}U'(r)T'(r) + T'' = -\frac{2}{\sigma}.$$

Note that $U(r) \sim r^2$ as $r \rightarrow \infty$. Multiplying both sides by $e^{-2U(r)/\sigma}$ yields

$$(T'e^{-2U(r)/\sigma})' = -\frac{2}{\sigma}e^{-2U(r)/\sigma}.$$

Integrate from r to ∞ and note that $T'(r)e^{-2U(r)/\sigma} \rightarrow 0$ as $r \rightarrow \infty$ so that

$$T' = \frac{2}{\sigma}e^{2U(r)/\sigma} \int_r^\infty e^{-2U(r')/\sigma} dr'.$$

Integrating again from r to a_0 and using the boundary condition $T(a_0) = 0$, we obtain

$$\begin{aligned} T(r) &= \frac{2}{\sigma} \int_{a_0}^r e^{2U(r')/\sigma} \left[\int_{r'}^\infty e^{-2U(r'')/\sigma} dr'' \right] dr' \\ &= \frac{2}{\sigma} \int_{a_0}^r \frac{1}{r'} e^{2U_0(r)/\sigma} \left[\int_{r'}^\infty r'' e^{-2U_0(r'')/\sigma} dr'' \right] dr'. \end{aligned} \quad (4.2)$$

It follows that the mean time to reach a_0 , starting at the stable bump solution is $T(a^*)$, and the corresponding mean extinction time is thus $2T(a^*)$. The mean time to bump extinction has been numerically studied in [16] and analytically approximated in [18] when the system is near a saddle-node bifurcation. However, to our knowledge, there has never been an exact formula.

In general, one has to evaluate the integrals in Eq. (4.2) numerically. However, if σ is sufficiently small, then one can approximate the integrals using steepest descents. Using standard arguments, one can approximate $T(a^*)$ by the product of two independent integrals [36]

$$T(a^*) = \frac{2}{\sigma} \left[\int_{a_0}^{a^*} \frac{1}{r'} e^{2U_0(r)/\sigma} dr' \right] \left[\int_{a_0}^\infty r'' e^{-2U_0(r'')/\sigma} dr'' \right].$$

The first integral is dominated by the peak of U_0 at $r = a_0$ and the second integral is dominated by the minimum of U_0 around $r = a^*$. Therefore, the potential is Taylor expanded as

$$U_0(r) = U_0(a_0) - \frac{1}{2} \left(\frac{r - a_0}{\delta} \right)^2 + \dots$$

near $r = a_0$ and

$$U_0(r) = U_0(a^*) + \frac{1}{2} \left(\frac{r - a^*}{\alpha} \right)^2 + \dots$$

near $r = a^*$, where

$$\delta = \frac{1}{\sqrt{|U_0''(a_0)|}}, \quad \alpha = \frac{1}{\sqrt{|U_0''(a^*)|}}.$$

The second integral is then approximated as

$$\begin{aligned} \int_{a_0}^\infty r'' e^{-2U_0(r'')/\sigma} dr'' &\approx a^* \int_{-\infty}^\infty e^{-2U_0(a^*)/\sigma} e^{-(r''-a^*)^2/\sigma\alpha^2} dr'' \\ &\approx \sqrt{\pi\sigma}\alpha^* e^{-2U_0(a^*)/\sigma}, \end{aligned}$$

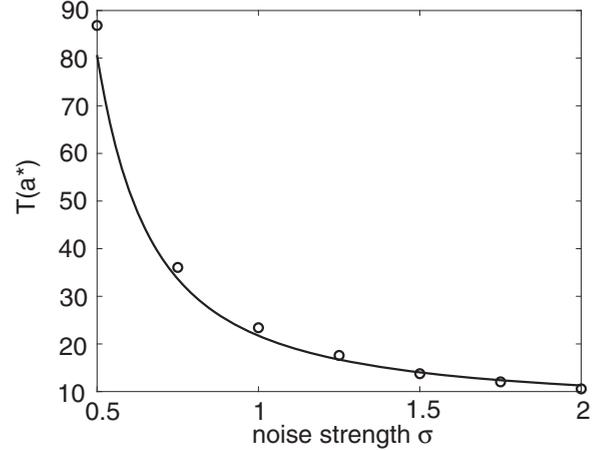


FIG. 12. Numerical simulation of mean escape time (circles) compared against the approximation from Eq. (4.3) (solid line) for the gradient system in Eq. (3.3) with $\bar{I} = 0$. Other parameters are the same as Fig. 7.

and the first integral as

$$\begin{aligned} \int_{a_0}^{a^*} \frac{1}{r'} e^{2U_0(r')/\sigma} dr' &\approx \frac{1}{a_0} \int_{a_0}^\infty e^{2U_0(a_0)/\sigma} e^{-(r'-a_0)^2/\sigma\delta^2} \\ &\approx \sqrt{\pi\sigma}\delta \frac{1}{2a_0} e^{2U_0(a_0)/\sigma}. \end{aligned}$$

Hence,

$$T(a^*) \approx \pi \frac{a^*}{a_0} \frac{e^{2[U_0(a_0) - U_0(a^*)]/\sigma}}{\sqrt{|U_0''(a_0)||U_0''(a^*)|}}. \quad (4.3)$$

[If we had taken the absorbing boundary to be at $a = 0$, then we would have found the mean extinction time to be $2T(a^*)$.] A plot of the mean escape time versus noise strength σ is shown in Fig. 12, and example plots of noise-induced transitions are shown in Fig. 13(a). It can be seen from Fig. 12 that the approximation in Eq. (4.3) is a good match with numerical simulations even for relatively large values of σ .

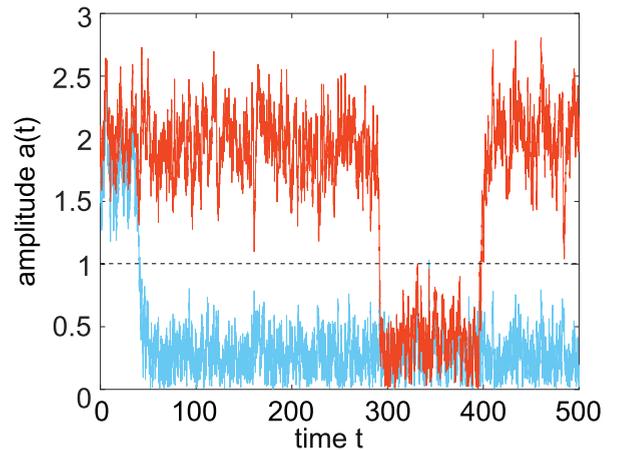


FIG. 13. Numerical simulations of a single noise-induced transition from a stable bump to the zero state for the full stochastic neural field model with $\bar{I} = 0$ (blue or lighter plot) and $\bar{I} = 0.2$ (red or darker plot). Other parameters are the same as Fig. 7.

As $\sigma \rightarrow 0$, the MFPT becomes exponentially large, so that the extinction of a bump is unlikely to be observed over physiological timescales. In Fig. 13, we compare single realizations of the amplitude of the neural field solution. We find that the input reduces the amount of time spent in the zero state, which is one of the factors in the input-dependent suppression of amplitude fluctuations discussed at the end of Sec. III A.

V. SPHERICAL MODEL

In this section we show how to extend our gradient analysis of ring attractor networks to the case of spherical attractor networks. First, note that analogous to Fourier series expansions, one can perform an expansion in spherical harmonics to show that a spherical attractor network also supports stationary bump solutions, and that these can also lock to weakly biased external stimuli [12,13]. In the absence of inputs, the bumps are marginally stable with respect to rotations of the sphere, reflecting equivariance of the neural field equation with respect to the action of the special orthogonal group $SO(3)$. One major application of spherical attractor networks is to model the joint orientation and spatial frequency tuning of neurons in a V1 hypercolumn [12,13]. Such a model was originally motivated by a number of optical imaging studies, which suggested that both orientation and spatial frequency preferences are distributed almost continuously across cortex, with iso-orientation and isofrequency contours being approximately orthogonal, so that they generate a local curvilinear coordinate system [37,38]. Although the existence of spatial frequency preference maps in V1 is still controversial, a more recent two-photon imaging study appears to be consistent with earlier studies [39]. Further evidence for spherical network structures in cortex has been provided by multielectrode data analysis of cortical activity patterns based on computational homology [40]. On larger spatial scales, the Nunez model for the generation of electroencephalogram (EEG) signals [41] has recently been formulated as a neural field model on a sphere with space-dependent delays [42].

Marginal stability of spherical bumps with respect to rotations of the sphere suggests that weak noise will induce a stochastic wandering of the bumps, characterized by two-dimensional Brownian motion of the peak of the bump on the surface of the sphere (in the absence of inputs). This has recently been established using a combination of group theory, harmonic analysis, and perturbation theory [43], which also highlighted one key difference between spherical attractor networks and ring attractor networks: the sphere S^2 is a curved manifold whereas the circle S^1 is intrinsically flat. Here, we show that under extra constraints on noise correlations, one can formulate a stochastic neural field equation on the sphere as a three-dimensional gradient system.

A. Deterministic model

In the absence of noise, a spherical attractor network can be modeled in terms of the neural field equation

$$\frac{\partial u}{\partial t} = -u + \int_{S^2} K(\theta, \phi | \theta', \phi') f(u(\theta', \phi')) \times d\mu(\theta', \phi') + I(\theta, \phi), \quad (5.1)$$

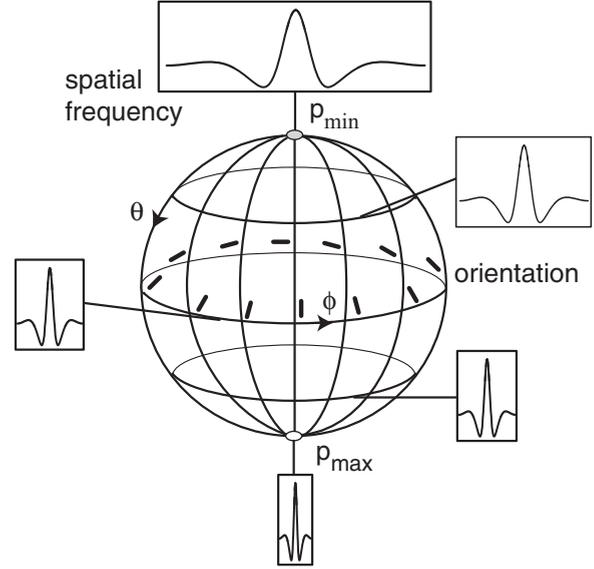


FIG. 14. Spherical network topology. Cells are labeled by the pair of angular coordinates (θ, ϕ) on the surface of a unit sphere, with $0 \leq \theta < \pi$ and $0 \leq \phi \leq 2\pi$. These coordinates could represent a pair of stimulus feature preferences such as spatial frequency and orientation.

where μ is the measure on the sphere and thus $d\mu = \sin(\theta')d\theta'd\phi'$. Within the context of orientation and spatial frequency tuning, $u(\theta, \phi)$ denotes the activity of neural population with orientation preference $\phi/2$ and spatial frequency preference p , which is related to θ according to $\theta = \pi \frac{\log(p/p_{\min})}{\log(p_{\max}/p_{\min})}$ [12,13] (see Fig. 14). The function K denotes the synaptic strength between neural populations (θ, ϕ) and (θ', ϕ') . We will take Eq. (5.1) to be equivariant with respect to the action of $SO(3)$, that is, rotations of the sphere. This requires K to be invariant under this action:

$$K(\gamma(\theta, \phi) | \gamma(\theta', \phi')) = K(\theta, \phi | \theta', \phi'), \quad \forall \gamma \in SO(3).$$

Hence, K will have an expansion in terms of spherical harmonics,

$$K(\theta, \phi | \theta', \phi') = \sum_{n=0}^{\infty} w_n \sum_{m=-n}^n Y_n^m(\theta, \phi) Y_n^{m*}(\theta', \phi'), \quad (5.2)$$

where

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m[\cos(\theta)] e^{im\phi}$$

for $n \geq 0$ and $-n \leq m \leq n$, and $P_n^m(\cos \theta)$ is an associated Legendre function. The functions $Y_n^m(\theta, \phi)$ constitute the angular part of the solutions of Laplace's equation in three dimensions, and thus form a complete orthonormal set. The orthogonality relation is

$$\int_{S^2} Y_{n_1}^{m_1*}(\theta, \phi) Y_{n_2}^{m_2}(\theta, \phi) d\mu(\theta, \phi) = \delta_{n_1, n_2} \delta_{m_1, m_2},$$

where z^* denotes complex conjugate of z .

For completeness, the solution u and input $I(\theta, \phi)$ can be written as

$$u(\theta, \phi, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^n a_{nm}(t) y_n^m(\theta, \phi),$$

$$I(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n I_{nm}(t) y_n^m(\theta, \phi),$$

where y_n^m are the corresponding real-valued spherical harmonics

$$y_n^m = \begin{cases} \frac{1}{\sqrt{2}} [Y_n^{-m} + (-1)^m Y_n^m], & m > 0 \\ Y_n^m, & m = 0 \\ \frac{i}{\sqrt{2}} [Y_n^m - (-1)^m Y_n^{-m}], & m < 0. \end{cases} \tag{5.3}$$

Note that y_n^m are also orthonormal and

$$\langle y_n^m, Y_i^j \rangle = \begin{cases} \frac{1}{\sqrt{2}} [\delta_{ni} \delta_{-m,j} + (-1)^m \delta_{n,i} \delta_{m,j}], & m > 0 \\ \delta_{n,i} \delta_{0,j}, & m = 0 \\ -\frac{i}{\sqrt{2}} [\delta_{n,i} \delta_{m,j} - (-1)^m \delta_{n,i} \delta_{-m,j}], & m < 0. \end{cases}$$

Substituting the expansions for u and I into Eq. (5.1) and using Eq. (5.2) yields

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{da_{nm}(t)}{dt} y_n^m(\theta, \phi) = - \sum_{n=0}^{\infty} \sum_{m=-n}^n a_{nm}(t) y_n^m(\theta, \phi) + \sum_{n=0}^{\infty} \sum_{m=-n}^n I_{nm} y_n^m(\theta, \phi) + \sum_{n=0}^{\infty} w_n \sum_{m=-n}^n (-1)^m Y_n^m(\theta, \phi) \int_{S^2} Y_n^{-m}(\theta', \phi') f(u(\theta', \phi')) d\mu(\theta', \phi').$$

Taking the inner product of both sides with each Y_n^m and using orthonormality yields

$$\frac{da_{nm}(t)}{dt} = -a_{nm}(t) + I_{nm} + w_n \int_{S^2} f \left(\sum_{i=0}^{\infty} \sum_{j=-i}^i a_{ij}(t) y_i^j(\theta', \phi') \right) y_n^m(\theta', \phi') d\mu(\theta', \phi'). \tag{5.4}$$

If K has a finite expansion, then we can pose this as an ODE in a finite-dimensional space. Suppose that $w_n = 0$ for $n > N$ so that a_{nm} exponentially approaches I_{nm} . Analogous to the ring model we can take $a_{nm} = I_{nm}$ for $n > N$ and focus on the finite-dimensional space

$$W_N^S = \left\{ \sum_{n=1}^N \sum_{m=-n}^n a_{nm} y_n^m(\theta, \phi) \right\} \cong \mathbb{R}^{N(N+1)+1}. \tag{5.5}$$

Defining the vectors

$$\mathbf{x}_n = \begin{bmatrix} a_{n,-n} \\ \vdots \\ a_{n,-1} \\ a_{n,0} \\ a_{n,1} \\ \vdots \\ a_{n,n} \end{bmatrix}, \quad \mathbf{I}_n = \begin{bmatrix} I_{n,-n} \\ \vdots \\ I_{n,-1} \\ I_{n,0} \\ I_{n,1} \\ \vdots \\ I_{n,n} \end{bmatrix}, \quad \mathbf{y}_n = \begin{bmatrix} y_n^{-n} \\ \vdots \\ y_n^{-1} \\ y_n^0 \\ y_n^1 \\ \vdots \\ y_n^n \end{bmatrix}$$

and the scalar function

$$V = \sum_{n=0}^N \frac{1}{w_n} \left[\frac{1}{2} \|\mathbf{x}_n\|^2 - \mathbf{x}_n \cdot \mathbf{I}_n \right] - \int_{S^2} F \left(\sum_{n=0}^N \mathbf{x}_n \cdot \mathbf{y}_n(\theta', \phi') + \sum_{n=N+1}^{\infty} \mathbf{I}_n \cdot \mathbf{y}_n(\theta', \phi') \right) d\mu(\theta', \phi'), \tag{5.6}$$

the system can be written as

$$\frac{d\mathbf{x}_n(t)}{dt} = -w_n \nabla_{\mathbf{x}_n} V, \quad n \leq N, \quad \nabla_{\mathbf{x}_n} V = (\partial_{a_{n,-n}} V, \dots, \partial_{a_{n,-1}} V, \partial_{a_{n,0}} V, \partial_{a_{n,1}} V, \dots, \partial_{a_{n,n}} V)^\top. \tag{5.7}$$

B. Three-dimensional gradient system

Following along similar lines to our analysis of ring attractor networks, we now consider the special case where $w_n = I_{nm} = 0$ for all $n \neq 1$. We then need only consider the first-order spherical harmonics

$$\begin{aligned} Y_1^0 &= \sqrt{\frac{3}{4\pi}} \cos(\theta), \\ Y_1^{-1} &= \frac{1}{2} \sqrt{\frac{6}{4\pi}} \sin(\theta) e^{-i\phi}, \\ Y_1^1 &= -\frac{1}{2} \sqrt{\frac{6}{4\pi}} \sin(\theta) e^{i\phi} \end{aligned}$$

such that

$$K(\theta, \phi) = w[\cos(\theta) \cos(\theta') + \sin(\theta) \sin(\theta') \cos(\phi - \phi')],$$

where we have absorbed a factor of $3\pi/4$ into w . Moreover, solutions take the form of spherical bumps

$$u(\theta, \phi, t) = \mathbf{x}(t) \cdot \mathbf{e}(\theta, \phi),$$

where $\mathbf{x} = (a_{1,1}, a_{1,-1}, a_{1,0}) \in \mathbb{R}^3$ and

$$\mathbf{e}(\theta, \phi) = (\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta))$$

is the unit vector on the sphere with angles θ and ϕ . We thus obtain the three-dimensional gradient system

$$\frac{d\mathbf{x}}{dt} = -\nabla V_0(\mathbf{x}), \quad (5.8)$$

where the potential is now defined as

$$V_0(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|^2 - \mathbf{x} \cdot \mathbf{I} - \int_{S^2} F(\mathbf{x} \cdot \mathbf{e}(\theta', \phi')) d\mu(\theta', \phi'), \quad (5.9)$$

with $\mathbf{I} = (I_{1,1}, I_{1,-1}, I_{1,0})$. Since the original neural field is SO(3) equivariant when $I = 0$, then $V_0(\mathbf{x})$ is SO(3) invariant and Eq. (5.8) is equivariant with respect to the standard action of SO(3) on \mathbb{R}^3 when $\mathbf{I} = \mathbf{0}$.

If we express \mathbf{x} in terms of spherical coordinates

$$\mathbf{x} = A\mathbf{e}(\Theta, \Phi), \quad (5.10)$$

then

$$\begin{aligned} u(\theta, \phi) &= \mathbf{x} \cdot \mathbf{e}(\theta, \phi) \\ &= A[\cos(\theta) \cos(\Theta) + \sin(\theta) \sin(\Theta) \cos(\phi - \Phi)] \end{aligned}$$

so that the solution obtains its maximum value A when $(\theta, \phi) = (\Theta, \Phi)$. Hence, A is the amplitude of the bump and (Θ, Φ) encodes the location of the peak of the bump. Similarly, if we set $\mathbf{I} = \bar{I}\mathbf{e}(\bar{\Theta}, \bar{\Phi})$, then \bar{I} is the input amplitude and $(\bar{\Theta}, \bar{\Phi})$ denotes the location of the input peak.

C. Stationary bump solutions and stability

First consider the homogeneous system with $\mathbf{I} = \mathbf{0}$. Stationary solutions are given by extremal values of $V_0(\mathbf{x})$. Since $V_0(\mathbf{x})$ is SO(3) symmetric, it follows that equilibria lie on spheres in \mathbb{R}^3 . We can therefore, without loss of generality, take $\mathbf{x}_0^* = a^*(0, 0, 1)$ and then apply the action of SO(3) to

obtain all other solutions. The implicit equation for a^* is

$$\begin{aligned} a^* &= \mathcal{G}(a^*) \equiv w \int_{S^2} f(a^* \cos(\theta')) \cos(\theta') d\mu(\theta', \phi') \\ &= 2\pi w \int_0^\pi f(a^* \cos(\theta')) \cos(\theta') \sin(\theta') d\theta'. \end{aligned} \quad (5.11)$$

Once a solution for a is established, the group orbit of solutions is given by the two-parameter family of solutions $\mathbf{x}^* = a^*(\sin(\Theta) \cos(\Phi), \sin(\Theta) \sin(\Phi), \cos(\Theta))$, $\forall \Theta \in [0, \pi)$, $\Phi \in [0, 2\pi)$. Linearizing the system about the equilibrium \mathbf{x}^* yields the eigenvalue problem $\lambda \mathbf{y} = H\mathbf{y}$, $\mathbf{y} \in \mathbb{R}^3$, where

$$\begin{aligned} H &= -\frac{\partial^2 V}{\partial x_i \partial x_j} \Big|_{\mathbf{x}=\mathbf{x}^*} \\ &= -I + w \int_{S^2} f'(\mathbf{x}^* \cdot \mathbf{e}(\theta, \phi)) \mathbf{e}(\theta, \phi) \mathbf{e}(\theta, \phi)^T d\mu(\theta, \phi). \end{aligned} \quad (5.12)$$

Since the system has SO(3) symmetry, if $\mathbf{x}^* \neq \mathbf{0}$, then there exist two null vectors for H , one for each independent direction of rotation. Let $R_i(s)$ be a rotation about the $i = x, y, z$ axes. Then, following an argument identical to the SO(2) case, the null vectors are given by

$$\frac{d}{ds} [R_x(s) \mathbf{x}^*] \Big|_{s=0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}^*,$$

$$\frac{d}{ds} [R_y(s) \mathbf{x}^*] \Big|_{s=0} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \mathbf{x}^*,$$

and

$$\frac{d}{ds} [R_z(s) \mathbf{x}^*] \Big|_{s=0} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x}^*.$$

If \mathbf{x}^* lies along one of the axes, then one of these vectors are zero and the other two necessarily nonzero. If it does not lie on any axis, then only two of these vectors are linearly independent. In general, let $\mathbf{x}_\theta^{*\perp}$ and $\mathbf{x}_\phi^{*\perp}$ denote the two nonzero independent vectors. In other words, these two vectors span the tangent space of the sphere at point \mathbf{x}^* . The analogous result for the original neural field is that if $u_0(\theta, \phi)$ is a solution, then $u_0(\theta - \theta_0, \phi - \phi_0)$ is also a solution for all θ_0, ϕ_0 . Thus, perturbations

$$u = u_0(\theta, \phi) + \varepsilon \partial_\theta u_0(\theta, \phi) + \varepsilon \partial_\phi u_0(\theta, \phi)$$

correspond to infinitesimally shifting θ and ϕ , while perturbations

$$\mathbf{x} = \mathbf{x}^* + \varepsilon \mathbf{x}_\theta^{*\perp} + \varepsilon \mathbf{x}_\phi^{*\perp}$$

correspond to infinitesimally rotating \mathbf{x}^* along the sphere.

We can then decompose the space into eigenspaces

$$\mathbb{R}^3 = \text{span}\{\mathbf{x}^*\} \oplus \text{span}\{\mathbf{x}_\theta^{*\perp}\} \oplus \text{span}\{\mathbf{x}_\phi^{*\perp}\}.$$

The nontrivial eigenvalue will correspond to eigenvectors in $\text{span}\{\mathbf{x}^*\}$, i.e., amplitude changing perturbations. Restricting

H to span $\{\mathbf{x}^*\}$ yields

$$H\mathbf{x}^* = -\mathbf{x}^* + w \int_{S^2} f'(\mathbf{x}^* \cdot \mathbf{e}(\theta, \phi)) \mathbf{e}(\theta, \phi) \times [\mathbf{x}^* \cdot \mathbf{e}(\theta, \phi)] d\mu(\theta, \phi).$$

Taking the dot product of both sides of $\lambda \mathbf{x}^* = H\mathbf{x}^*$ with \mathbf{x}^* yields

$$\begin{aligned} \lambda &= -1 + w \int_{S^2} f'(\mathbf{x}^* \cdot \mathbf{e}(\theta, \phi)) \cos^2(\alpha(\theta, \phi)) d\mu(\theta, \phi) \\ &= -1 + 2\pi w \int_0^\pi f'(a^* \cos(\theta)) \cos^2(\theta) \sin(\theta) d\theta \\ &= -1 + \mathcal{G}'(a^*), \end{aligned} \quad (5.13)$$

where $\alpha(\theta, \phi)$ is the angle between \mathbf{x}^* and point at (θ, ϕ) , the second equality follows from SO(3) invariance of the integral and the last equality follows from the definition of $\mathcal{G}(a)$ in Eq. (5.11). The dependence of existence and stability of equilibria on parameters is qualitatively the same as the ring model.

Let us now turn to the inhomogeneous spherical model. Following the same argument as for the ring model, we conclude that if \mathbf{x}^* is an equilibrium when $\mathbf{I} = \mathbf{I}_0$, then $R\mathbf{x}^*$ is an equilibrium solution when $\mathbf{I} = R\mathbf{I}_0$ for any rotation R in SO(3). Moreover, \mathbf{x}^* must still be collinear with \mathbf{I} . Thus, we may take $\mathbf{I} = \mathbf{I}_0 \equiv (0, 0, \bar{I})$ with $\bar{I} > 0$ and $\mathbf{x}_0^* = (0, 0, a^*)$ with a^* satisfying

$$a^* - \bar{I} = \mathcal{G}(a^*).$$

Again, positive values of a^* will lead to bump solutions that are in phase with the input while negative values lead to bump solutions that are antiphase.

There is no longer a two-dimensional null space, however, we can derive eigenvalues corresponding to phase changing perturbations. Let $\mathbf{I}_{0,\theta}^\perp$ and $\mathbf{I}_{0,\phi}^\perp$ be the vectors orthogonal to \mathbf{I}_0 with the same magnitude, i.e., the two linearly independent vectors from

$$\begin{aligned} \left. \frac{d}{ds} [R_x(s)\mathbf{I}_0] \right|_{s=0} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{I}_0, \\ \left. \frac{d}{ds} [R_y(s)\mathbf{I}_0] \right|_{s=0} &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \mathbf{I}_0, \\ \left. \frac{d}{ds} [R_z(s)\mathbf{I}_0] \right|_{s=0} &= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{I}_0. \end{aligned}$$

Using identical arguments as the ring model, it can easily be shown that

$$H\mathbf{x}_\theta^{*\perp} = -\mathbf{I}_{0,\theta}^\perp, \quad H\mathbf{x}_\phi^{*\perp} = -\mathbf{I}_{0,\phi}^\perp,$$

where $\mathbf{x}_\theta^{*\perp}$ and $\mathbf{x}_\phi^{*\perp}$ are the corresponding vectors orthogonal to \mathbf{x}^* . Therefore, $\mathbf{x}_\theta^{*\perp}$ and $\mathbf{x}_\phi^{*\perp}$ are eigenvectors with eigenvalue

$\lambda_o = -\bar{I}/a^*$, which is exactly the same as the ring model. All other stability results are qualitatively identical to the ring model.

D. Stochastic spherical model

We now consider the stochastic spherical model

$$du = \left[-u + \int_{-\pi}^\pi K(\theta, \phi|\theta', \phi') f(u(\theta', \phi')) d\mu(\theta', \phi') + I(\theta, \phi) \right] dt + dW(\theta, \phi, t), \quad (5.14)$$

with $\mathbb{E}[dW] = 0$ and

$$\mathbb{E}[dW(\theta, \phi, t) dW(\theta', \phi', t')] = C(\theta, \phi|\theta', \phi') \delta(t - t') dt dt'.$$

Analogous to the ring model, we take C to be SO(3) invariant, i.e., $C(\gamma^{-1}(\theta, \phi)|\gamma^{-1}(\theta', \phi')) = C(\theta, \phi|\theta', \phi')$ and thus C has the expansion

$$C(\theta, \phi|\theta', \phi') = \sum_{n=0}^{\infty} \sigma_n \sum_{m=-n}^n (-1)^m Y_n^m(\theta, \phi) Y_n^{m*}(\theta', \phi') \quad (5.15)$$

with $\sigma_n > 0$ for $n \leq N$ so that the corresponding operator $\int_{S^2} C(\theta, \phi|\theta', \phi') u(\theta', \phi') d\mu(\theta', \phi')$ is positive definite on W_N^S . Analogous to the ring model we obtain

$$d\mathbf{x}_n = -w_n \nabla_{\mathbf{x}_n} V dt + \sqrt{\sigma_n} d\mathbf{W}_n, \quad n \leq N \quad (5.16a)$$

$$d\mathbf{x}_n = [-\mathbf{x}_n + \mathbf{I}_n] dt + \sqrt{\sigma_n} d\mathbf{W}_n, \quad N < n \leq M \quad (5.16b)$$

where each \mathbf{W}_n is a $(2n+1)$ -dimensional uncorrelated Wiener process with unit variance.

Let us consider the simplest case $N = M = 1$ and $\mathbf{I}_n = \mathbf{0}$ for $n \neq 1$, so that we have the 3D stochastic gradient system

$$d\mathbf{x} = -\nabla V_0(\mathbf{x}) dt + \sqrt{\sigma} d\mathbf{W}, \quad \mathbf{x} = (a_{1,1}, a_{1,-1}, a_{1,0}).$$

The corresponding FP equation for the probability density $p(\mathbf{x}, t)$ is given by Eq. (3.4), except now ∇^2 is the 3D Laplacian and $V_0(\mathbf{x})$ is given by Eq. (5.9). Moreover, the stationary density has the form of Eq. (3.5) with the appropriate normalization. Given the fact that we can express the stochastic dynamics of the spherical model as a gradient system means that we can generalize the various results obtained for the ring.

In particular, in the absence of an external input, $V_0(\mathbf{x}) \rightarrow U_0(\mathbf{x})$ with $U_0(\mathbf{x})$ spherically symmetric. It follows that the marginal density for Θ, Φ is given by the uniform distribution on the sphere. Converting the full time-dependent FP equation (3.4) into spherical polar coordinates (5.10) with

$$p(\mathbf{x}, t) d\mathbf{x} = p_s(A, \Theta, \Phi, t) A^2 \sin \Theta dA d\Theta d\Phi$$

gives

$$\begin{aligned} \frac{\partial p_s}{\partial t} = & \frac{1}{A^2} \frac{\partial}{\partial A} \left\{ A^2 \left[\frac{\partial U_0(A)}{\partial A} - \bar{I} \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}) \right] p_s \right\} - \frac{\bar{I}}{A \sin \Theta} \frac{\partial}{\partial \Theta} [\sin \Theta \partial_\Theta \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}) p_s] \\ & - \frac{\bar{I}}{A \sin^2 \Theta} \frac{\partial}{\partial \Phi} [\partial_\Phi \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}) p_s] + \frac{\sigma}{2} \left[\frac{1}{A^2} \frac{\partial}{\partial A} \left(A^2 \frac{\partial}{\partial A} \right) + \frac{1}{A^2 \sin \Theta} \frac{\partial}{\partial \Theta} \left(\sin \Theta \frac{\partial}{\partial \Theta} \right) + \frac{1}{A^2 \sin^2 \Theta} \frac{\partial^2}{\partial \Phi^2} \right] p_s. \end{aligned} \quad (5.17)$$

The corresponding SDE in spherical polar coordinates is thus

$$dA = \left[-\partial_A U_0(A) + \frac{\sigma}{A} + \bar{I} \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}) \right] dt + \sqrt{\sigma} dW_A, \quad (5.18a)$$

$$\begin{aligned} d\Theta = & \left[\frac{\bar{I}}{A \sin \Theta} \partial_\Theta \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}) + \frac{\sigma}{2A^2} \cot(\Theta) \right] dt \\ & + \frac{\sqrt{\sigma}}{A} dW_\Theta, \end{aligned} \quad (5.18b)$$

$$d\Phi = \frac{\bar{I}}{A \sin^2 \Theta} \partial_\Phi \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}) dt + \frac{\sqrt{\sigma}}{A \sin \Theta} dW_\Phi, \quad (5.18c)$$

where $W_A(t)$ and $W_\Theta(t)$ and $W_\Phi(t)$ are independent Wiener processes with unit variance.

Similar to the ring model, in the absence of an input, the amplitude dynamics decouples from the phase dynamics with the stationary phase distribution being uniform over the sphere and the marginal stationary density for the amplitude is

$$p_A^h(A) = \frac{1}{Z_h} A^2 e^{-2U_0(A)/\sigma}, \quad (5.19)$$

with

$$U_0(A) = \frac{1}{2} A^2 - 2\pi w \int_0^\pi F(A \cos \theta) \sin \theta d\theta, \quad (5.20)$$

and $Z_h = \int_0^\infty A^2 e^{-2U_0(A)/\sigma} dA$. Moreover, having obtained the homogeneous density, one can determine moments of the amplitude and phases in the inhomogeneous case along analogous lines to the ring, using a higher-dimensional version of the von Mises distribution (see below). Finally, the mean bump extinction time can be obtained in the same manner as

the ring model. The result is identical to Eq. (4.3) except for changing $a^*/a_0 \rightarrow (a^*/a_0)^2$.

Mathematically speaking, there is one interesting difference between the sphere and the ring, namely, the former has intrinsic curvature. This is reflected by the nature of Brownian motion on the surface of the sphere. Focusing on the phase variables, the corresponding SDE is obtained by setting $\bar{I} = 0$ in Eqs. (5.18b) and (5.18c) which reduces to the form

$$d\Theta = \frac{\sigma}{2A^2} \cot(\Theta) dt + \frac{\sqrt{\sigma}}{A} dW_\Theta, \quad d\Phi = \frac{\sqrt{\sigma}}{A \sin \Theta} dW_\Phi. \quad (5.21)$$

The drift term in the first equation is a consequence of the fact that the sphere is a curved manifold. Finally, note that Eqs. (5.18b) and (5.18c) for $A = a^*$ and $\bar{I} \neq 0$ are consistent with the phase equations recently derived for a more general class of correlation functions using perturbation theory [43].

E. Moments of inhomogeneous network

We now show how to relate the homogeneous and inhomogeneous probability densities of the stochastic spherical model, using a higher-dimensional analog of the von Mises distribution. This provides compact analytical expressions for the various moments in a similar fashion to the ring model. In the inhomogeneous case, we have

$$V_0(A, \Theta, \Phi) = U_0(A) - \bar{I} A \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}).$$

Therefore,

$$\begin{aligned} p^{ih}(A, \Theta, \Phi) & \equiv A^2 \sin(\Theta) p(A, \Theta, \Phi) \\ & = \frac{Z_h}{Z} \sin(\Theta) e^{2\bar{I} A \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi})/\sigma} p_A^h(A), \end{aligned}$$

with $p_A^h(A)$ given by Eq. (5.19). The moments of A and directional moments of $\mathbf{e}(\Theta, \Phi)$ are given by

$$\mathbb{E}[A^n \mathbf{e}(\Theta, \Phi)]_{ih} = \zeta \int_0^\infty dA A^n p_A^h(A) \left[\frac{1}{4\pi} \int_{S^2} \mathbf{e}(\Theta, \Phi) e^{2\bar{I} A \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi})/\sigma} d\mu(\Theta, \Phi) \right],$$

$$\mathbb{E}[A^n \mathbf{e}(\Theta, \Phi) \mathbf{e}(\Theta, \Phi)^T] = \zeta \int_0^\infty dA A^n p_A^h(A) \left[\frac{1}{4\pi} \int_{S^2} \mathbf{e}(\Theta, \Phi) \mathbf{e}(\Theta, \Phi)^T e^{2\bar{I} A \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi})/\sigma} d\mu(\Theta, \Phi) \right],$$

where $\zeta = 4\pi Z_h/Z$. The term inside the brackets can be related to the Fisher distribution on the sphere given by [44]

$$p_F(\mathbf{x}) = \frac{\kappa}{4\pi \sinh(\kappa)} e^{\kappa \mathbf{n} \cdot \mathbf{x}}, \quad \mathbf{x}, \mathbf{n} \in S^2$$

with \mathbf{n} denoting the mean direction, and κ is a parameter. The directional moments of the Fisher distribution are given by [44]

$$m_1(\kappa)\mathbf{n} \equiv \int_{S^2} \mathbf{x} p_F(\mathbf{x}) d\mathbf{x} = \left(\coth \kappa - \frac{1}{\kappa} \right) \mathbf{n},$$

$$m_2(\kappa) \equiv \int_{S^2} \mathbf{x}\mathbf{x}^T p_F(\mathbf{x}) d\mathbf{x} = \frac{m_1(\kappa)}{\kappa} I_{3 \times 3} + \left(1 - \frac{3m_1(\kappa)}{\kappa} \right) \mathbf{n}\mathbf{n}^T,$$

where $I_{3 \times 3}$ is the 3×3 identity matrix. Setting $\kappa = \beta A$ for $\beta = 2\bar{l}/\sigma$, and $\mathbf{n} = \mathbf{e}(\bar{\Theta}, \bar{\Phi})$ in the expressions for $m_1(\kappa)$ and $m_2(\kappa)$, we see that

$$\frac{1}{4\pi} \int_{S^2} e^{2\bar{l}A\mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi})/\sigma} d\mu(\Theta, \Phi) = C(\beta A),$$

$$\frac{1}{4\pi} \int_{S^2} \mathbf{e}(\Theta, \Phi) e^{2\bar{l}A\mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi})/\sigma} d\mu(\Theta, \Phi) = C(\beta A) m_1(\beta A) \mathbf{e}(\bar{\Theta}, \bar{\Phi}),$$

$$\frac{1}{4\pi} \int_{S^2} \mathbf{e}(\Theta, \Phi) \mathbf{e}(\Theta, \Phi)^T e^{2\bar{l}A\mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi})/\sigma} d\mu(\Theta, \Phi) = C(\beta A) m_2(\beta A),$$

where $C(\kappa) = \sinh(\kappa)/\kappa$.

It follows that

$$\mathbb{E}[A^n]_{ih} = \zeta \mathbb{E}[A^n C(\beta A)]_h, \quad (5.22a)$$

$$\mathbb{E}[A^n \mathbf{e}(\Theta, \Phi)]_{ih} = \zeta \mathbb{E}[A^n C(\beta A) m_1(\beta A)]_h \mathbf{e}(\bar{\Theta}, \bar{\Phi}), \quad (5.22b)$$

$$\mathbb{E}[A^n \mathbf{e}(\Theta, \Phi) \mathbf{e}(\Theta, \Phi)^T]_{ih} = \zeta \mathbb{E}[A^n C(\beta A) m_2(\beta A)]_h. \quad (5.22c)$$

For example, the mean, variance of A , and covariance matrix of $\mathbf{e}(\Theta, \Phi)$ are given by

$$\langle A \rangle = \zeta \mathbb{E}[AC(\beta A)]_h, \quad (5.23a)$$

$$\langle \mathbf{e}(\Theta, \Phi) \rangle = \zeta \mathbb{E}[C(\beta A) m_1(\beta A)]_h \mathbf{e}(\bar{\Theta}, \bar{\Phi}), \quad (5.23b)$$

and

$$\text{Var}[A] = \zeta \mathbb{E}[A^2 C(\beta A)]_h - \zeta^2 \mathbb{E}[AC(\beta A)]_h^2, \quad (5.24a)$$

$$\text{cov}[\mathbf{e}(\Theta, \Phi)] = \zeta \mathbb{E}[C(\beta A) m_2(\beta A)]_h - \zeta^2 \mathbb{E}[C(\beta A) m_1(\beta A)]_h^2 \mathbf{e}(\bar{\Theta}, \bar{\Phi}) \mathbf{e}(\bar{\Theta}, \bar{\Phi})^T, \quad (5.24b)$$

respectively. The mean and covariance of $\mathbf{e}(\Theta, \Phi)$ depend on the input phases $(\bar{\Theta}, \bar{\Phi})$. It would be useful to compute the mean and covariance of a quantity that is invariant under changes in the input phase. To that end, consider the vector

$$\tilde{\mathbf{e}}(\Theta, \Phi) = R(\bar{\Theta}, \bar{\Phi})^{-1} \mathbf{e}(\Theta, \Phi),$$

where R is the rotation matrix such that

$$R(\bar{\Theta}, \bar{\Phi}) \mathbf{e}_3 = \mathbf{e}(\bar{\Theta}, \bar{\Phi}), \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Then,

$$\langle \tilde{\mathbf{e}}(\Theta, \Phi) \rangle = \zeta \mathbb{E}[C(\beta A) m_1(\beta A)]_h \mathbf{e}_3, \quad (5.25)$$

$$\text{cov}[\tilde{\mathbf{e}}(\Theta, \Phi)] = \zeta \mathbb{E}[C(\beta A) D_2(\beta A)] - \zeta^2 \mathbb{E}[C(\beta A) m_1(\beta A)]_h^2 \mathbf{e}_3 \mathbf{e}_3^T, \quad (5.26)$$

where

$$D_2(\kappa) = \begin{bmatrix} \frac{m_1(\kappa)}{\kappa} & 0 & 0 \\ 0 & \frac{m_1(\kappa)}{\kappa} & 0 \\ 0 & 0 & 1 - \frac{2m_1(\kappa)}{\kappa} \end{bmatrix}.$$

Thus, the covariance matrix is diagonal and the directions are uncorrelated. In particular, the phase variables Θ and Φ are uncorrelated.

In the ring model, we calculated statistics of the quantity $\cos(\Delta - \bar{\Delta})$. The analogous quantity here is $\cos(\alpha(\Theta, \Phi | \bar{\Theta}, \bar{\Phi}))$, where α is the coplanar angle between (Θ, Φ) and $(\bar{\Theta}, \bar{\Phi})$. Since

$$\cos(\alpha) = \mathbf{e}(\Theta, \Phi) \cdot \mathbf{e}(\bar{\Theta}, \bar{\Phi}),$$

it follows that the first two moments are given by

$$\langle \cos(\alpha) \rangle = \mathbf{e}(\bar{\Theta}, \bar{\Phi})^T \langle \mathbf{e}(\Theta, \Phi) \rangle = \zeta \mathbb{E}[C(\beta A) m_1(\beta A)]_h, \quad (5.27)$$

$$\langle \cos^2(\alpha) \rangle = \mathbf{e}(\bar{\Theta}, \bar{\Phi})^T \mathbb{E}[\mathbf{e}(\Theta, \Phi) \mathbf{e}(\Theta, \Phi)^T]_{ih} \mathbf{e}(\bar{\Theta}, \bar{\Phi}) = \zeta \mathbb{E}[C(\beta A) \tilde{m}_2(\beta A)]_h, \quad (5.28)$$

where

$$\tilde{m}_2(\beta A) = \mathbf{e}(\bar{\Theta}, \bar{\Phi})^T m_2(\beta A) \mathbf{e}(\bar{\Theta}, \bar{\Phi}) = 1 - \frac{2m_1(\beta A)}{\beta A}.$$

Therefore, the variance is

$$\text{Var}[\cos(\alpha)] = \zeta \mathbb{E}[C(\beta A) \tilde{m}_2(\beta A)]_h - \zeta^2 \mathbb{E}[C(\beta A) m_1(\beta A)]_h^2. \quad (5.29)$$

Finally, we obtain the covariance

$$\begin{aligned} \text{cov}(A, \cos(\alpha)) &= \mathbf{e}(\bar{\Theta}, \bar{\Phi})^T \mathbb{E}[A \mathbf{e}(\Theta, \Phi)]_{ih} - \langle A \rangle \langle \cos(\alpha) \rangle \\ &= \zeta \mathbb{E}[AC(\beta A) m_1(\beta A)]_h - \zeta^2 \mathbb{E}[AC(\beta A)]_h \mathbb{E}[C(\beta A) m_1(\beta A)]_h. \end{aligned} \quad (5.30)$$

In the absence of input, taking $\beta \rightarrow 0$, we obtain

$$\zeta \rightarrow 1, \quad C(\beta A) \rightarrow 1, \quad m_1(\beta A) \rightarrow 0, \quad \tilde{m}_2(\beta A) \rightarrow \frac{1}{3}$$

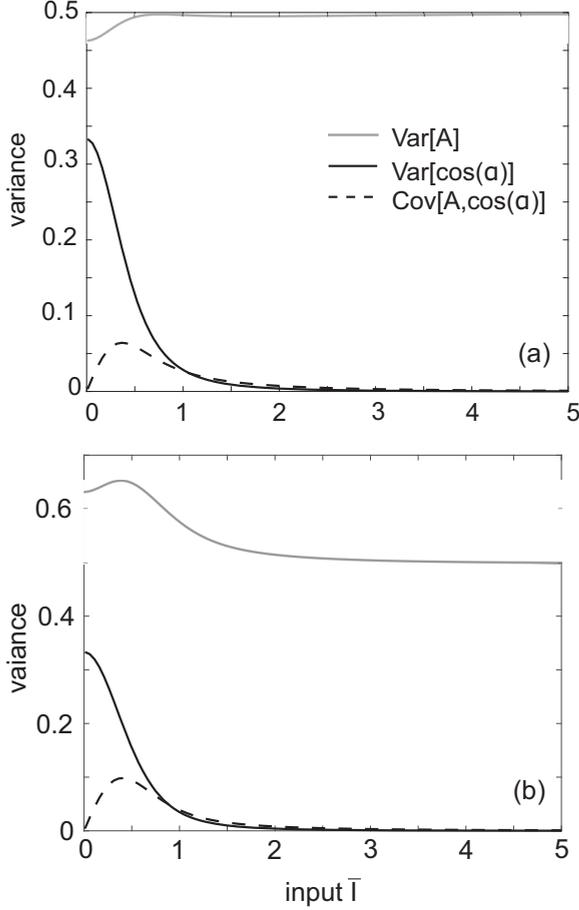


FIG. 15. Variances and covariances of A and $\cos(\alpha)$ as a function of input strength for (a) $\kappa = 0.5$ and (b) $\kappa = 0.9$. All other parameters are the same as the ring model except we set $w = 0.7$ so that the stable amplitude is relatively the same.

and thus

$$\text{Var}[\cos(\alpha)] = \frac{1}{3}, \quad \text{cov}(A, \cos(\alpha)) = 0,$$

with the latter showing that the amplitude and phase are uncorrelated. Including an input introduces correlations.

Finally, the marginal densities are given by

$$p_A(A) = \zeta C(\beta A) p_A^h(A), \quad (5.31a)$$

$$p_{\Theta, \Phi}(\Theta, \Phi) = \frac{\zeta \sin(\Theta)}{4\pi} \mathbb{E}[e^{2\bar{I}A \cos(\alpha)/\sigma}]_h. \quad (5.31b)$$

We see that the marginal phase density is similar in form to the phase density for the ring model, with α analogous to $\Delta - \bar{\Delta}$ and the former having the prefactor of $\sin(\Theta)$ which is due to the curvature of the sphere. Note that, although Θ and Φ are uncorrelated for any \bar{I} , when $\bar{I} \neq 0$ they are dependent since the joint density is not separable. On the other hand, when $\bar{I} \rightarrow 0$ we obtain

$$p_{\Theta, \Phi}(\Theta, \Phi) \rightarrow \frac{\sin(\Theta)}{4\pi}, \quad p_{\Theta}(\Theta) \rightarrow \frac{\sin(\Theta)}{2}, \quad p_{\Phi}(\Phi) \rightarrow \frac{1}{2\pi},$$

which are the joint and marginal densities for the uniform distribution over a sphere. In this case, the phases are uncorrelated and independent.

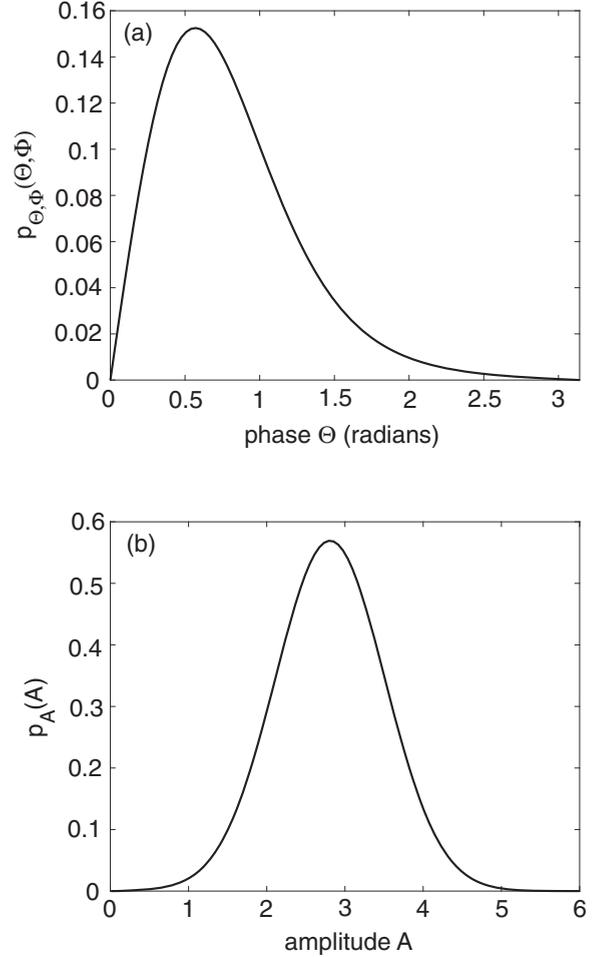


FIG. 16. Marginal densities for (a) phase and (b) amplitude. Note that the density for the phase is constant in Φ for input phase $\bar{\Theta} = \bar{\Phi} = 0$.

Finally, the phase density conditioned on the amplitude is

$$\begin{aligned} p_{(\Theta, \Phi)|A}(\Theta, \Delta) &\equiv \frac{p_{th}(A, \Theta, \Phi)}{p_A(A)} \\ &= \frac{\sin(\Theta)}{C(\beta A)} e^{2\bar{I}A e^{i(\Theta, \Phi)} \cdot e^{i(\bar{\Theta}, \bar{\Phi})}/\sigma}, \end{aligned}$$

which has the same form as the Fisher distribution after converting to angular coordinates, i.e., $\mathbf{x} = \mathbf{e}(\theta, \phi)$, $\mathbf{n} = \mathbf{e}(\bar{\theta}, \bar{\phi})$.

In Fig. 15 we plot the steady-state variance and covariance as a function of the input strength \bar{I} . The results are qualitatively similar to the ring model. Finally in Fig. 16 we show example plots of the marginal densities $p_A(A)$ and $p_{\Theta, \Phi}(\Theta, \Phi)$ for $\bar{\Theta} = \bar{\Phi} = 0$. Note that, for this choice of input, the phase density is constant in Φ and thus we plot it against Θ only. These are the spherical analogs of the densities shown in Figs. 10(b) and 11(b) for the ring model.

VI. DISCUSSION

In this paper we have demonstrated how a combination of stochastic analysis and group theoretic methods provides a powerful tool for investigating the effects of noise in continuous attractor networks. In particular, by imposing certain

constraints on the synaptic weight function and intrinsic noise, we derived an exact SDE for both the amplitude and phase of bump solutions, which was valid at arbitrary levels of noise. Moreover, the SDE could be expressed as a stochastic gradient dynamical system, which allowed us to derive an exact expression for the steady-state probability density and its moments. We focused on the example of a ring attractor, and used our theory to investigate the noise-induced suppression of amplitude and phase fluctuations. We showed that increasing the input greatly suppresses steady-state phase fluctuations by locking the location of the peak of the bump, consistent with a recent study based on perturbation methods [25]. This differs from the mechanism for noise-induced suppression of amplitude fluctuations, which is based on the suppression of noise-induced transitions of the bump to other attractors of the system. In cases where such transitions are rare events, input suppression of amplitude fluctuations is negligible. We also calculated the mean time for bump extinction due to a noise-induced transition to a zero state or a low amplitude bump. These transitions contribute to the steady-state variance in the amplitude, but would be rare events in the case of weak noise. Finally, we showed how our theory could be extended to higher-dimensional spherical attractor networks, where we exploited the underlying SO(3) symmetry of the homogeneous network to analyze the resulting gradient dynamical system.

In order to use the theory of gradient dynamical systems, we have assumed that the weight kernel can be expanded as a finite Fourier series with a single dominant mode $\cos\theta$ (in the case of ring attractor networks). This is consistent with the observation that population tuning curves in visual cortex tend to be unimodal. If the weight kernel is just a single mode, then we have an exact gradient system, whereas the inclusion of a finite number of higher-order modes in the weight kernel leads to a “scaled gradient” system coupled with an Ornstein-Uhlenbeck process [see Eq. (3.2)]. (Note that our analytical methods would not be applicable to weight kernels that have infinite Fourier series expansions. However, it is common to approximate such kernels by truncated Fourier series, which could then be incorporated into our framework.) One possibility for analyzing the effects of higher-order modes is to treat the higher-order coefficients as small compared to the first-order coefficient and combine perturbation theory with the methods presented here. Including higher modes in the weight kernel also requires the inclusion of higher-order modes in the noise. (Below we describe a possible method for analyzing the effects of higher-order noise terms.) Interestingly, the restriction to low-order noise in order to formulate the dynamics as a gradient system does not appear to be a severe constraint, since higher-order noise yields qualitatively, and in many cases, quantitatively similar behavior.

There are a number of possible extensions of the gradient analysis developed in this paper, as we now highlight.

A. Modeling the effects of higher-order noise as a fluctuating barrier

If higher-order noise terms are small, $\sigma_n \ll \sigma_1$ for all $n \neq 1$, then it is possible to reformulate the scaled gradient system (3.2) for $N = 1$ as a planar gradient system with a fluctuating

barrier. For the sake of illustration, suppose that $\sigma_n = \sigma \delta_{n,1} + q\delta_{n,M}$ with $q \ll \sigma$. Equation (3.2) then becomes

$$d\mathbf{x} = -\nabla_{\mathbf{x}}V(\mathbf{x}, \mathbf{y})dt + \sqrt{\sigma}d\mathbf{W}_1, \quad (6.1a)$$

$$d\mathbf{y} = -\mathbf{y}dt + \sqrt{q}d\mathbf{W}_2, \quad (6.1b)$$

where

$$V(\mathbf{x}, \mathbf{y}) = \left[\frac{1}{2}\|\mathbf{x}\|^2 - \mathbf{x} \cdot \mathbf{I} \right] - w \int_{-\pi}^{\pi} F[\mathbf{x} \cdot \mathbf{e}(\theta') + \mathbf{y} \cdot \mathbf{e}(M\theta')]d\theta'. \quad (6.2)$$

For small q we can treat \mathbf{y} as small with high probability and thus Taylor expand the potential to first order in \mathbf{y} . This gives

$$V(\mathbf{x}, \mathbf{y}) \approx V_0(\mathbf{x}) + \mathbf{y} \cdot \mathbf{Z}(\mathbf{x}), \quad (6.3)$$

with $V_0(\mathbf{x})$ given by Eq. (2.9) and

$$\mathbf{Z}(\mathbf{x}) = -w \int_{-\pi}^{\pi} f(\mathbf{x} \cdot \mathbf{e}(\theta'))\mathbf{e}(M\theta')^T d\theta'. \quad (6.4)$$

Applying this approximation to Eq. (6.1), we obtain the SDE

$$d\mathbf{x} = [-\nabla_{\mathbf{x}}V_0(\mathbf{x}) - \mathbf{y}^T \mathbf{B}(\mathbf{x})]dt + \sqrt{\sigma}d\mathbf{W}_1, \quad (6.5a)$$

$$d\mathbf{y} = -\mathbf{y}dt + \sqrt{q}d\mathbf{W}_2, \quad (6.5b)$$

where

$$\mathbf{B}(\mathbf{x}) = -w \int_{-\pi}^{\pi} f'(\mathbf{x} \cdot \mathbf{e}(\theta'))\mathbf{e}(\theta')\mathbf{e}(M\theta')^T d\theta'.$$

Since we have assumed that $q \ll \sigma$, we can introduce a second small parameter ε with $\sqrt{q} = \varepsilon\sqrt{\sigma}$. Following Ref. [45], perform the rescaling $\tilde{\mathbf{y}}(t) = \mathbf{y}(t)/\varepsilon$ so that Eqs. (6.5) become, after dropping the tildes and setting $\mathbf{I} = \mathbf{0}$,

$$d\mathbf{x} = [-\nabla_{\mathbf{x}}U_0(\mathbf{x}) - \varepsilon\mathbf{y}^T \mathbf{B}(\mathbf{x})]dt + \sqrt{\sigma}d\mathbf{W}_1, \quad (6.6a)$$

$$d\mathbf{y} = -\mathbf{y}dt + \sqrt{\sigma}d\mathbf{W}_2. \quad (6.6b)$$

Note that the systems \mathbf{x}, \mathbf{y} decouple when $\varepsilon = 0$. The corresponding FP equation for the probability density $p(\mathbf{x}, \mathbf{y}, t)$ takes the form

$$\begin{aligned} \frac{\partial p}{\partial t} = & \nabla_{\mathbf{x}} \cdot [\nabla_{\mathbf{x}}U_0(\mathbf{x})p(\mathbf{x}, \mathbf{y}, t)] + \frac{\sigma}{2}\nabla_{\mathbf{x}}^2 p(\mathbf{x}, \mathbf{y}, t) \\ & + \varepsilon\nabla_{\mathbf{x}} \cdot [\mathbf{y}^T \mathbf{B}(\mathbf{x})p(\mathbf{x}, \mathbf{y}, t)] \\ & + \nabla_{\mathbf{y}} \cdot [\mathbf{y}p(\mathbf{x}, \mathbf{y}, t)] + \frac{\sigma}{2}\nabla_{\mathbf{y}}^2 p(\mathbf{x}, \mathbf{y}, t). \end{aligned} \quad (6.7)$$

Equations (6.5) or (6.6) are examples of overdamped Brownian motion in the presence of a fluctuating barrier. Such systems have previously arisen in a wide variety of contexts (see the review [46]). In the 1990's, fluctuating barriers generated considerable interest following the observation by Doering and Gouda [47] that the mean escape time across a fluctuating barrier may exhibit a nonmonotonous dependence on the characteristic timescale of the fluctuations, so-called *resonant activation*. The particular fluctuating barrier model that most resembles Eqs. (6.5) was considered in Refs. [45,48], and consisted of a one-dimensional gradient dynamical system with additive noise, with fluctuations in the potential a linear function of a scalar OU process. However, there are a number of major differences between this model

and Eqs. (6.5). First, both \mathbf{x} and \mathbf{y} are two-dimensional rather than one-dimensional stochastic processes. Second, the characteristic timescale of the fluctuating barrier in Eq. (6.5b) is fixed. That is, the relaxation rates of both \mathbf{x} and \mathbf{y} are determined by the membrane time constant τ , which we have set to unity. Hence, one cannot independently modify the correlation time of the colored noise. Third, in the absence of noise, the four-dimensional dynamical system has a continuous circle of fixed points, rather than isolated fixed points, which are given by $(\mathbf{x}, \mathbf{y}) = (a^* \cos \theta, a^* \sin \theta, 0, 0)$ with a^* a stationary point of the radially symmetric potential $U(a)$. [It is not possible to reduce the \mathbf{x} dynamics to radial dynamics since the matrix $\mathbf{B}(\mathbf{x})$ is not radially symmetric.]

In future work it would be interesting to calculate how the fluctuating barrier due to higher-order noise affects the mean escape time (4.3) for bump extinction. Calculating escape times for multidimensional models with fluctuating barriers is generally intractable, unless some additional approximation is made such as a separation of timescales or a weak noise limit [45]. Even then, one can usually only determine the exponentially leading part (the quasipotential).

B. Other topological attractor networks

Another possible extension of our work would be to consider other topological attractor networks. Examples include the torus with $SO(2) \times SO(2)$ symmetry and the Poincaré disk with its group of isometries. (The latter is a model of two-dimensional hyperbolic geometry in which the points of the geometry are inside the unit disk, and the straight lines consist of all segments of circles contained within that disk that are orthogonal to the boundary of the disk, plus all diameters of

the disk. It has been incorporated into a neural field model of texture processing [14].) Yet another example is a spatially extended ring (or spherical) model [49]. Suppose that the associated neural field equations on $\mathbb{R}^d \times S^1$ take the product form [50,51]

$$\frac{\partial u(x, \theta, t)}{\partial t} = -u(x, \theta, t) + \int_{\mathbb{R}^2 \times S^1} J(\mathbf{x} - \mathbf{x}') K(\theta - \theta') \times f(u(\mathbf{x}', \theta', t)) d\mathbf{x}' d\theta'. \quad (6.8)$$

If we take $K(\theta) = w \cos(\theta)$, then we can also write this as a gradient system in the following way. Define the vector-valued function $\mathbf{v}(\mathbf{x}, t) = (a(\mathbf{x}, t), b(\mathbf{x}, t))^T$ and potential

$$V(\mathbf{v}(\mathbf{x})) = \frac{1}{2} \|\mathbf{v}(\mathbf{x})\|^2 - w \int_{\mathbb{R}^2 \times S^1} J(\mathbf{x} - \mathbf{x}') F(\mathbf{v}(\mathbf{x}') \cdot \mathbf{e}(\theta')) d\theta' d\mathbf{x}'.$$

Equation (6.8) can then be rewritten as

$$\frac{\partial \mathbf{v}(\mathbf{x}, t)}{\partial t} = -\nabla_{\mathbf{v}} V.$$

However, it remains to be seen to what extent such a formulation is useful when the Fourier components evolve according to a partial differential equation, rather than a finite-dimensional system of ordinary differential equations.

ACKNOWLEDGMENT

P.C.B. and S.R.C. were supported by the National Science Foundation (Grant No. DMS 1613048).

-
- [1] X. J. Wang, Synaptic reverberation underlying mnemonic persistent activity, *Trends Neurosci.* **24**, 455 (2001).
- [2] C. E. Curtis, Prefrontal and parietal contributions to spatial working memory, *Neuroscience* **139**, 173 (2006).
- [3] M. Camperi and X. J. Wang, A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability, *J. Comput. Neurosci.* **5**, 383 (1998).
- [4] K. Zhang, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory, *J. Neurosci.* **16**, 2112 (1996).
- [5] 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**, 910 (2000).
- [6] D. Durstewitz, J. K. Seamans, and T. J. Sejnowski, Neurocomputational models of working memory, *Nat. Neurosci.* **3**, 1184 (2000).
- [7] B. Gutkin, C. Laing, C. Colby, C. Chow, and G. B. Ermentrout, Turning on and off with excitation: The role of spike-timing asynchrony and synchrony in sustained neural activity, *J. Comput. Neurosci.* **11**, 121 (2001).
- [8] C. Constantinidis and X. J. Wang, A neural circuit basis for spatial working memory, *Neuroscientist* **10**, 553 (2004).
- [9] R. Ben-Yishai, R. Lev Bar-Or, and H. Sompolinsky, Theory of orientation tuning in visual cortex, *Proc. Natl. Acad. Sci. USA* **92**, 3844 (1995).
- [10] D. C. Somers, S. Nelson, and M. Sur, An emergent model of orientation selectivity in cat visual cortical simple cells, *J. Neurosci.* **15**, 5448 (1995).
- [11] P. C. Bressloff and J. D. Cowan, An amplitude approach to contextual effects in primary visual cortex, *Neural Comput.* **14**, 493 (2002).
- [12] P. C. Bressloff and J. D. Cowan, An $SO(3)$ Symmetry Breaking Mechanism for Orientation and Spatial Frequency Tuning in Visual Cortex, *Phys. Rev. Lett.* **88**, 078102 (2002).
- [13] P. C. Bressloff and J. D. Cowan, Spherical model of orientation and spatial frequency tuning in a cortical hypercolumn, *Philos. Trans. R. Soc. London B* **358**, 1643 (2003).
- [14] G. Faye, P. Chossat, and O. Faugeras, Analysis of a hyperbolic geometric model for visual texture perception, *J. Math. Neurosci.* **1**, 4 (2011).
- [15] C. R. Laing and C. C. Chow, Stationary bumps in networks of spiking neurons, *Neural Comput.* **13**, 1473 (2001).
- [16] Z. P. Kilpatrick and G. B. Ermentrout, Wandering bumps in stochastic neural fields, *SIAM J. Appl. Dyn. Syst.* **12**, 61 (2013).
- [17] S. Carroll, K. Josić, and Z. P. Kilpatrick, Encoding certainty in bump attractors, *J. Comp. Neurosci.* **37**, 29 (2014).
- [18] Z. P. Kilpatrick, Ghosts of bump attractors in stochastic neural fields: Bottlenecks and extinction, *Discrete Contin. Dyn. Syst. Ser. B* **21**, 2211 (2016).
- [19] Z. P. Kilpatrick, Synaptic mechanisms of interference in working memory, *Sci. Rep.* **8**, 7879 (2018).

- [20] P. C. Bressloff and M. A. Webber, Front propagation in stochastic neural fields, *SIAM J. Appl. Dyn. Syst.* **11**, 708 (2012).
- [21] O. Faugeras and J. Inglis, Stochastic neural field theory: A rigorous footing, *J. Math. Biol.* **71**, 259 (2014).
- [22] M. Kruger and W. Stannat, Front propagation in stochastic neural fields: A rigorous mathematical framework, *SIAM J. Appl. Dyn. Syst.* **13**, 1293 (2014).
- [23] J. Inglis and J. MacLaurin, A general framework for stochastic traveling waves and patterns, with application to neural field equations, *SIAM J. Appl. Dyn. Syst.* **15**, 195 (2016).
- [24] M. A. Webber and P. C. Bressloff, The effects of noise on binocular rivalry waves: A stochastic neural field model, *J. Stat. Mech.* (2013) P03001.
- [25] P. C. Bressloff, Stochastic neural field model of stimulus-dependent neural variability, *PLoS Comput. Biol.* **15**, e1006755 (2019).
- [26] J. Fiser, C. Chiu, and M. Weliky, Small modulation of ongoing cortical dynamics by sensory input during natural vision, *Nature (London)* **431**, 573 (2004).
- [27] A. Kohn and M. A. Smith, Stimulus dependence of neuronal correlation in primary visual cortex of the macaque, *J. Neurosci.* **25**, 3661 (2005).
- [28] M. M. Churchland, B. M. Yu, J. P. Cunningham, L. P. Sugrue, M. R. Cohen *et al.*, Stimulus onset quenches neural variability: A widespread cortical phenomenon, *Nat. Neurosci.* **13**, 369 (2010).
- [29] D. Deco and E. Hugues, Neural network mechanisms underlying stimulus driven variability reduction, *PLoS Comput. Biol.* **8**, e1002395 (2012).
- [30] A. Litwin-Kumar and B. Doiron, Slow dynamics and high variability in balanced cortical networks with clustered connections, *Nat. Neurosci.* **15**, 1498 (2012).
- [31] A. Ponce-Alvarez, A. Thiele, T. D. Albright, G. R. Stoner, and G. Devo, Stimulus-dependent variability and noise correlations in cortical MT neurons, *Proc. Natl. Acad. Sci. USA* **110**, 13162 (2013).
- [32] A. Kohn, R. Coen-Cagli, I. Kanitscheider, and A. Pouget, Correlations and neuronal population information, *Annu. Rev. Neurosci.* **39**, 237 (2016).
- [33] B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker, and K. Josic, The mechanics of state-dependent neural correlations, *Nat. Neurosci.* **19**, 383 (2016).
- [34] C. Huang, D. A. Ruff, P. Ryan, R. Rosenbaum, M. R. Cohen, and B. Doiron, Circuit models of low-dimensional shared variability in cortical networks, *Neuron* **101**, 337 (2019).
- [35] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 2nd ed. Wiley Series in Probability and Statistics (Wiley, Chichester, 2000).
- [36] C. W. Gardiner *Handbook of Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
- [37] M. Hübener, D. Shoham, A. Grinvald, and T. Bonhoeffer, Spatial relationships among three columnar systems in cat area 17, *J. Neurosci.* **17**, 9270 (1997).
- [38] N. P. Issa, C. Trepel, and M. P. Stryker, Spatial frequency maps in cat visual cortex, *J. Neurosci.* **20**, 8504 (2000).
- [39] I. Nauhaus, K. J. Nielsen, A. A. Disney, and E. M. Callaway, Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex, *Nat. Neurosci.* **15**, 1683 (2012).
- [40] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach, Topological analysis of population activity in visual cortex, *J. Vision* **8**, 11 (2008).
- [41] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, 2nd ed. (Oxford University Press, New York, 2005).
- [42] S. Visser, R. Nicks, and O. Faugeras, and S. Coombes, Standing and traveling waves in a spherical brain model: The Nunez model revisited, *Phys. D (Amsterdam)* **349**, 27 (2017).
- [43] P. C. Bressloff, Stochastic neural field theory of wandering bumps on a sphere, *Physica D: Nonlin. Phenom.*, doi: 10.1016/j.physd.2019.04.010 (2019).
- [44] T. Hillen, K. J. Painter, A. C. Swan, and A. D. Murtha, Moments of von Mises and Fisher distributions and applications, *Math. Biosci. Eng.* **14**, 673 (2017).
- [45] P. Reimann, R. Bartussek, and P. Hanggi, Reaction rates when barriers fluctuate: A singular perturbation approach, *Chem. Phys.* **235**, 11 (1998).
- [46] P. Reimann and P. Hanggi, in *Stochastic Dynamics*, edited by L. Schimansky-Geier and Th. Poschel, Lecture Notes in Physics Vol. 484 (Springer, Berlin, 1997), pp. 127–139.
- [47] C. R. Doering and J. C. Gadoua, Resonant Activation Over a Fluctuating Barrier, *Phys. Rev. Lett.* **69**, 2318 (1992).
- [48] P. Reimann, Thermally activated escape with potential fluctuations driven by an Ornstein-Uhlenbeck process, *Phys. Rev. E* **52**, 1579 (1995).
- [49] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas and M. Wiener, Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex, *Philos. Trans. R. Soc. London B* **356**, 299 (2001).
- [50] P. C. Bressloff and S. R. Carroll, Pattern-forming instabilities in neural fields on product spaces, *SIAM J. Appl. Dyn. Syst.* **13**, 1620 (2014).
- [51] S. R. Carroll and P. C. Bressloff, Symmetric bifurcations in a neural field model for encoding the direction of spatial contrast gradients, *SIAM J. Appl. Dyn. Syst.* **17**, 1 (2018).