Neural network model of the primary visual cortex: From functional architecture to lateral connectivity and back

Barak Blumenfeld · Dmitri Bibitchkov · Misha Tsodyks

Abstract The role of intrinsic cortical dynamics is a debatable issue. A recent optical imaging study (Kenet et al., 2003) found that activity patterns similar to orientation maps (OMs), emerge in the primary visual cortex (V1) even in the absence of sensory input, suggesting an intrinsic mechanism of OM activation. To better understand these results and shed light on the intrinsic V1 processing, we suggest a neural network model in which OMs are encoded by the intrinsic lateral connections. The proposed connectivity pattern depends on the preferred orientation and, unlike previous models, on the degree of orientation selectivity of the interconnected neurons. We prove that the network has a ring attractor composed of an approximated version of the OMs. Consequently, OMs emerge spontaneously when the network is presented with an unstructured noisy input. Simulations show that the model can be applied to experimental data and generate realistic OMs. We study a variation of the model with spatially restricted connections, and show that it gives rise to states composed of several OMs. We hypothesize that these states can represent local properties of the visual scene.

Keywords Visual cortex · Spontaneous activity · Orientation selectivity · Neural network model · Optical imaging

Action Editor: Jonathan D. Victor

B. Blumenfeld · D. Bibitchkov · M. Tsodyks (✉)
Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel
e-mail: misha@weizmann.ac.il

1. Introduction

Processing of visual information in the mammalian primary visual cortex (V1) is thought to arise from an interplay between the pattern of external projections and intrinsic cortical dynamics. Even though the intrinsic processing has been a subject of an extensive research, its role and the mechanisms underlying it remain poorly understood. A powerful tool for understanding the intrinsic processing is the study of spontaneous cortical activity, that is, activity not evoked by visual stimulation. Even in the absence of visual stimulation, V1 exhibits a rich and complex spontaneous activity (Lampl et al., 1999; Tsodyks et al., 1999; Kenet et al., 2003; Fiser et al., 2004). This activity is thought to originate from mechanisms intrinsic to V1. Thus, spontaneous activity can reveal the underlying V1 architecture and shed light on its function.

In this study, we focus on the link between spontaneous activity and the activity evoked by visual stimulation. Since the work of Hubel and Wiesel (1959) it is believed that one of the main functions of V1 is to encode and process oriented stimuli. A major progress in studying the functional architecture of V1 was achieved by optical imaging (Blasdel and Salama, 1986; Grinvald et al., 1986). In a widely used experimental protocol (Grinvald et al., 1999), an animal is shown full field moving gratings of different orientations. Each orientation yields a single condition *orientation map* (OM), a two-dimensional representation of the neuronal activity across the cortical sheet evoked by the stimulus. In cats, as well in many other mammals (including primates, ferrets, tree shrews and sheep, but not rodents; see Van Hooser et al., 2005) neurons with similar orientation preference tend to be clustered. Thus, OMs are characterized by patches of high

neuronal activity separated by regions of low neuronal activity. The location and shape of the patches change with the gratings' orientation.

A recent work (Kenet et al., 2003) used optical imaging with voltage sensitive dye to study spontaneous activity in V1 of an anaesthetized cat, measured when the animal's eyes were closed. Even though no visual stimulus was presented, some of the spontaneous cortical states were similar to the OMs of the same region. When such spontaneous OMs emerged, they spanned several hypercolumns, and were often followed by states similar to OMs of a proximal orientation. It was speculated that the spontaneous activity reflects the internal state of the brain, which might be influenced by context, attention, or perceptual memory. For example, the spontaneous state can reflect an expectation of the orientation of a forthcoming input, based on prior activity. In any case, the striking similarity between spontaneous and evoked states suggests that there exists a mechanism intrinsic to V1 capable of generating the OMs. If such mechanism exists, the evoked OMs originate from an interplay between the intrinsic mechanism and the pattern of afferent LGN input, known to have some degree of orientation tuning.

What is the intrinsic V1 mechanism underlying the spontaneous emergence of OMs? Many experimental studies suggest an important role for the pattern of lateral connectivity in V1. A large body of anatomical and electrophysiological research has shown that the V1 intracortical connectivity is correlated with orientation preference. The general trend is that long range lateral projections connect neurons with similar orientation preference (Gilbert and Wiesel, 1989; Malach et al., 1993; Weliky et al., 1995; Bosking et al., 1997; Kisvárday et al., 1997; Buzás et al., 1998). As the data is highly variable, this trend is true only on average, and the exact relation between intracortical connectivity, orientation preference, and other functional maps remains a subject of active research (Kisvárday et al., 2002; Ben-Shahar and Zucker, 2004). Direct evidence for the involvement of the lateral connections in orientation tuning comes from studies in which inactivation of GABAergic neurons altered the orientation tuning curves of remotely located neurons (Crook et al., 1998; Kisvárday et al., 2000). Developmental studies in newborn animals have shown that the overall geometry of the OMs is stable from a very early stage, and that it is resistant to manipulations of the visual input (Crair et al., 1998; Löwel et al., 1998). On the other hand, cortical development involves sharpening of the OMs with a time course that closely matches the time course for expression and refinement of long range lateral connections (Gödecke et al., 1997). Developmental experiments in which the visual input was rerouted into the *auditory* cortex (Sharma et al., 2000), yielded an auditory cortex that was similar, in many aspects, to V1. In particular, it exhibited both OMs and long range lateral projections that preferentially connected cells with
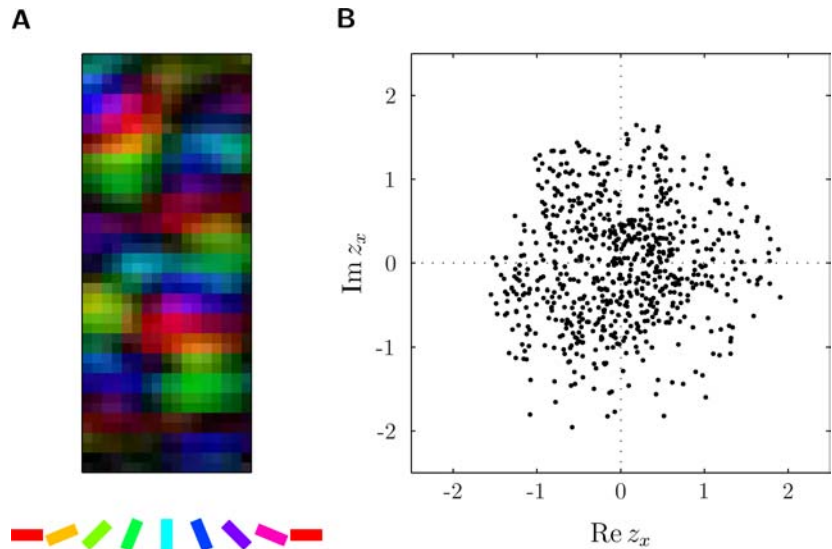
similar orientation preference. Taken together, these developmental studies provide further support for the strong link between OMs and lateral connections.

From a theoretical point of view, V1 can be viewed as a complex, nonlinear, dynamical system. Several theoretical studies suggested that OMs are *attractor* states of the cortical dynamics (Somers et al., 1995; Sompolinsky and Shapley, 1997; Ernst et al., 2001). This line of thought is consistent with the spontaneous emergence of orientation maps because OMs can still be attractors of the intracortical dynamics even without the stimulus-encoding afferent input. In this family of models the pattern of lateral connectivity plays an important role in the formation of the attractor landscape, consistent with the experimental findings listed above. A well studied example of such a model is the *ring model* (Ben-Yishai et al., 1995; Hansel and Sompolinsky, 1998). This model considers a population of neurons in one hypercolumn in V1. Each neuron is characterized by a preferred orientation, and the overall distribution of preferred orientations is uniform over the entire range of possible orientations. The neurons receive both afferent LGN inputs and lateral recurrent inputs. When a stimulus is presented to the network via the afferent connections, the network develops an activity profile that peaks at the neuron whose preferred orientation matches the stimulus' orientation. However, the precise shape of the stationary profile is determined, to a large extent, by recurrent connections. The ring model therefore suggests a mechanism that links the cortical response evoked by an oriented stimulus and the pattern of lateral connections.

Moreover, if the recurrent connections are strong enough, states similar to those evoked by a stimulus emerge even if a uniform input (representing lack of an external visual stimulus) is applied. This occurs because the uniform solution becomes unstable, and a continuum of states that are similar to the states evoked by the different stimulus orientations emerges as an attractor of the network's dynamics. Collectively, these states are known as the *ring attractor*. Thus, the ring model suggests a mechanism that explains how states similar to OMs can emerge spontaneously. This mechanism was suggested to underlie the patterns of spontaneous activity observed in Kenet et al. (2003).

The ring model, however, cannot directly explain the generation of experimental OMs. It is inherently a one-dimensional model, proposed as a model for one hypercolumn, and cannot account for the generation of the 2-dimensional OMs. One can of course consider a straightforward 2-dimensional implementation of the ring model, by simply placing the neurons on a 2-dimensional sheet (Goldberg et al., 2004); however, the attractor states of such implementation do not necessarily match the experimental OMs. We will further discuss the limitations of the ring model in Section 2.

Fig. 1 (A) The polar map (PM) for the data obtained in Kenet et al. (2003). The preferred orientation at each pixel is encoded by its color according to the scheme at the bottom of the figure. The selectivity is encoded by brightness (a dark pixel represents low selectivity, a bright pixel represents high selectivity; The brightness scale is linearly stretched between zero selectivity and the maximal selectivity in this PM) (B) Distribution over the complex plane of the values of z_x of the PM in (A)

This leaves the following question open: what type of lateral connectivity can lead to the spontaneous emergence of the OMs? In this contribution we answer this question by suggesting a neural network model with a simple connectivity rule that supports a ring attractor; this ring attractor is composed of states similar to the experimental OMs. To construct this network we first consider an approximation to the OMs (Section 2). Then, we introduce the model and solve it analytically (Section 3). Later, we take an experimental data set, and apply to it the theory we developed in the previous section (Section 4). We continue by studying the effect of the limited spatial extent of the lateral connections (Section 5), and conclude with a discussion of our results and predictions of the models (Section 6).

2. The polar map

As a first step toward constructing the network, we will consider an approximation to the OMs that is based on the "polar map" (PM; Bonhoeffer and Grinvald, 1993). The PM is a functional map that assigns each location x, a complex number, z_x. The values of z_x are typically calculated from the OMs as follows. Let $\varphi_1, \varphi_2, \ldots, \varphi_p$ be p equidistant orientations presented to the animal.[1] Let $S_x^{\varphi_j}$ denote the response at location x for the OM evoked by orientation φ_j. Then, z_x is defined by:

---

[1] For the theoretical analysis we let orientations take values over the interval $[0, 2\pi)$ rather than the natural encoding over the interval $[0, \pi)$. This encoding induces a 2π periodicity that simplifies the equations we present. Thus, φ_j represents a stimulus whose actual orientation was $\varphi_j/2$.

$$z_x \equiv r_x e^{i\theta_x} = \frac{2}{p} \sum_{j=1}^{p} S_x^{\varphi_j} r^{i\varphi_j} \qquad (1)$$

where $i = \sqrt{-1}$. Equation (1) can be viewed as a summation of p 2-dimensional vectors whose angle, φ_j, represents the orientation of the stimulus, and their length, $S_x^{\varphi_j}$ represents the magnitude of the cortical response to that orientation. This view gives rise to the standard interpretation of the angle of z_x, i.e. θ_x, as the *preferred orientation* at location x. The magnitude of z_x, i.e. r_x, measures the degree to which the response at location x is modulated by the stimulus' orientation. Following other authors, we will refer to this variable as the *selectivity* of location x.

Throughout the article we will use the data set obtained in Kenet et al. (2003) to demonstrate and validate the theory we develop. This data set contained 8 OMs obtained from an anaesthetized cat using voltage-sensitive dye optical imaging. After a two-fold downsampling of the original data set, each map contained $42 \times 17 = 714$ pixels, corresponding to an area of approximately $11.7\,\text{mm}^2$ (the downsampling was preformed so that we can run efficient simulations in the later stages of the analysis). The PM calculated from the 8 OMs using Eq. (1) is depicted in Fig. 1(A). The color coding represents the above polar interpretation of z_x. The color of each pixel represents the preferred orientation, θ_x at that location, and the brightness represents the selectivity, r_x. Pinwheels are represented as dark pixels, i.e. pixels where r_x is close to zero, around which all preferred orientations are represented. A closer examination reveals that when moving from a point far from the pinwheel ("linear zone") towards the pinwheel, the selectivity is gradually reduced, resulting in a large range of selectivity values. The variability of the selectivities can

be also observed in Fig. 1(B), where distribution of the values of z_x over the complex plane is depicted. In this figure each pixel in the PM is represented by a single point. The distance of each point from the origin represents the selectivity of the pixel, and the angle it makes with the real axis represents its preferred orientation. It can be seen that for each preferred orientation there is a wide range of selectivities. This variability is not captured by previous models such as the ring model, which implicitly assume a constant selectivity. As a result, these models are limited in their ability to account for realistic OMs.

Equation (1) can also be viewed as the (discrete) Fourier transform of the "tuning curve" of location x, where the term "tuning curve" here refers to the optical imaging response, $S_x^{\varphi_j}$, as a function φ_j. Thus, z_x is simply the first Fourier coefficient of the tuning curve of location x. Since these tuning curves tend to be unimodal and centered around the pixel's preferred orientation, we can expect the first Fourier component to contain much of the tuning curve's energy. Therefore, the tuning curve of each location can be approximated with this single Fourier component. This approximation can be written explicitly by defining the *approximated OM* for orientation φ as:

$$M_x^{\varphi} = r_x \cos(\theta_x - \varphi) \qquad (2)$$

and $S_x^{\varphi_j} \approx M_x^{\varphi_j}$, based on the arguments presented above. We suggest two ways of measuring the quality of this approximation and apply them to the Kenet et al. (2003) data set. First, the similarity between each of the experimental OMs and its approximated version can be we quantified. A standard measure of similarity between OMs is their Pearson correlation coefficient. For the experimental data set, the average correlation between the 8 pairs of experimental and approximated OMs was 0.815 (min: 0.746, max: 0.887). These are high correlations that validate the approximation. Another measure for the quality of the approximation is the percent of variance explained, which we denote by γ. With r_x defined by Eq. (1), γ can be defined as:

$$\gamma = \frac{\sum_x r_x^2}{2 \sum_x \sigma_x^2} \qquad (3)$$

where σ_x^2 is the variance of the cortical response over all orientations at location x. The values that γ takes are between 0 and 1, with $\gamma = 1$ if and only if the approximation is precise. For the experimental data set, $\gamma = 0.89$, again validating the usage of the approximation. We obtained similar results for other data sets as well.

From the above analysis we conclude that experimental OMs can be well-approximated using the PM. For the purpose of this contribution, that is, constructing a neural

network with OMs as being its attractor states, this approximation offers two advantages. First, by associating each location with only two real numbers (r_x, θ_x), rather than the entire tuning curve, the dimensionality of the problem is greatly reduced. Second, the approximation suggests a natural way to interpolate OMs for orientations other than the ones presented during the experiment. We can thus consider a ring of approximated OMs, formally defined by assigning arbitrary values to φ in Eq. (2). Our modeling goal can be now reformulated as constructing a model whose attractor states will coincide with this ring of approximated OMs. This goal is achieved by the model we introduce in the next section.

3. Model

Let $m_x(t)$ denote the average firing rate of the cortical minicolumn at location x at time t. Since OMs tend to be smooth functions of space, we use a continuous approximation and let x take real numbers rather than discrete values. The evolution of m_x is described by the standard rate equation (Wilson and Cowan, 1973):

$$\tau \dot{m}_x = -m_x + \left[I_x^{\text{rec}} + I_x^{\text{aff}} - T \right]_+ \qquad (4)$$

where $\dot{m}_x$ is the time derivative of m_x, I_x^{rec} is the input to location x due to the recurrent connections, I_x^{aff} is the afferent input to that location, τ is a time constant, T is the firing threshold, $[\ ]_+$ denotes the ramp gain function ($[\xi]_+ = \xi$ if $\xi > 0$ and $[\xi]_+ = 0$ otherwise), and the time argument has been suppressed for brevity. The recurrent input, I_x^{rec}, is defined by:

$$I_x^{\text{rec}} = \frac{1}{A} \int dy W_{xy} m_y \qquad (5)$$

The integral is taken over the whole region of V1 that is modeled, and A is the area of that region. A is used as a normalization factor that simplifies the mathematical formulations. The synaptic weights connecting locations x and y, W_{xy}, are defined by:

$$W_{xy} = J_2 r_x r_y \cos(\theta_x - \theta_y) + J_0 \qquad (6)$$

The information about the PM is given by the term $r_x r_y \cos(\theta_x - \theta_y)$. The parameter $J_2 > 0$, is a global scaling factor of this term. Since scaling the values of r is equivalent to scaling J_2, we assume that the values of r are normalized so that the mean of r_x^2 over the area is 1. The parameter J_0 represents global excitation (if $J_0 > 0$) or global inhibition (if $J_0 < 0$). It should be noted that the connectivity given by Eq. (6) is similar to the connectiv-

ity of the ring model (Ben-Yishai et al., 1995). However, unlike in the ring model, the connectivity we suggest depends not only on the preferred orientations of the pre- and post- synaptic locations, but also on their selectivities.

The connectivity pattern induced by Eq. (6) is illustrated in Fig. 2(A). The connectivity matrix was calculated using the PM of the Kenet et al. (2003), and setting $J_0 = -2$, $J_2 = 5$, as used in subsequent simulations. The connectivity strength between one specific location (black dot) and all other locations is visualized by color coding, with shades of green representing positive weights ("excitatory" connections), shades of red representing negative weights ("inhibitory" connections), and white representing connections close to zero. Note the patchy structure of the connectivity pattern, which follows directly from the patchy nature of the PM. Moreover, the "excitatory patches" correspond to regions with a similar orientation preference; Thus, the patchy pattern in Fig. 2(A) reflects the patchy layout of long range connections in V1 found in many anatomical studies (Gilbert and Wiesel, 1989; Malach et al., 1993; Bosking et al., 1997; Kisvárday et al., 1997). However, in contrast with known anatomy, the connectivity does not depend on the distance between the pre- and post-synaptic locations. This is a simplification we make at this point to allow an analytical investigation. Later, in Section 5, we discuss a model with spatially restricted connections and show that it is similar, in many aspects, to the model we consider here. The spatially restricted connections are illustrated in Fig. 2(B), with the corresponding parameters used in subsequent simulations.

We next consider the system given by Eqs. (4)–(6) for two types of afferent input: one corresponding to spontaneous activity (Section 3.1), and the other representing the afferent input evoked by the gratings stimulus (Section 3.2).

3.1. Spontaneous activity

To model spontaneous activity, we take the afferent input to be constant for all locations, i.e., $I_x^{aff} = C$. We also assume $C > T$, otherwise all locations are below threshold, and the network does not develop any activity.

3.1.1. Reduced dynamics

We start the analysis of the model by reducing the dynamics given in (4) to the dynamics of two order parameters. To this end, we observe that the synaptic weights defined by Eq. (6) can be written as $J_2 \, \text{Re}(\bar{z}_x z_y) + J_0$, with z_x, z_y defined by Eq. (1). Thus, Eq. (5) can rewritten as:

$$I_x^{rec} = J_2 \text{Re}\left(\bar{z}_x \frac{1}{A} \int dy \, z_y m_y\right) + J_0 \frac{1}{A} \int dy \, m_y \qquad (7)$$
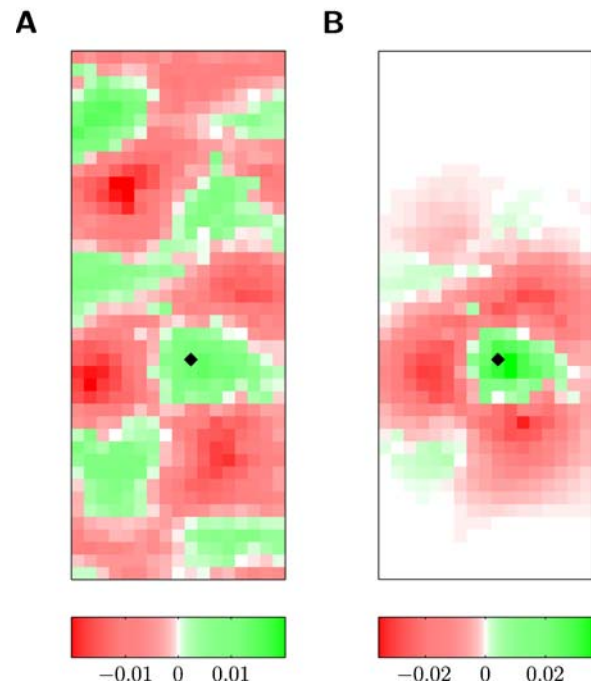

Fig. 2 Connectivity pattern of the model. (A) Connection strength between one location, marked by the black dot, and all other locations. The connection strength was calculated by Eq. (6), using the data of Kenet et al. (2003), with $J_0 = -2$, $J_2 = 5$. (B) Connection strength for the model with spatially restricted connections. The connection strength was calculated by Eq. (48), using the same data set, with $J_0 = 3$, $J_2 = 3.5$, $\sigma = 0.6 \, \text{mm}$

This form leads to the following definitions of the order parameters μ and Z:

$$\mu = \frac{1}{A} \int dx \, m_x \qquad (8)$$

$$Z \equiv \rho e^{i\psi} = \frac{1}{A} \int dx \, r_x e^{i\theta_x} m_x \qquad (9)$$

These variables represent global properties of the network activity: μ is the average firing rate and Z is the inner product of the firing rates with the PM. Equation (7) can be now rewritten as:

$$I_x^{rec} = J_2 \text{Re}(\bar{z}_x Z) + J_0 \mu = J_2 \rho r_x \cos(\theta_x - \psi) + J_0 \mu \qquad (10)$$

To obtain equations for the evolution of μ and Z, we differentiate Eqs. (8) and (9) with respect to time, and then express $\dot{m}_x$ as a function of μ and Z using Eqs. (4) and (10). With these substitutions, we replace the spatial integration, i.e., the integration over y, with an integration over r and θ. As a

result, the following equations are obtained:

$$\tau \dot{\mu} = -\mu + \int \int dr d\theta P(r,\theta)[J_2 \rho r(\theta - \psi)$$
$$+ J_0 \mu + C - T]_+ \qquad (11)$$

$$\tau \dot{Z} = -Z + \int \int dr d\theta P(r,\theta) r e^{i\theta}[J_2 \rho r \cos(\theta - \psi)$$
$$+ J_0 \mu + C - T]_+ \qquad (12)$$

where the integrals are taken over all possible values of r and θ (i.e., $r \in [0, \infty], \theta \in [-\pi, \pi]$) and $P(r, \theta)$ is the joint probability density function of r and θ. Equations (11) and (12) are closed equations for the dynamics of μ and Z that do not depend directly on the particular values of m_x. However, in general, these equations still depend on the angle of Z, i.e., ψ. Nevertheless, if isotropy is assumed, that is, the distribution of θ is uniform and independent of the distribution of r, this dependency upon ψ is lost. The validity of these assumptions is discussed later, in Section 4. In this case, we can rewrite Eqs. (11) and (12) as:

$$\tau \dot{\mu} = -\mu + \int \int dr d\theta \frac{P(r)}{2\pi}[J_2 \rho r \cos\theta + J_0 \mu + C - T]_+ \qquad (13)$$

$$\tau \dot{\rho} = -\rho + \int \int dr d\theta \frac{P(r)}{2\pi} r \cos\theta [J_2 \rho r \cos\theta + J_0 \mu$$
$$+ C - T]_+ \qquad (14)$$

$$\tau \rho \dot{\psi} = 0 \qquad (15)$$

As usual, to find the fixed points for the dynamics, we set the time derivative in Eqs. (13) – (15) to zero. We discuss uniform (with $\rho = 0$), and non-uniform (with $\rho > 0$) solutions separately.

3.1.2. Uniform solutions

In the case of a uniform solution, Eq. (14) implies that the time derivative of ρ is zero for any μ. Equation (13) implies $\mu = [J_0 \mu + C - T]_+$ which can be solved for any $J_0 < 1$. The solution for the uniform fixed point is therefore:

$$\mu = \frac{C - T}{1 - J_0} \qquad (16)$$

$$\rho = 0 \qquad (17)$$

The stability analysis is straightforward, as all neurons are above threshold, and the system near the fixed point is linear.

We find that the fixed point is stable in the region $J_0 < 1$ and $J_2 < 2$.

3.1.3. Non-uniform solutions

To find the non-uniform solutions, we pull ρ out of the integrals in Eqs. (13) and (14) and rewrite them as:

$$\tau \dot{\mu} = -\mu + J_2 \rho \mathcal{F}_0(X) \qquad (18)$$

$$\tau \dot{\rho} = -\rho + J_2 \rho \mathcal{F}_2(X) \qquad (19)$$

with,

$$X = \frac{J_0 \mu + C - T}{J_2 \rho} \qquad (20)$$

and,

$$\mathcal{F}_0(X) = \int \int dr d\theta \frac{P(r)}{2\pi}[r \cos\theta + X]_+ \qquad (21)$$

$$\mathcal{F}_2(X) = \int \int dr d\theta \frac{P(r)}{2\pi} r \cos\theta [r \cos\theta + X]_+ \qquad (22)$$

In order to find the fixed point solutions for μ and ρ using Eqs. (18) and (19), it is necessary to know the functions $\mathcal{F}_0$ and $\mathcal{F}_2$ that in turn depend on the concrete choice of the selectivities' distribution $P(r)$ (Eqs. (21) and (22)). To study the properties of these functions, we first plot them for 3 different choices of $P(r)$ (Fig. 3). In these examples, we see that the precise values that $\mathcal{F}_0$ and $\mathcal{F}_2$ take indeed depend on $P(r)$. However, in all 3 cases, $\mathcal{F}_0$ is a non-decreasing function of X, approaching zero for large negative X, and increasing with a slope of 1 for large positive X. Similarly, in all 3 cases, $\mathcal{F}_2$ is a sigmoid-shaped function that saturates at 0 and 1/2. In fact, these observations about $\mathcal{F}_0$ and $\mathcal{F}_2$ are true in the general case. Formally, we show a set of properties of $\mathcal{F}_0$ and $\mathcal{F}_2$ that hold for any $P(r)$. These properties are listed in Table 1, and their proof is given in the appendix. As we will show, they allow us to obtain solutions for the fixed point values of μ and ρ for an arbitrary $P(r)$.

We first find the fixed point value of X by setting the time derivative in Eq. (19) to zero. We find that the fixed point value of X is defined by:

$$1 - J_2 \mathcal{F}_2(X) = 0 \qquad (23)$$

Property (vi) in Table 1 directly implies that when $J_2 < 2$, Eq. (23) cannot be satisfied, therefore a non-uniform solution does not exist (see also Fig. 4(C)). On the other hand, properties (v), (vi), (vii) imply that when $J_2 > 2$ Eq. (23) has

Fig. 3 The functions $\mathcal{F}_0$ and $\mathcal{F}_2$. (A) $\mathcal{F}_0$ plotted for 3 choices of P(r); Dashed: constant r, P(r) = $\delta(r-1)$. Dash-dotted: exponential distribution, P(r) = $\sqrt{2}e^{-\sqrt{2r}}$. Solid: P(r) according to the isotropy-adjusted experimental PM (see Section 4 and Fig. 7 for details). (B) $\mathcal{F}_2$ plotted for the same distributions of r
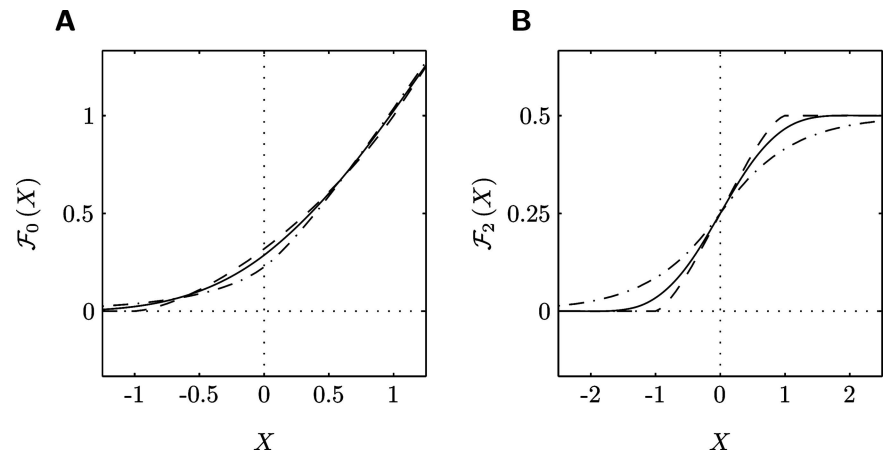
Table 1

No	Property
(i)[a]	$\mathcal{F}_0(X) \geq 0$
(ii)[b]	$\mathcal{F}_0(X) \geq X$
(iii)[a]	$\mathcal{F}_0'(X) \geq 0$
(iv)[b]	$\mathcal{F}_0'(X) \leq 1$
(v)[a]	$\mathcal{F}_2(X) \geq 0$
(vi)[b]	$\mathcal{F}_2(X) \leq \frac{1}{2}$
(vii)[c]	$\mathcal{F}_2'(X) \geq 0$
(viii)	$\mathcal{F}_0(X) - X\mathcal{F}_0'(X) = \mathcal{F}_2(X)$
(ix)[b]	$\mathcal{F}_2(X) - X\mathcal{F}_2'(X) \leq \frac{1}{2}$
(x)[a]	$\mathcal{F}_0'(X)\mathcal{F}_2(X) - \mathcal{F}_0(X)\mathcal{F}_2'(X) \geq 0$

The inequalities hold with equality if and only if:

[a] r is bounded and $X \leq -r_{\max}$ or r is unbounded and $X \to -\infty$

[b] r is bounded and $X \geq r_{\max}$ or r is unbounded and $X \to \infty$

[c] r is bounded and $|X| \geq r_{\max}$ or r is unbounded and $|X| \to \infty$

a unique solution (Fig. 4(D)). Therefore, when $J_2 > 2$ there is a unique solution for the fixed point value of X. Assuming this solution is known, the following expressions for the fixed point value of μ and ρ in terms of the fixed point value of X can be derived form Eqs. (18), (20), and (23):

$$\mu = \frac{C - T}{X - J_0\mathcal{F}_0(X)}\mathcal{F}_0(X) \qquad (24)$$

$$\rho = \frac{C - T}{X - J_0\mathcal{F}_0(X)}\mathcal{F}_2(X) \qquad (25)$$

Since μ and ρ are by definition positive, it must hold that:

$$X - J_0\mathcal{F}_0(X) > 0 \qquad (26)$$

otherwise, the solution diverges. Again, we can use the properties in Table 1 to find the conditions under which this inequality is satisfied. Property (ii) implies that inequality

(26) never holds when $J_0 > 1$ (Fig. 4(A)). On the other hand, if $J_0 < 1$, $X - J_0\mathcal{F}_0(X)$ is a monotonically increasing function of X (properties (iii), (iv)), taking negative values for large negative values of X, and positive values for large positive values of X (properties (i), (ii)). Therefore, when $J_0 < 1$, $X - J_0\mathcal{F}_0(X) = 0$ has a unique solution. We denote this solution by X_0 (Fig. 4(B)), and the solution of Eq. (23) by X_2 (Fig. 4(D)). Using these notations, we find that inequality (26) holds for $X = X_2$ if and only if $X_0 < X_2$. Since X_0 is a function of J_0 and X_2 is a function of J_2, the condition $X_0 < X_2$, together with the conditions $J_0 < 1$, $J_2 > 2$, define a region in the J_0J_2 plane, where the non-uniform fixed point exists. Stability analysis given in the appendix shows that when this fixed point exists, it is always stable.
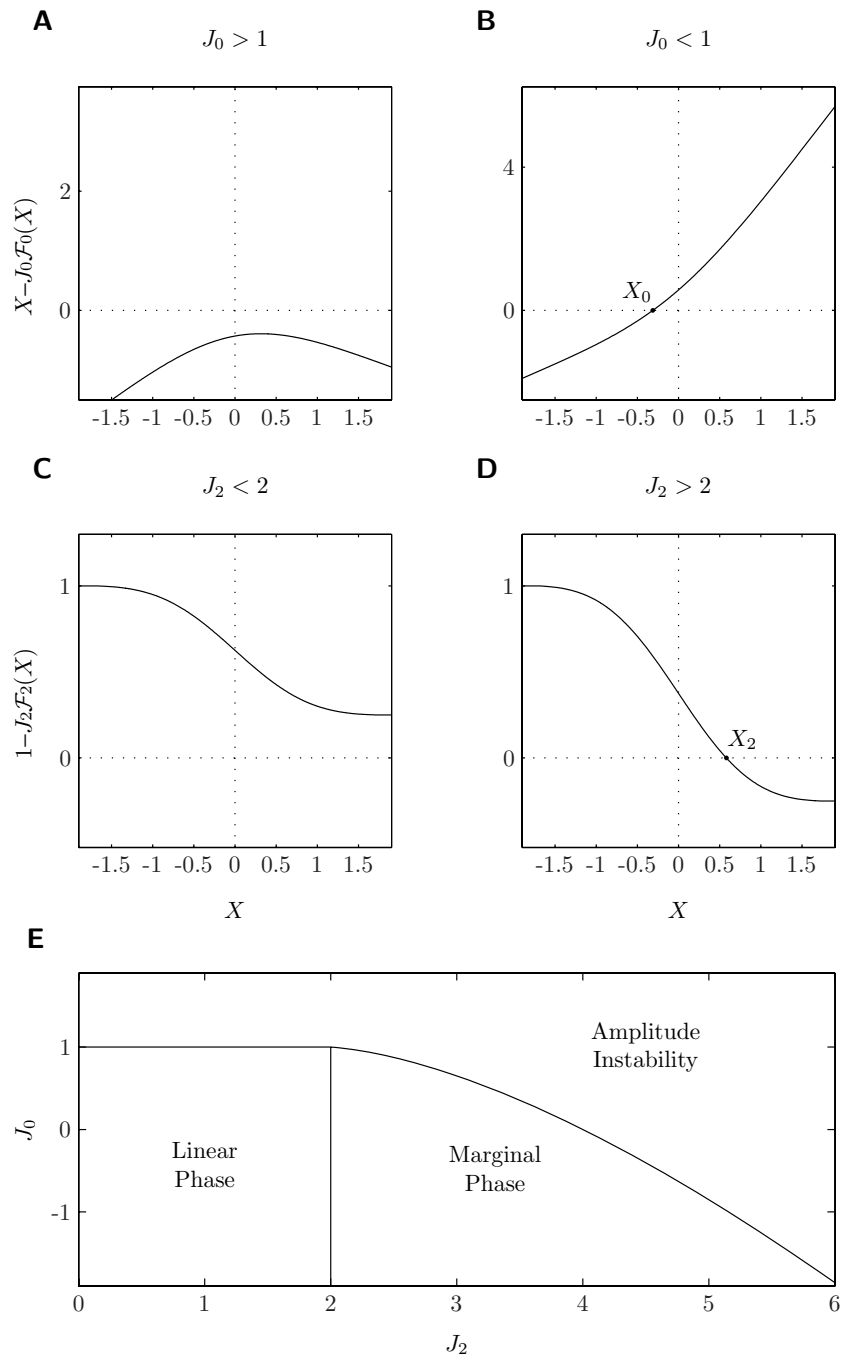
3.1.4. Phase diagram

We turn to summarize the different regimes of the model and relate the fixed points in these regimes to the OMs. Since the voltage-sensitive dye optical imaging signal corresponds to the instantaneous post synaptic membrane potential rather than firing rate (Sterkin et al., 1999; Grinvald et al., 1999; Sharon and Grinvald, 2002; Petersen et al., 2003), we compare OMs to the total synaptic input, denoted by $I^{\text{tot}} = I^{\text{rec}} + I^{\text{aff}}$. The shape of I^{tot} is given by (see Eq. (10)):

$$I_x^{\text{tot}} = J_2\rho r_x \cos(\theta_x - \psi) + J_0\mu + C \qquad (27)$$

with μ, ρ, and ψ taking their fixed point values. The different phases of the model are summarized in the phase diagram (Fig. 4(E)). In the *linear phase*, $J_0 < 1$, $J_2 < 2$, the only solution is $\rho = 0$, and it is stable. In this case, Eq. (27) implies that I_x^{tot} is uniform and does not correspond to any OM. However, in the *marginal phase*, $J_0 < 1$, $J_2 > 2$, $X_0 < X_2$ (where X_0 is a function of J_0 and X_2 is a function of J_2), the

Fig. 4 Solution and phase diagram of the spontaneous activity model. (A, B) The function $X - J_0\mathcal{F}_0(X)$ (see Eq. (26)). In (A), $J_0 = 1.5$, representing the case where $J_0 > 1$. The function $X - J_0\mathcal{F}_0(X)$ is negative for all X. In (B), $J_0 = -2$, representing $J_0 < 1$. The function $X - J_0\mathcal{F}_0(X)$ crosses zero at a unique point, which is denoted by X_0. (C, D) The function $1 - J_2\mathcal{F}_2(X)$ (see Eq. (23)). In (C) $J_2 = 1.5$, representing the case where $J_2 < 2$. The function $1 - J_2\mathcal{F}_2(X)$ is positive for all X. In (D) $J_2 = 2.5$ representing $J_2 > 2$. The function $1 - J_2\mathcal{F}_2(X)$ crosses zero at a single point, which is denoted by X_2. (E) Phase diagram for the spontaneous activity. The border of the linear phase is given by the lines $J_0 = 1$, $J_2 = 2$. The line separating the marginal phase from the amplitude instability is determined by the condition $X_0 = X_2$. In all examples, $\mathcal{F}_0$ and $\mathcal{F}_2$ were calculated using the experimental P(r) (solid lines in Fig. 3)

solution with $\rho = 0$ is unstable, but solutions with $\rho > 0$ are stable. In fact, there are infinitely many solutions, because the value of ψ is arbitrary (Eq. (15)). By comparing Eq. (27) to Eq. (2), we see that in this case I_x^{tot} equals a scaled version of the approximated OM of orientation ψ, up to an additive constant shift. Thus, in the marginal phase, OMs of arbitrary orientations emerge as fixed points of the network dynamics. Collectively, these fixed points form a ring attractor, that

is, a ring of states where each state is an attractor of the dynamics.

Outside the regions of linear phase and marginal phase no stable fixed point exists and the network develops amplitude instability. Unlike the boundaries of the linear phase, the line $X_0 = X_2$ that separates the marginal phase from the amplitude instability region depends on P(r). However, some of its properties do not depend on P(r). In particular

it can be checked that it always passes through the points $(J_0 = 1, J_2 = 2)$, and $(J_0 = 0, J_2 = 4)$.

3.2. Evoked activity

The case of a non-uniform input is more complex than the uniform input case, because the steady state solutions are sensitive to the fine structure of the input. However, some general observations about the fixed point can still be made. Since the synaptic input I^{rec} is still given by Eq. (10), it equals a scaled/shifted version of one of the approximated OMs, and the total synaptic input is the sum of that map, and the afferent input, I^{aff}. The orientation of the approximated OM, and its scaling and translation parameters are still determined by ψ, ρ, and μ. Thus, the role of I^{aff} reduces to selecting the values of μ, ρ, and ψ, which will put the system in a steady state.

In this section we study a simple structure of the input; we add to the uniform input we studied in the previous section a small modulation that is proportional to one of the approximated OM. Formally we define:

$$I_x^{aff} = C(1 + \epsilon r_x \cos(\theta_x - \psi_{aff})) \tag{28}$$

where ψ_{aff} is the orientation encoded in the input and ε is the modulation of the orientation encoding term. With this form, each location receives maximal afferent input when the orientation of the stimulus coincides with its preferred orientation. The main motivation for choosing this particular form is that it allows a relatively simple analytical study. Using simulations, we also studied different variations of the above input, and obtained essentially the same results. An example of such variation, is given in Section 4.

3.2.1. Reduced dynamics

Rewriting the right-hand side of Eq. (28) as $C + \text{Re}(\bar{z}_x Z_{aff})$ with $Z_{aff} = C\epsilon e^{i\psi_{aff}}$, we find that:

$$I_x^{rec} + I_x^{aff} = \text{Re}(\bar{z}_x(J_2 Z + Z_{aff})) + J_0\mu + C \tag{29}$$

where μ and Z are defined by the same equations given for the spontaneous activity (Eqs. (8) and (9)). Equation (29) leads to the following definition:

$$Z_{tot} \equiv \rho_{tot} e^{i\psi_{tot}} = J_2 Z + Z_{aff} \tag{30}$$

and with these notations:

$$\tau \dot{m}_x = -m_x + [\rho_{tot} r_x \cos(\theta_x - \psi_{tot}) + J_0\mu + C - T]_+ \tag{31}$$

As in the case of spontaneous activity, Eq. (31) can be used to derive dynamical equations for μ and Z:

$$\tau \dot{Z} = -Z + e^{i\psi_{tot}} \int\int dr d\theta \frac{P(r)}{2\pi} r e^{i(\theta - \psi_{tot})}$$
$$[\rho_{tot} r \cos(\theta - \psi_{tot}) + J_0\mu + C - T]_+ \tag{32}$$

$$\tau \dot{\mu} = -\mu + \int\int dr d\theta \frac{P(r)}{2\pi}[\rho_{tot} r \cos(\theta - \psi_{tot}) + J_0\mu$$
$$+ C - T]_+ \tag{33}$$

This can be rewritten as:

$$\tau \dot{\mu} = -\mu + \rho_{tot} \mathcal{F}_0(X) \tag{34}$$

$$\tau \dot{\rho} = -\rho + \rho_{tot} \cos(\psi_{tot} - \psi) \mathcal{F}_2(X) \tag{35}$$

$$\tau \rho \dot{\psi} = \rho_{tot} \sin(\psi_{tot} - \psi) \mathcal{F}_2(X) \tag{36}$$

with,

$$X = \frac{J_0\mu + C - T}{\rho_{tot}} \tag{37}$$

In this formulation, it is assumed that $\rho_{tot} > 0$. However it can be easily checked that $\rho_{tot} = 0$ is not a fixed point solution, therefore the case of the uniform solution can be disregarded.

3.2.2. Solution

At fixed point, Eq. (36) implies that $\sin(\psi_{tot} - \psi) = 0$. Therefore, it must hold that either $\psi_{tot} = \psi$ or $\psi_{tot} = \psi + \pi$. However, in the second case, Eq. (35) implies $\rho = -\rho_{tot}\mathcal{F}_2(X)$, which cannot hold because both ρ and ρ_{tot} are positive, and $\mathcal{F}_2(X)$ is non-negative (property (v) in Table 1). Thus, as follows from Eq. (30), the possible solutions are $\psi_{tot} = \psi = \psi_{aff}$, and $\psi_{tot} = \psi = \psi_{aff} + \pi$. We focus on the first case, which corresponds, as the analysis in the appendix shows, to the only stable fixed point. In this case, $\rho_{tot} = J_2\rho + C\epsilon$. Setting $\dot{\mu} = 0$, $\dot{\rho} = 0$, we rewrite Eqs. (34), (35), (37) as:

$$\mu = (J_2\rho + C\epsilon)\mathcal{F}_0(X) \tag{38}$$

$$\rho = (J_2\rho + C\epsilon)\mathcal{F}_2(X) \tag{39}$$

$$X = \frac{J_0\mu + C - T}{(J_2\rho + C\epsilon)} \tag{40}$$

Thus, a fixed point requires that this system of 3 equations for the 3 variables μ, ρ, and X has a solution with positive μ

and ρ. Solving this system, we find that the fixed point value of X is determined by:

$$\frac{1 - J_2 \mathcal{F}_2(X)}{X - J_0 \mathcal{F}_0(X)} = \Upsilon \tag{41}$$

where Υ is the *effective stimulus tuning* (Hansel and Sompolinsky, 1998), defined by:

$$\Upsilon = \frac{\epsilon C}{C - T} \tag{42}$$

The solution for μ and ρ is given by the same equations we derived for the spontaneous activity (Eqs. (24) and (25)), and, as it was in that case, these equations imply that inequality (26) must hold. In addition, in the evoked case, because $\Upsilon > 0$, it must also hold that:

$$1 - J_2 \mathcal{F}_2(X) > 0 \tag{43}$$

Taken together, a fixed point solution requires finding X that satisfies Eq. (41) and inequalities (26), (43). Thus, we check the existence of such a solution in the different regions of the $J_0 J_2$ plane.

As in the case of the spontaneous activity, when $J_0 > 1$, inequality (26) never holds, and no fixed point exists. In the region corresponding to the linear phase of the spontaneous activity, i.e., $J_0 < 1$, $J_2 < 2$, inequality (43) holds for any X, and inequality (26) holds for any $X > X_0$, where X_0 is a function of J_0 (Section 3.1.3). On the interval $X > X_0$ the left hand side of (41) is a monotonically decreasing function of X, approaching ∞ at $X = X_0$, and 0 at $X \to \infty$ (Fig. 5(A)). Therefore, in this region, Eq. (41) has a unique solution for any (positive) Υ, and a unique fixed point exists.

In the region corresponding to the marginal phase of the activity, i.e., $J_0 < 1$, $J_2 > 2$, $X_0 < X_2$, where X_0 is a function of J_0 and X_2 is a function of J_2, (Section 3.1.3) inequalities (26) and (43) imply $X_0 < X < X_2$. On the interval $X_0 < X < X_2$, the left hand side of Eq. (41) is monotonically decreasing of X, approaching ∞ at $X = X_0$, and taking the value of 0 at $X = X_2$ (Fig. 5(B)). Therefore, in this region, Eq. (41) has a unique solution for any (positive) Υ, and a unique fixed point exists. Finally, in the region $J_0 < 1$, $J_2 > 2$, $X_0 > X_2$, inequalities (26) and (43) cannot be satisfied simultaneously, therefore, no fixed point exists.

3.2.3. Phase diagram

In sum, the above analysis shows that the region where a fixed point exists for the evoked activity, coincides precisely with the regions where a fixed point exists for the spontaneous activity, i.e., the regions of the linear and marginal

phases of the spontaneous activity. The stability analysis in the appendix shows that this fixed point is stable, and that any other fixed point in this region is unstable. Similar arguments to those presented for the marginal phase of the spontaneous activity imply that the total synaptic current is a scaled/shifted version of an OM. However, unlike the case of the spontaneous activity where the orientation of the OM was arbitrary, for the evoked activity, the orientation of the OM is bound to match the orientation of the afferent input.

Although I^{tot} is always a scaled/shifted OM, when firing rates are taken into account, we find that two types of solutions exist. The fixed point solution for the firing rate is given by:

$$m_x = \rho_{\text{tot}}[r_x \cos(\theta_x - \psi_{\text{aff}}) + X]_+ \tag{44}$$

Thus, if the selectivities (r_x) are bounded, i.e., there exists $r_{\max}$ with $P(r > r_{\max}) = 0$, and X is large enough, all locations are above threshold, and the system at fixed point is effectively linear. Otherwise, at least some locations are below threshold, and the system is non-linear. This distinction between linear and non linear solutions is important because linear solutions are associated with broad tuning curves, whereas non-linear solutions are associated with narrow tuning curves (Hansel and Sompolinsky, 1998). If $J_2 > 2$, the system is non-linear for any Υ, because the fixed point solution for X is smaller than X_2, which in turn is smaller that $r_{\max}$ (property (vi)). However, if $J_2 < 2$, the fixed point solution for X can take arbitrarily large values, and both linear and non-linear solutions are possible. The line separating the linear regime from the non linear one for the evoked case, can be obtained by plugging $X = r_{\max}$ in Eq. (41). Using the fact that $\mathcal{F}_0(r_{\max}) = r_{\max}$ and $\mathcal{F}_2(r_{\max}) = \frac{1}{2}$ (properties (ii), (vi)) we find that:

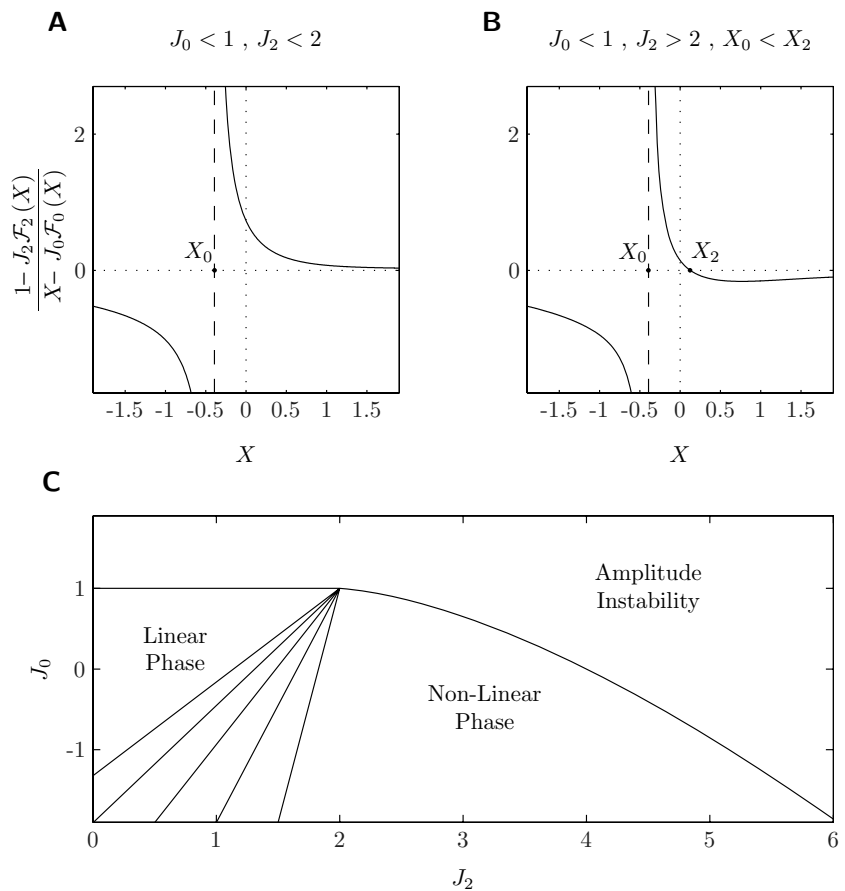$$J_0 - 1 = \frac{J_2 - 2}{2 \Upsilon r_{\max}} \tag{45}$$

Thus, we find that this line always passes through the point $(J_0 = 1, J_2 = 2)$ and its slope (where the line is viewed as a function of J_2) is inversely proportional to Υ and $r_{\max}$. This line for a few choices of Υ is drawn in the phase diagram of the evoked activity (Fig. 5(C)). In the limit $\Upsilon \to 0$, this line becomes $J_2 = 2$, i.e., the line separating the linear phase from the marginal phase of the spontaneous activity. On the other hand, when $\Upsilon \to \infty$ the system switches from linear solutions to non-linear ones, for any choice of J_0 and J_2.

3.3. Energy function

The connectivity matrix of our model is symmetric; therefore, an energy function for the model can be constructed.

Fig. 5 Solution and phase diagram of the evoked activity model. (A, B) The function $\frac{1-J_2\mathcal{F}_2(X)}{X-J_0\mathcal{F}_0(X)}$ (see Eq. (41)). In (A) $J_0 = -3$, $J_2 = 1.5$, representing the case where $J_0 < 1$ and $J_2 < 2$. On the interval $X > X_0$ the function takes all possible positive values, and a unique solution for X exists for every value of Υ. In (B) $J_0 = -3$, $J_2 = 3.5$, representing the case where $J_0 < 1$, $J_2 > 2$, and $X_0 < X_2$. On the interval $X_0 < X < X_2$ the function takes all possible positive values, and a unique solution for X exists for every value of Υ as well. (C) Phase diagram for the evoked activity. The line separating the linear phase from the non-linear phase is plotted for few values of Υ (0.05, 0.1, 0.15, 0.2, 0.25, bottom-right to top-left). In all examples, $\mathcal{F}_0$ and $\mathcal{F}_2$ were calculated with the experimental P(r) (solid lines in Fig. 3)

The energy function decreases during the evolution of the system, and the fixed points are local minima of the energy function. The appropriate energy function is given by Hopfield, (1984) and Hertz et al. (1991):

$$H = -\frac{1}{2A^2}\iint dx\,dy\,W_{xy}R_x R_y + \frac{1}{2A}\int dx\,R_x^2$$
$$- \frac{1}{A}\int dx\,R_x\left(I_x^{\text{aff}} - T\right) \qquad (46)$$

where $R_x = [I_x^{\text{rec}} + I_x^{\text{aff}} - T]_+$. Using a derivation similar to the one we used to reduce the dynamics of the system to the dynamics of the order parameters, Eq. (46) can be rewritten as a function of the order parameters μ and Z:

$$H = -\frac{1}{2}\rho_{\text{tot}}^2(J_0\mathcal{F}_0(X)^2 + J_2\mathcal{F}_2(X)^2)$$
$$+ \frac{1}{2}(\rho_{\text{tot}}(J_0\mu - C + T)\mathcal{F}_0(X)$$
$$+ ((J_2\rho)^2 - (\epsilon C)^2)\mathcal{F}_2(X)) \qquad (47)$$

with ρ_{tot} and X being functions of μ and Z, according to the definitions given for the evoked activity (Eqs. (30) and (37)),

and the spontaneous activity treated as a particular case with $\epsilon = 0$.

Usually, energy functions of neural networks are high dimensional and cannot be directly visualized. However, Eq. (47) reveals that the energy function for this system is a function of one complex and one real variable (Z and μ). In the case of $J_0 = 0$ the energy is a function of Z only, and therefore can be easily visualized. Figure 6 summarizes the main results we obtained in previous sections, by showing how the energy function looks like in the different regimes of the system. In Fig. 6(A)–(C) the energy is drawn as a function of Z for *spontaneous* activity. In Fig. 6(A), the model is in the linear phase, and the energy function is similar to a cone. Only one local minimum exists and it is at the center of the "cone", i.e., at $Z = 0$. This local minimum corresponds to the uniform solution with $\rho = 0$. In Fig. 6(B), the model is in the marginal phase, and the energy function is similar to a "mexican hat". The ring of local minima corresponds to the infinite number of solutions with $\rho > 0$. In Fig. 6(C), no local minima exist, and the system undergoes amplitude instability. In Fig. 6(D), the case of *evoked* activity is depicted. The energy function is similar to the energy function of the spontaneous activity in the marginal phase, Fig. (6B), but "tilted" in the direction defined by the angle

ψ_{aff}. Thus, the continuum of local minima is lost, and only one local minimum remains. While the afferent input we used to generate Fig. 6(D) is of the type we studied in Section 3.2, qualitatively similar results can be obtained for many other types of inputs. This is because the ring attractor, as depicted in Fig. 6(B), is singular; any small anisotropy in the afferent input will destroy its prefect symmetry, and will cause the network to converge to particular values of Z, representing specific OMs.

4. Simulations

To illustrate the theoretical results obtained in the previous section and validate their applicability to experimental data, we applied the model to the Kenet et al. (2003) data set. An important assumption we made in the previous section is that of isotropy of orientation preference (Section 3.1.1). The isotropy assumption requires the distribution of the values of z_x to be rotationally symmetric about the origin. The experimental distribution of z_x, which is shown in Fig. 1(B), does not seem to deviate strongly form the rotational symmetry assumption. This is consistent with some previous optical imaging studies in cats that either did not find any systematic over-representations of specific orientations (Bonhoeffer and Grinvald, 1993; Kenet et al., 2003), or found small, albeit significant in some cases, over representation of cardinal orientations (Müller et al., 2000; Yu and Shou, 2000; Dragoi et al., 2001b; Wang et al., 2003). However, there are many factors that might distort the precise distribution and mask the true extent of the anisotropy. In particular, the fact that only a small fraction of V1 is imaged can increase the anisotropy in the imaged map compared to the true one, whereas measurement noise of the optical imaging data, can artificially induce an opposite effect.

In this work, we avoid the precise characterization of the anisotropy in the cat data. Instead, we first adjusted the polar map (PM) to match the isotropy assumption, and then quantified the deviation between the experimental and the adjusted PMs. The adjustment was performed in two steps. In the first step, we adjusted the values of r_x such that for any angle θ_x the conditional distribution of r_x given θ_x will match, approximately, the marginal distribution of r_x. This was achieved by considering for every pixel x, the set of pixels $\{y\}$ such that $|\theta_x - \theta_y| \leq \frac{\pi}{10}$. Next, we calculated p_x, defined as the percent of pixels with $r_y \leq r_x$. Then, r_x was set to the p_x-th percentile of the distribution of the original values of r_x of all the pixels. In the second step, the pixels were sorted according to the value of r_x and divided into 12 groups of approximately equal size. The values of r at each group were set to the average value at the group. Then, the pixels at each group were sorted according to θ_x, and the values of θ_x were adjusted such that difference between two adjacent θs

was made constant. The distribution of the values of z_x after the adjustment is shown in Fig. 7(B). This distribution is rotationally symmetric, and meets the isotropy assumption.

The PM after isotropy adjustment is shown in Fig. 7(A) and it is very similar to the experimental PM shown in Fig. 1(A). Since our main motivation of using the PM is its ability to provide an approximation to the OMs (Eq. (2)), we quantified the distortion caused by the isotropy adjustment by calculating for every orientation the correlation between the OM approximated with the experimental PM and the OM approximated with the adjusted PM. The average correlation over all possible orientations was 0.971 (min: 0.968, max: 0.973). These high correlations indicate that the original map was not very far from meeting the isotropy assumption in the first place. The average correlation between the 8 experimental OMs and the OMs approximated with the adjusted PM was high as well (0.791; min: 0.725, max: 0.855). We therefore used the adjusted PM to construct the connectivity matrix (Eq. (6)) and tested the model in the spontaneous activity and the evoked activity regimes.

Our analysis of the spontaneous activity in the marginal phase showed that the steady state synaptic input ($I^{\text{rec}} + C$) is a scaled/shifted version of an approximated OM. The orientation of the OM is arbitrary and therefore depends only on the initial state, i.e., m_x at $t = 0$. We thus performed 10000 simulations with random initial states. The simulations were performed using MATLAB$^{\circledR}$, using first order Euler method to solve Eq. (4). Each simulation was let run until convergence. The parameters of the model were fixed for all simulations, with J_0 and J_2 selected to be in the marginal phase (see Fig. 8 for a full list of parameters' values).

An example of a single simulation is presented in Fig. 8(A)–8(C). Figure 8(A) shows the initial state. Figure 8(B) shows the total synaptic input, $I^{\text{rec}} + C$, at the fixed point. It is a linear transformation of the approximated OM of $90°$. The highly similar experimental OM of $90°$ is presented in Fig. 8(C) for comparison (correlation = 0.85). The steady states of all simulations where similar to one of the experimental OMs (correlation >0.7). The high similarity between the simulated and experimental maps demonstrates that the model produced a good approximation for the experimental OM. It also implies that any violations of the approximation assumptions, (ability to approximate OMs with the PM and isotropy) did not cause any significant distortion of the OM. In Fig. 8(D) we show the distribution of the orientations of the steady state OMs for all 10000 simulations. The orientation was calculated by taking the angle of the projection of the synaptic input at the fixed point on the PM. Consistently with our theoretical analysis, this figure shows that the OM at the fixed point was arbitrary, with equal probability for all orientations.

We next tested the model with a tuned input, with $\epsilon = 0.1$ and ψ_{aff} selected randomly from a uniform distribution. In

Fig. 6 Energy function. In the examples given in this figure we assume $J_0 = 0$, and the energy, H, depends only on the complex order parameter Z. (A, B, C) The energy as a function of Z for spontaneous activity. (A) $J_2 = 1$, linear phase. (B) $J_2 = 3$, marginal phase. (C) $J_2 = 5$, amplitude instability. (D) The energy as a function of Z for evoked activity with $J_2 = 3$, $\epsilon = 0.04$. In all examples $\mathcal{F}_0$ and $\mathcal{F}_2$ were calculated with the experimental P(r) (solid lines in Fig. 3). The values of other parameters were $C = 3$, $T = 1$. For clarity the energy was drawn only for values of Z with $|Z| < 3.3$
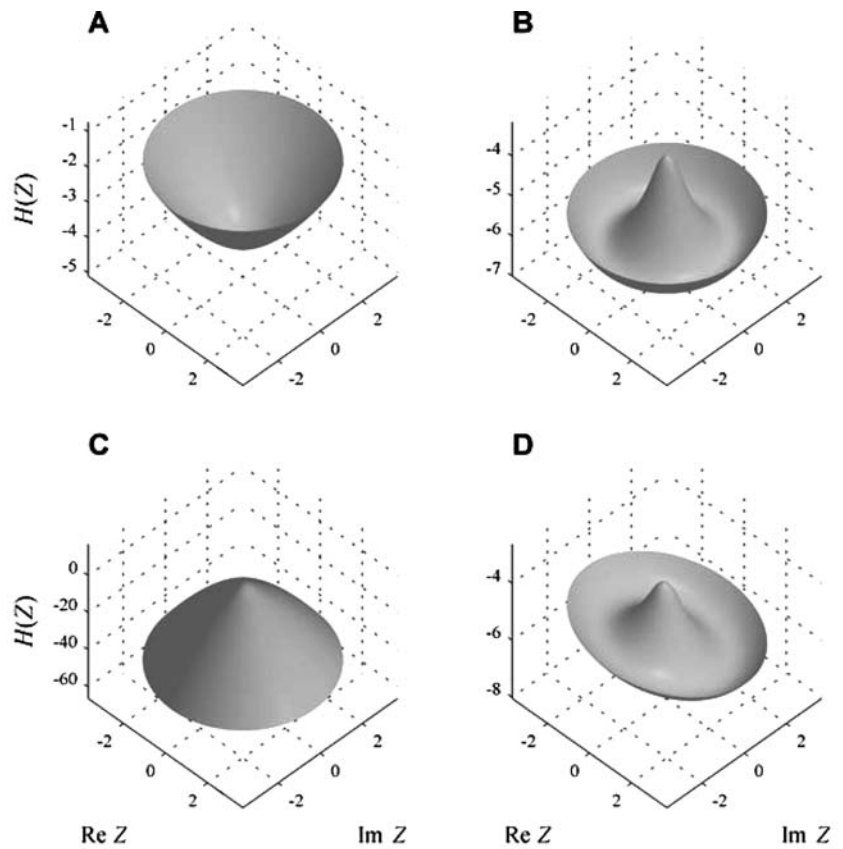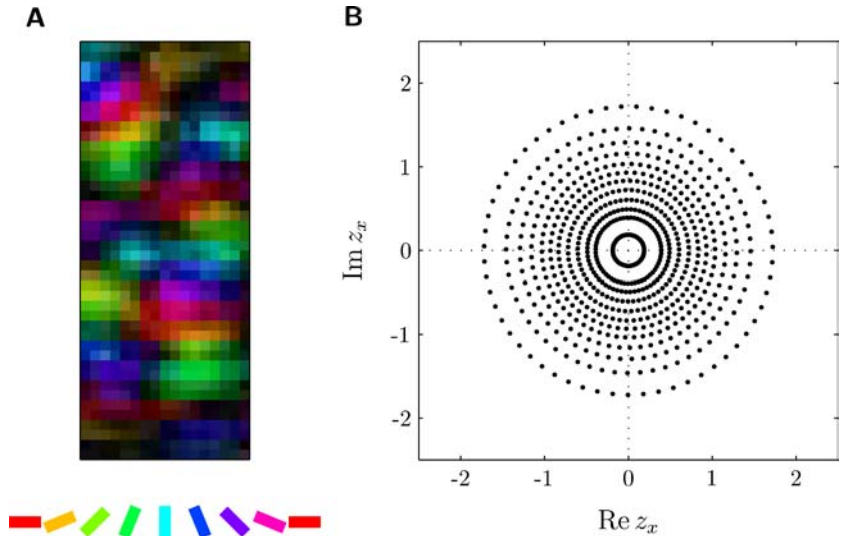

Fig. 7 (A) PM for the data obtained in Kenet et al. (2003) after isotropy adjustment. The figure is displayed on the same brightness scale as in 1A to facilitate comparison. (B) Distribution over the complex plane of the values of z_x of the PM in (A)

order to test the model with a more complex afferent input than the one we analyzed in the previous section, we added to the afferent input gaussian noise with zero mean and a standard deviation of 0.1. Thus, noise in the afferent input was equal in amplitude to ϵ. The other parameters of the simulation were identical to those of the spontaneous activity simulation, and the simulation was also performed 10000 times. The results are shown in Fig. 9. Figures 9(A)–(C) show an ex-

ample of a single simulation. Figure 9(A) shows the afferent input. The orientation used to construct the input was $0°$. The steady state is shown in Fig. 9(B). It is very similar to the experimental OM of $0°$ degrees which is shown in 9C. Note that the noise we added in the afferent input is filtered out by the recurrent connections. Finally, in Fig. 9(D), we show the distribution of the difference between the afferent input orientation and the steady state orientation. The narrow distribution
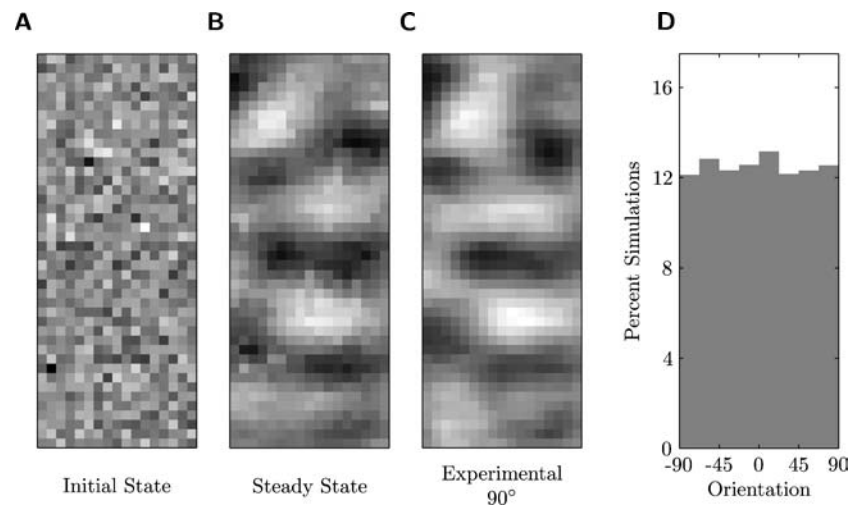
Fig. 8 Simulations of spontaneous activity. (A) Initial state in one simulation. (B) The total synaptic input at steady state for the same simulation. (C) The experimental OM of 90°, which is very similar to the state in (B). (D) Distribution of the orientations of the approximated OMs at steady state in 10000 simulations. The number of bins was 8. The following values of parameters were used in all simulations:

$J_0 = -2$, $J_2 = 5$, $\tau = 10$, $C = 2$, $T = 1$. The random initial state was taken from a gaussian distribution with mean 1 and variance 0.25. Equation (4) was solved using the first order Euler method with $\Delta t = 1$. The simulation was run until $t = 500$, a value that was large enough to guarantee the convergence to a fixed point. Grayscale range (arbitrary units): (A) 0.2–2, (B) −9–7, (C) −2.4–2.7

around zero (SD = 2.2°) shows that the input orientation and the steady state orientation are very close to one another despite the noise. A comparison of Fig. 8(D) and Fig. 9(D) demonstrates the main difference between the spontaneous and evoked conditions in the marginal phase. In the former, the OM is chosen randomly whereas in the latter, it is chosen by the input. However, in both cases, the steady states are similar to OMs.

5. Model with spatially restricted connections

The connectivity pattern we considered so far (Eq. (6)) allowed a thorough mathematical analysis, which we presented in Section 3. However, this pattern is not realistic when large cortical areas are considered because the synaptic weights do not fall off with the distance between the pre- and post-synaptic locations. In contrast, lateral cortical projections in the cortex are spatially restricted. In the cat V1 these projections do not exceed 3.5 mm, with most axons projecting up to a distance of 0.5 mm (Gilbert and Wiesel, 1989; Kisvárday et al., 1997; Yousef et al., 2001). In this section we extend the model by imposing a restriction on the spatial extent of the lateral connections. This restriction is achieved by multiplying the synaptic strength given by Eq. (6) with a spatially isotropic gaussian of the distance between the pre- and post-synaptic locations, i.e.:

$$W_{xy} = (J_2 r_x r_y \cos(\theta_x - \theta_y) + J_0) G_\sigma(\|x - y\|) \qquad (48)$$

where σ denotes the standard deviation of the gaussian that determines the spatial range of the connections, G_σ denotes a 2-dimensional gaussian function ($G_\sigma(\xi) = \frac{1}{2\pi\sigma^2} e^{-\xi^2/(2\sigma^2)}$) and $\|x - y\|$ is the distance between locations x and y. An example of the connectivity pattern induced by Eq. (48) for the Kenet et al. (2003) data set, is given in Fig. 2(B). Like the connectivity pattern induced by the spatially unrestricted model (Eq. (6), Fig. 2(A)), it exhibits a patchy structure, with patches corresponding to regions with similar orientation preference. However, now the patches are spatially restricted, and their density decays as a function of the distance between the interconnected locations. Both properties have been observed experimentally (Gilbert and Wiesel, 1989; Malach et al., 1993; Bosking et al., 1997; Kisvárday et al., 1997).

While the connectivity given by Eq. (48) is more realistic than the one given by Eq. (6), the behavior of the model is now much more complicated. This is because it depends not only on the distribution of preferred orientations and selectivities, but also on the spatial structure of the PM. One implication of this additional complexity is the loss of isotropy. While the PM can be still isotropic over the entire region simulated, each location is now connected to only a small subregion; in this subregion isotropy does not necessarily hold. One might expect that this would disrupt the formation of the ring attractor, and indeed simulations of the spatially restricted model with a uniform input yielded only one or two attractor states, depending on the model parameters (e.g., J_2, σ). These attractor states were usually similar to OMs, with the identity of the attractor-OM also
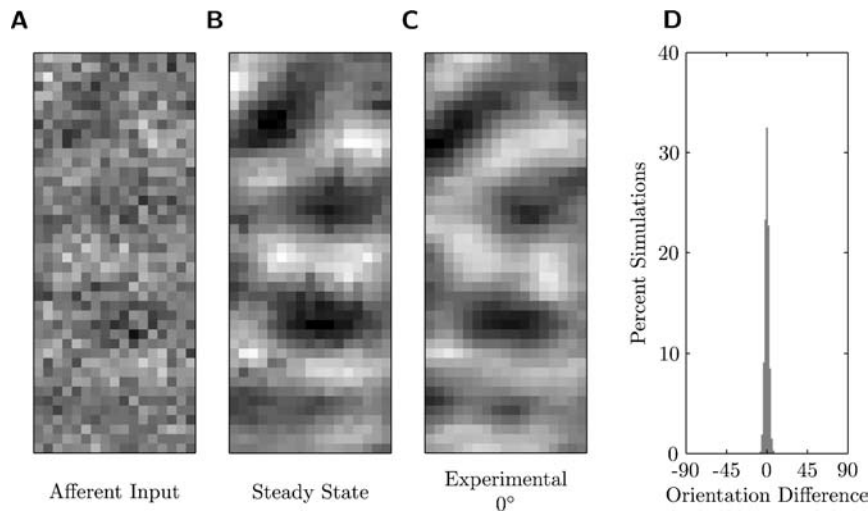
Fig. 9 Simulations of evoked activity. (A) Afferent input in one simulation. The input is constant, with a small modulation in the shape of the approximated OM of 0° ($\epsilon = 0.1$, $\psi_{aff} = 0$) and a random gaussian noise (mean = 0, std = 0.1). (B) The total synaptic input at steady state for the same simulation. The 0° OM is extracted from the afferent input. (C) The experimental OM of 0°, which is very similar to the state in (B). (D) Distribution of the difference between the orientation of the approximated OM at steady state and the orientation encoded in the afferent input for 10000 simulations. The number of bins was 100. The parameters for these simulations were identical to the parameters used for the spontaneous activity simulations. Grayscale range (arbitrary units): (A) 1.5–2.5, (B) −9–7, (C) −2.4–2.7

depending on the precise selection of the model parameters. A similar phenomenon, in which the ring attractor collapses into few isolated attractors, occurs in the ring model when some degree of anisotropy is introduced, i.e. when the distribution of preferred orientations is not uniform (Tsodyks and Sejnowski, 1995; Zhang, 1996; Renart et al., 2003; Goldberg et al., 2004). Thus, the spatially restricted model does not support a ring attractor in the sense we discussed so far. However, if the spatial range of the interactions is large enough, such that the recurrent input received by each location is *approximately* isotropic, the network has an unlimited number of states that have energy levels similar to the attractors. We therefore expected that these states could be revealed by adding a random noise to the uniform input. To explore this possibility we simulated the network with many realizations of the random noise. The random noise was taken to be a gaussian noise with small spatial correlations induced by filtering it with a 2-dimensional spatially isotropic gaussian kernel with a small standard deviation (0.1 mm) (see a similar approach in Goldberg et al. (2004)). An example of two realizations of the input is given in Fig. 10(A, D). To allow comparison between the spatially restricted and unrestricted models under the same conditions, we repeated the simulations of spontaneous activity in the spatially unrestricted model (Section 4) using the noisy input we used for the spatially restricted model. The results for the unrestricted model with the noisy input were very similar to the results with the uniform input, implying that our characterization of spontaneous activity in Section 4 holds also for the noisy input.

For the spatially restricted model we constructed the lateral connectivity with Eq. (48), using the same PM we used for the simulations in Section 4. We took $\sigma = 0.6$ mm, which is approximately the diameter of a hypercolumn. It has been estimated that in cats most projections are approximately within this distance (Kisvárday et al., 1997; Schummers et al., 2002). The global parameters of the lateral connections were $J_0 = -3$ and $J_2 = 3.5$, which correspond to the marginal phase of the spatially unrestricted model. Any single realization of the input, such as the ones in Fig. 10(A) and (D), yielded one particular steady state solution. However, for different realizations of the input, the network had an unlimited number of attractors states. Some of these states were similar to OMs, like in the case of the spatially unrestricted model. Still, other steady states were not highly correlated with OMs. An example of two such states is given in Fig. 10(B) and (E). A close examination of these states revealed that they were mosaics of OMs, i.e., they were composed of different OMs in different regions of the modeled area. The structure of these states is demonstrated in Fig. 10(C, F). For example, in Fig. 10(C), the OM of 22.5° is presented in the upper part, and the OM of 67.5° is presented in the lower part. This combination of OMs presented in Fig. 10(C) produces a state that is very similar to the steady state shown in Fig. 10(B). We conclude that with noisy afferent input the model with spatially restricted connections is qualitatively similar to the model of spontaneous activity we considered in Section 3.1, except that steady states can be mosaics of OMs, rather than a single OM.
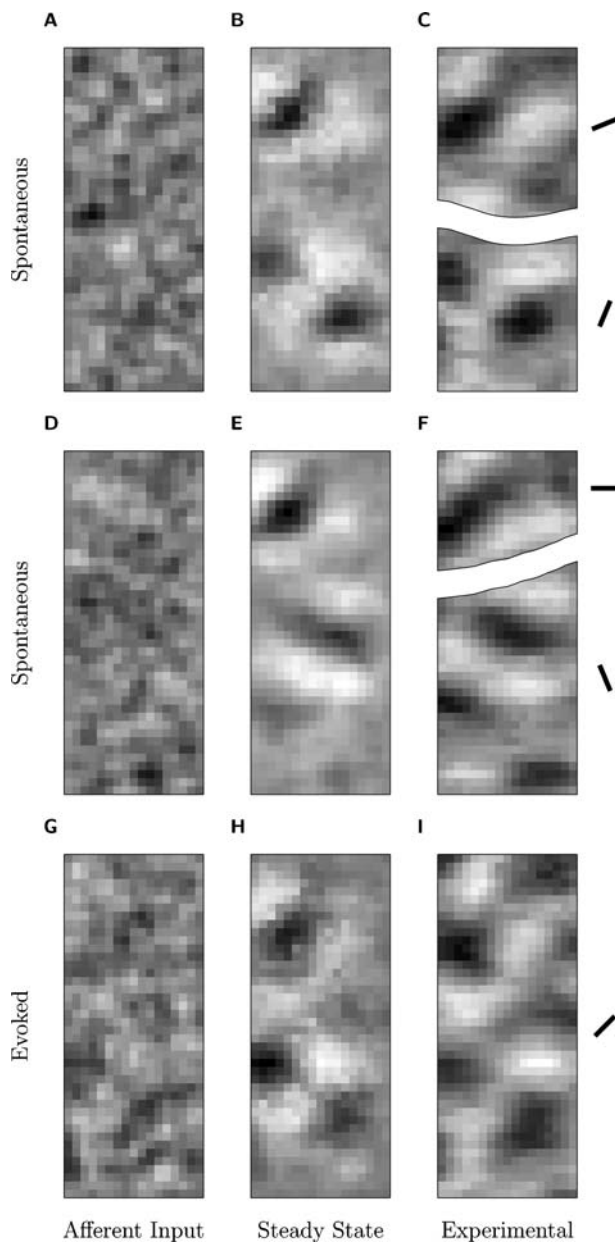
Fig. 10 Simulations of the model with spatially restricted connectivity. (A, D) Afferent input for two simulations of spontaneous activity. (B, E) The total synaptic input at steady state for the same simulations. (C, F) Mosaics of experimental OMs that are highly similar to the steady states in (B) and (E). (G) Afferent input in a simulation of evoked activity encoding an orientation of 45°. (H) Steady state for the same simulation. (I) The experimental OM of 45°. The small bars on the right-hand side of the figure indicate the orientations of the OMs in (C), (F), (I). The values of the parameters were $J_0 = -3$, $J_2 = 3.5$, $\sigma = 0.6$ mm, $C = 2$, $T = 1$. The noise in the input was random gaussian with zero mean, 0.1 standard deviation, and it was spatially filtered with a spatially isotropic gaussian with a standard deviation of 0.1 mm. Grayscale range (arbitrary units): (A, D, G) 1.1–2.9, (B, E) −0.8–4.6, (H) −1.3–4.3, (C, F, I) −2.4–2.7

We also tested this model in the evoked activity case, by adding a small modulation in the shape of an approximated OM to the noisy input described above (similarly to our approach in Section 4). An example of such input is presented in Fig. 10(F) where the orientation of the OM was 45°. The steady state for this simulation is presented in Fig. 10(G), and the experimental OM for 45° is presented in 10H. The steady state is similar to the OM encoded in the input, and the noise is filtered out. This was the case for all simulations. Thus, for evoked activity, the behavior of the model with spatially restricted connections is very similar to the behavior of the model with unrestricted connections.

6. Discussion

The emergence of OMs during spontaneous activity in anaesthetized cats has led to the suggestion that OMs are attractors of the intracortical network (Kenet et al., 2003). Adopting this view, we posed the following question: what type of intracortical dynamics and connectivity can lead to the formation of attractors at the OMs? In this work, we suggested a rate model endowed with a simple connectivity rule (Eq. (6)), and showed that it yields attractor states that are highly similar to OMs. Specifically, we showed analytically that in the proper parameter regime (marginal phase), and given a uniform input, the model has a ring attractor formed by approximated OMs of arbitrary orientations. This property explains the formation of OMs during spontaneous activity where the afferent input is assumed to be unstructured. We also considered the case where the activity is evoked by a visual stimulus and showed how a structured afferent input can select the OM that matches the stimulus' orientation. The model therefore suggests that OMs are encoded in the lateral connections, and that these connections can generate OMs both when the activity is spontaneous and when it is evoked by a visual stimulus.

6.1. Selectivity

The core of our model is its connectivity pattern (Eqs. (6), (48)). This connectivity pattern depends on three properties of the interconnected sites: the selectivity, the preferred orientation, and the distance between the pre- and post- synaptic locations. The main theoretical innovation in our model is the explicit incorporation of the selectivity variable into the connectivity pattern. We treat the selectivity as a structural attribute of every cortical location, similarly to how the preferred orientation is dealt with by many models (including ours). We have shown that this approach is sufficient for explaining the correspondence between OMs and intrinsically preferred states of the cortical network. The model therefore predicts a different connectivity pattern for neurons near

pinwheels and neurons in linear zones. There are some experimental evidence in favor of this view. Yousef et al. (2001) reported that connections of neurons near pinwheels have shorter lateral extent that neurons in linear zones, and that these connections are not as orientation-specific as those of neurons in linear zones. The latter result is consistent with our connectivity matrix, in which the orientation-dependent term in the equation of synaptic strength (Eq. (6)) is multiplied by the selectivity.

In the context of studying orientation selectivity, it is important to distinguish between two types tuning curves: of membrane potential tuning curves, on which we focused in this contribution, and commonly considered firing rate tuning curves. These two types of tuning curves give rise to two different measures of selectivity, and the closely related measures of tuning width. In our model, the shape of the membrane potential tuning curve is the same for all locations (a cosine shape). The different selectivity of different locations comes from different scaling of the tuning curve. As for firing rate tuning curves, the picture is more complicated. Equation (44) implies that the selectivity and width of firing rate tuning curves vary with membrane potential selectivity. The precise characteristics of the firing rate tuning curve depends on the model parameters (J_0, J_2), and both broad and narrow tuning curves are possible. In the linear regime, all locations have broad tuning curves. In the non-linear regime, both broad and narrow tuning curves are possible.

Several experimental studies used recordings of single neurons to study the dependence of selectivity and other tuning curve properties on the cortical location. Before we compare our model with these studies, it is important to note that the low selectivity near pinwheels seen with optical imaging, which is at the heart of our modeling approach, is difficult to interpret in terms of single neuron properties. This is because it can arise from either lack of orientation selectivity of single neurons near the pinwheel, or from averaging highly selective neurons with a highly variable orientation preference. The second option received support form extracellular recordings (Maldonado et al., 1997; Dragoi et al., 2001a) that found that as far as firing rates are concerned, neurons near pinwheels are as sharply tuned as neurons in linear zones. However, intracellular recordings (Schummers et al., 2002) have shown that when tuning curves of subthreshold membrane potential are considered, the selectivity of neurons is indeed reduced with the selectivity given by the PM. Our model, in which selectivity varies both for synaptic-input and firing-rate tuning curves, is therefore consistent with membrane potential tuning curves, but not with firing rate tuning curves. As suggested by Schummers et al. (2002), this difference between the tuning of membrane potential and the tuning of firing rate might be accounted for by different parameters of the gain function (e.g., firing threshold) for neurons near pinwheels. This suggestion can be easily incorporated into our model. Other mechanisms were suggested, some of them, like sensitivity to the temporal structure of the membrane potential fluctuations (Volgushev et al., 2002), are beyond the framework of the rate models we discussed here.

The notion of tuning curves, is clearly associated with evoked activity. However, the principles laid down in this work can lead to the definition of "spontaneous tuning curves". In an experimental setup, if optical imaging of spontaneous activity is performed simultaneously with single unit recordings, a spontaneous tuning curve can be defined as the average firing rate (measured electrophysiologically) as a function of the orientation encoded by the cortical state (measured by the optical imaging). It would be interesting to characterize spontaneous tuning curves and compare them with the classical, evoked, tuning curves. In one particular scenario our model predicts a qualitative difference between spontaneous and evoked tuning curves. A comparison between the phase diagrams for spontaneous and evoked activity (Figs. 4(E), 5(C)) implies that if V1 operates near the phase transition between the linear and marginal phases of spontaneous activity, (J_2 close to 2), spontaneous tuning curves are expected to be significantly broader than the evoked ones. This prediction can be tested experimentally.

6.2. Orientation preference

Similarly to other models, our connectivity pattern suggests a more excitatory (or less inhibitory) coupling between neurons with similar orientation preference. This is in agreement with many experimental works that have shown that neurons tend to make connections with other neurons with a similar orientation preference. It has been argued that this is only a weak trend that is true only on average (Kisvárday et al., 1997). This does not necessarily contradict our model because the model considers populations of neurons and not single neurons. In addition, both in the linear and in the marginal phases, the ratio between the preferred-orientation dependent and independent components can be made arbitrarily small, by taking large negative value for J_0.

Another challenge for our model is the fact that the dependence of the connectivity on orientation preference is primarily true for long range connections, i.e. connections between different hypercolumns. Moreover, it has been suggested that short-range connections are *spatially* isotropic (Ts'o et al., 1986; Das and Gilbert, 1999) (note that isotropy here refers to invariance with respect to direction on the cortical sheet, unlike the isotropy in the feature space we considered before). In contrast, our model does not distinguish between long-range and short-range connections, and in particular it does not include explicitly isotropic short range connections. We focused on the long-range type

connections because isotropic connections alone, are hard to reconcile with the spontaneous emergence of OMs. This type of connectivity is invariant with respect to spatial rotation/translation. Consequently, the spontaneous patterns it induces are also invariant with respect to rotation/translation. However, experimentally observed patterns of spontaneous activity are correlated with the OMs and do not exhibit this invariance (Kenet et al., 2003). It is also important to note that because of the dependence of the connectivity on the selectivity, our model does not exhibit a strong spatial anisotropy of the short-range connections. In the linear zones, neurons are surrounded by neurons with similar orientation preference, and therefore they exhibit roughly isotropic connections. Near pinwheels this is not true anymore; However there the connectivity depends only weakly on the orientation preference, and the global component (J_0) dominates the connectivity, producing again a roughly isotropic connectivity. Nonetheless, an explicit incorporation of isotropic short range connections might be considered in a future work.

6.3. Spatial structure of the OMs

Two assumptions about the structure of the OMs were required for the formation of the ring attractor. The first assumption is that OMs can be well-approximated using the PM (Eq. (2)), that is, by considering at each location only the preferred orientation and selectivity. This assumption can be expressed by the condition that on average, a large fraction of the variance of the pixels' tuning curves can be explained by its first Fourier component (Eq. (3)). We found that this condition is met in experimental data sets, and that individual OMs are highly correlated with their approximated versions. We note however, that the approximation of orientation tuning curves with cosine functions, as in Eq. (2), was criticized recently on the ground that it cannot describe the height and width of the tuning curves independently (Swindale et al., 2003). While deviations from cosine tuning curves can be functionally important, our analysis indicates that the PM provides a good first-order approximation for the OMs. The examples in Figs. 8–10 of individual experimental OMs compared with the output of the model, further support this claim.

The second assumption we made about the OMs is that the representation of orientation preference in the PM is isotropic, that is, that the distribution of preferred orientations is uniform and independent of the distribution of selectivities. For the experimental data set, we showed that orientation preference did not deviate strongly from the isotropy assumption. As discussed in Section 4, this result is consistent with some experimental studies in cats. Other studies, however, found a small bias of the distribution of preferred orientations towards the cardinal orientations. Interestingly,

Kenet et al. (2003) reported a similar cardinal-orientations bias in the distribution of spontaneous OMs. It is thus conceivable that a small anisotropy in the distribution of preferred orientations indeed exists in cats' V1, and that this anisotropy induces a significant anisotropy in spontaneous activity patterns. While our current model does not account for anisotropy, it provides a framework in which the effects of anisotropy in V1 could be addressed in future works.

6.4. Relation to the ring model

The model we presented here is strongly related to the ring model for feature selectivity (Ben-Yishai et al., 1995; Hansel and Sompolinsky, 1998). In fact, the ring model can be viewed as a special case of our model, in which all locations have the same selectivity. Therefore, these two models share some features that do not depend on the distribution of selectivities P(r), e.g., the general structure of the phase diagrams (Figs. 8(E) and 9(C)). Since in our model multiple selectivities are allowed, the attractor states can exhibit a much richer spatial structure and approximate the experimental OMs.

6.5. Lateral extent of the connections

For most of the work, the connectivity pattern that we considered (Eq. (6)), did not depend on the distance between the pre- and post- synaptic locations. In contrast, neuronal connections in the cortex have a limited lateral extent. Therefore, in Section 5, we suggested a modification of the connectivity rule in which the connections were spatially restricted. Using simulations, we showed that this version of the model has properties similar to the spatially unrestricted one, except for one important difference: the states generated during spontaneous activity are not necessarily OMs, but can be a mosaic of several OMs. Consequently, we predict that such states exist in spontaneous activity data. Preliminary analysis we of the data obtained in Kenet et al. (2003), indicates that mosaic states indeed emerge spontaneously.

It is important to note that the distinction between OM states and mosaic states is somewhat blurred. An OM state might in fact be a small part of a larger mosaic state, that would have been observed had a larger cortical area been considered. It is also possible, that what looks as an OM state is a composition of several OMs of similar orientations, that cannot be distinguished with current techniques. We can thus expect that if a large cortical region will be imaged during spontaneous activity, spontaneous states will correspond more closely to mosaic states, rather than to OM states. From a theoretical point of view, the prevalence and characteristics of mosaic states in spontaneous activity is an interesting question because mosaic states provide a signature of the effective strength and extent of the

lateral connections in V1. Spontaneous states that correspond to OM states rather than to mosaic states indicate a strong lateral connectivity with large spatial extent. A further analysis, which is beyond the scope of this contribution, is required to fully characterize the behavior of the spatially restricted model, and its implications on spontaneous activity.

6.6. Dynamics

Our approach was based on identifying the attractor states of the intracortical network with the OMs. Because we used a static input and simple rate dynamics, attractor states always existed; the network always converged to a stationary activity profile. In contrast, spontaneous activity in anaesthetized cats is highly dynamic, continuously switching between states. This gap between the dynamics of spontaneous activity in cats and the dynamics of our model can be bridged by considering various mechanisms that can prevent the asymptotic convergence of the model to a fixed point. The dynamical switching between states can be induced by a dynamical external input. Specifically, Goldberg et al. (2004) suggested that spatio-temporally correlated noise coming from the LGN may drive the cortical dynamics. In the *marginal phase*, where the shape of the activity is dominated by recurrent connections, this type of afferent noise would cause the network to perform a random walk in the space of OMs. An analysis of experimental data reveals that signatures of its dynamics are different from those of a simple random walk along the ring attractor. Another possibility, which might be more compatible with experimental data, is that the system is in the *linear phase*. The observed cortical states in this case are the sum of the afferent input and a component in the shape of the OMs, induced by the recurrent connections (Goldberg et al., 2004).

The dynamical switching between states can be also induced by mechanisms *intrinsic* to V1. For example, firing rate adaptation or synaptic depression/facilitation can destabilize the attractor states on a slow time-scale. Including a term for firing rate adaptation in the dynamical equation (Eq. (4)), as in Hansel and Sompolinsky (1998), causes the network, in the marginal phase, to perform a limit-cycle in the space of OMs. This possibility seems too simplistic to account for the spontaneous cortical dynamics. However, a more complicated dynamical behavior can emerge if synaptic depression/facilitation is included. Such complicated behavior indeed emerges in the ring model with synaptic dynamics (Tsodyks, unpublished results), and we expect similar results for our model. It is also likely that spontaneous activity in V1 is driven by a combination of several of the mechanisms we mentioned above. Our model provides the framework in which the contribution of each of these mechanisms could be studied.

6.7. The stimulus

Throughout the paper, we assumed that the evoked OMs encode the *orientation* of the stimulus. This might not be the case. Basole et al. (2003) found that states highly similar to OMs, can also be evoked by texture stimuli containing short line segments. The OM to which the resulting state corresponded depended on both the orientation of the line segments and the direction in which they moved. The authors concluded that OM-like states encode the spatio-temporal properties of the stimulus, rather than its (spatial) orientation. The issue of which stimulus property is encoded by the OMs is largely avoided in our model because we did not consider in details how a particular visual stimulus is transformed into a pattern of LGN input to V1. For the types of input we considered, the LGN-V1 mapping is assumed to extract some information about features of the visual scene (e.g. orientation) and the intrinsic V1 processing acts to amplify and denoise the afferent signal. Input patterns of higher complexity may lead to a more intricate transformation by the recurrent V1 network.

In a broader context, our model can be considered as a general model for population encoding of a sensory or motor variable in the presence of a weak or noisy input. This variable can be some spatio-temporal characteristic of a sensory stimulus, but also direction of a planned movement, or direction of gaze. The variable encoded must however be *periodic*, otherwise the concept of a ring attractor is not relevant. In fact, the generality of the model is inherited from the ring model, which our model generalizes. Extensions of the ring model and other similar models supporting ring attractors have been considered in several systems such as spatial working-memory (Camperi and Wang, 1998; Compte et al., 2000), head direction system (Skaggs et al., 1995; Redish et al., 1996; Zhang, 1996; Xie et al., 2002; Boucheny et al., 2005), hippocampal place cells (Tsodyks and Sejnowski, 1995), general statistical inference (Pouget et al., 1998; Deneve et al., 1999) and TMS-induced perceptual suppression (Miyawaki and Okada, 2004). Some of these models build directly on the formulation of Ben-Yishai et al. (1995); others use different formulations but are similar in many aspects (Ermentrout, 1998). Our model adds to this class of models the possibility of variability in response amplitude, i.e. selectivity, of different neurons. As a result, the ring attractor is not associated with a simple spatial form of the activity pattern, like a single localized bump in the ring model, but rather includes the complex spatial structure of the OMs.

Regardless of the precise nature of the features represented by the OMs, the spatially restricted version of the model suggests how local features of the stimulus can be encoded when the stimulus properties, e.g. orientation, vary across the visual field. Since the coarse grain mapping of

the visual field to the cortex is retinotopic, we can expect the activity patterns at a small cortical regions to represent the properties of the stimulus in the corresponding regions of the visual field. The states generated by the model, e.g. the mosaic states in Fig. 10, when evoked, achieve precisely this. We thus predict that states similar to the mosaic states will play important role in encoding complex stimuli whose features vary across the visual field.

Appendices

A.1. Properties of $\mathcal{F}_0$ and $\mathcal{F}_2$

In this appendix, we provide proofs for the properties of $\mathcal{F}_0$ and $\mathcal{F}_2$ given in Table 1. We provide proofs only for the weak inequalities given in that table.

A.1.1. Properties of $\mathcal{F}_0$

$\mathcal{F}_0$ is defined by Eq. (21). Property (i) is trivial, as $\mathcal{F}_0$ is defined by integrating a non-negative function. The derivative of $\mathcal{F}_0$, is given by:

$$\mathcal{F}_0'(X) = \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} \tag{49}$$

The integral on the right-hand side can be interpreted as the probability that $r\cos\theta > -X$, which immediately gives properties (iii) and (iv). In addition, it can be verified that:

$$\iint\limits_{r\cos\theta>X} dr\,d\theta \frac{P(r)}{2\pi} = 1 - \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} \tag{50}$$

therefore $\mathcal{F}_0'(-X) = 1 - \mathcal{F}_0'(X)$. Integrating both sides of the equation from 0 to X, we obtain: $\mathcal{F}_0(-X) = \mathcal{F}_0(X) - X$. Therefore, as follows from property (i), $\mathcal{F}_0(X) - X \geq 0$, which gives property (ii).

A.1.2. Properties of $\mathcal{F}_2$

$\mathcal{F}_2$ is defined by Eq. (22). Its derivative is given by:

$$\mathcal{F}_2'(X) = \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} r\cos\theta \tag{51}$$

Equation (51) immediately implies that property (vii) holds for non-positive X, as the integrated expression is non-negative. For positive X, it can be verified that:

$$\iint\limits_{r\cos\theta>X} dr\,d\theta \frac{P(r)}{2\pi} r\cos\theta = \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} r\cos\theta \tag{52}$$

Therefore, $\mathcal{F}_2'(X) = \mathcal{F}_2'(-X)$, and property (vii) holds for positive X as well. Property (vii) implies that $\mathcal{F}_2$ is non-decreasing, therefore, $\mathcal{F}_2(-\infty)$ and $\mathcal{F}_2(+\infty)$ give, respectively, lower and upper bounds for $\mathcal{F}_2$. Calculating these expressions, properties (v) and (vi) are obtained.

A.1.3. Other properties

For property (viii), we observe that the integral expression defining $\mathcal{F}_0$ (Eq. (21)) can be written as the sum of two integrals as follows:

$$\mathcal{F}_0(X) = \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} r\cos\theta + \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} X \tag{53}$$

Noting that the first term on the right-hand side is $\mathcal{F}_2'(X)$ and that the second term is $X\mathcal{F}_0'(X)$, property (viii) is obtained.

For property (ix), we observe that the integral expression defining $\mathcal{F}_2$ (Eq. (22)) can be written as the sum of two integrals as follows:

$$\mathcal{F}_2(X) = \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} (r\cos\theta)^2 + \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} Xr\cos\theta \tag{54}$$

Noting that the second term on the right-hand side is $X\mathcal{F}_2'(X)$, we obtain:

$$\mathcal{F}_2(X) - X\mathcal{F}_2'(X) = \iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} (r\cos\theta)^2 \tag{55}$$

The expression on the right hand side can be checked to be a non-decreasing function of X, and calculating its value for $X \to \infty$, gives property (ix).

For property (x), we first write the following application of the Cauchy-Schwarz inequality:

$$\left(\iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} r\cos\theta \right)^2 \leq \left(\iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} \right) \left(\iint\limits_{r\cos\theta>-X} dr\,d\theta \frac{P(r)}{2\pi} (r\cos\theta)^2 \right) \tag{56}$$

which reads $(\mathcal{F}_2'(X))^2 \leq \mathcal{F}_0'(X)(\mathcal{F}_2(X) - X\mathcal{F}_2'(X))$. Substituting $\mathcal{F}_2'(X)$ with the left-hand side of property (viii), property (x) is obtained.

A.2. Stability

In this section we study the stability of the non-uniform fixed point. We discuss only the case of the tuned input model (Section 3.2.2). The stability in the case of the spontaneous activity model (Section 3.1.3) can be easily obtained by repeating similar arguments with $\epsilon = 0$.

Let μ, ρ take their fixed point values (Eqs. (24),(25)), and let $\psi = \psi_{\text{aff}}$. We look at a small perturbation about the fixed point: $(\mu + \delta\mu), (\rho + \delta\rho), (\psi + \delta\psi)$. The linearized dynamics of Eqs. (34)–(36) are given by $\tau[\dot{\delta\mu}\ \dot{\delta\rho}\ \dot{\delta\psi}]^T = S[\delta\mu\ \delta\rho\ \delta\psi]^T$ with S being the 3×3 Jacobian matrix given by:

$$S =$$
$$\begin{bmatrix} J_0\mathcal{F}_0'(X) - 1 & J_2(\mathcal{F}_0(X) - X\mathcal{F}_0'(X)) & 0 \\ J_0\mathcal{F}_2'(X) & J_2(\mathcal{F}_2(X) - X\mathcal{F}_2'(X)) - 1 & 0 \\ 0 & 0 & -\epsilon C\mathcal{F}_2(X) \end{bmatrix}$$
$$(57)$$

This matrix implies that the evolution of $\delta\psi$ is decoupled from that of $\delta\mu, \delta\rho$, and that $\delta\psi$ decays with time. To show the decay of $\delta\mu, \delta\rho$, we study the eigenvalues of 2×2 upper-left sub-matrix of S, which we denote by S'. For the eigenvalues of S' to have a negative real part, it is enough to show that the determinant of S', det S', is positive, and that its trace, tr S', is negative. The determinant is given by:

$$\det S' = J_2\mathcal{F}_2'(X)(X - J_0\mathcal{F}_0(X))$$
$$+ (1 - J_0\mathcal{F}_0'(X))(1 - J_2\mathcal{F}_2(X)) \quad (58)$$

For the fixed point value of X it holds that $X - J_0\mathcal{F}_0(X) > 0$ and $1 - J_2\mathcal{F}_2(X) > 0$ (inequalities (26), (43)). In addition, for any X, $\mathcal{F}_2'(X) \geq 0$ (property (vii), Table 1) and when $J_0 < 1$, $1 - J_0\mathcal{F}_0'(X) > 0$ (properties (iii),(iv), Table 1). Therefore, the determinant is positive. The trace is given by:

$$\text{tr } S' = -1 + J_0\mathcal{F}_0'(X) - 1 + J_2\mathcal{F}_2(X) - J_2X\mathcal{F}_2'(X) \quad (59)$$

The inequalities we mentioned above for the determinant imply that when $X \geq 0$, the trace is negative. In addition, if $J_2 < 2$, by property (ix), $J_2(\mathcal{F}_2(X) - X\mathcal{F}_2'(X)) \leq 1$ and the trace is negative as well. The last case to check is when $X < 0$ and $J_2 > 2$. In this case, from the definitions of X_0

and X_2, it follows that $J_0 = \frac{X_0}{\mathcal{F}_0(X_0)}$ and $J_2 = \frac{1}{\mathcal{F}_2(X_2)}$. Thus, we rewrite the trace as:

$$\text{tr } S' = -2 + J_2\mathcal{F}_2(X) - \frac{(X - X_0)\mathcal{F}_0'(X)}{\mathcal{F}_0(X_0)}$$
$$+ \frac{X(\mathcal{F}_0'(X)\mathcal{F}_2(X_2) - \mathcal{F}_2'(X)\mathcal{F}_0(X_0))}{\mathcal{F}_0(X_0)\mathcal{F}_2(X_2)} \quad (60)$$

Since $-1 + J_2\mathcal{F}_2(X) < 0$, and $X > X_0$, the sum of the first three terms is negative. Therefore, to show that the entire summation is negative, it remains to show that $\mathcal{F}_0'(X)\mathcal{F}_2(X_2) - \mathcal{F}_2'(X)\mathcal{F}_0(X_0) > 0$. First, we note that because $X_0 < X < X_2$ it holds that:

$$\mathcal{F}_0'(X)\mathcal{F}_2(X_2) - \mathcal{F}_2'(X)\mathcal{F}_0(X_0)$$
$$\geq \mathcal{F}_0'(X)\mathcal{F}_2(X) - \mathcal{F}_2'(X)\mathcal{F}_0(X) \quad (61)$$

By property (x), the term on the right-hand side is non-negative. Therefore, the trace is negative.

Finally, our solution for evoked activity has shown that a fixed point with $\psi = \psi_{\text{aff}} + \pi$ can exist. Solving for this point and obtaining the corresponding stability matrix one finds that it is very similar to the one given in Eq. (57), with the evolution of $\delta\psi$ decoupled form that of $\delta\mu, \delta\rho$. However in this case it is given by $\dot{\delta\psi} = \epsilon C\mathcal{F}_2(X)\delta\psi$, which implies that the fixed point is unstable to perturbations in ψ.

References

Basole A, White LE, Fitzpatrick D (2003) Mapping multiple features in the population response of visual cortex. Nature 423: 986–990.

Ben-Shahar O, Zucker S (2004) Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. Neural Comput. 16: 445–476.

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. USA 92: 3844–3848.

Blasdel GG, Salama G (1986) Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. Nature 321: 579–585.

Bonhoeffer T, Grinvald A (1993) The layout of iso-orientation domains in area 18 of cat visual cortex: Optical imaging reveals a pinwheel-like organization. J. Neurosci. 13: 4157–4180.

Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J. Neurosci. 17: 2112–2127.

Boucheny C, Brunel N, Arleo A (2005) A continuous attractor network model without recurrent excitation: Maintenance and integration in the head direction cell system. J. Comput. Neurosci. 18: 205–227.

Buzás P, Eysel UT, Kisvárday ZF (1998) Functional topography of single cortical cells: An intracellular approach combined with optical imaging. Brain, Res. Brain. Res. Protoc. 3: 199–208.

Camperi M, Wang XJ (1998) A model of visuospatial working memory in prefrontal cortex: Recurrent network and cellular bistability. J. Comput. Neurosci. 5: 383–405.

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

Crair MC, Gillespie DC, Stryker MP (1998) The role of visual experience in the development of columns in cat visual cortex. Science 279: 566–570.

Crook JM, Kisvárday ZF, Eysel UT (1998) Evidence for a contribution of lateral inhibition to orientation tuning and direction selectivity in cat visual cortex: Reversible inactivation of functionally characterized sites combined with neuroanatomical tracing techniques. Eur. J. Neurosci. 10: 2056–2075.

Das A, Gilbert CD (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature 399: 655–661.

Deneve S, Latham PE, Pouget A (1999) Reading population codes: a neural implementation of ideal observers. Nat. Neurosci. 2: 740–745.

Dragoi V, Rivadulla C, Sur M (2001a) Foci of orientation plasticity in visual cortex. Nature 411: 80–86.

Dragoi V, Turcu CM, Sur M (2001b) Stability of cortical responses and the statistics of natural scenes. Neuron 32: 1181–1192.

Ermentrout B (1998) Neural networks as spatio-temporal pattern-forming systems. Rep. Prog. Phys. 61: 353–430.

Ernst UA, Pawelzik KR, Sahar-Pikielny C, Tsodyks MV (2001) Intracortical origin of visual maps. Nat. Neurosci. 4: 431–436.

Fiser J, Chiu C, Weliky M (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature 431: 573–578.

Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. J. Neurosci. 9: 2432–2442.

Gödecke I, Kim DS, Bonhoeffer T, Singer W (1997) Development of orientation preference maps in area 18 of kitten visual cortex. Eur. J. Neurosci. 9: 1754–1762.

Goldberg JA, Rokni U, Sompolinsky H (2004) Patterns of ongoing activity and the functional architecture of the primary visual cortex. Neuron 42: 489–500.

Grinvald A, Lieke E, Frostig RD, Gilbert CD, Wiesel TN (1986) Functional architecture of cortex revealed by optical imaging of intrinsic signals. Nature 324: 361–364.

Grinvald A, Shoham D, Shmuel A, Glaser D, Vanzetta I, Shtoyermann E, Slovin H, Sterkin A, Wijnberg C, Hildesheim R, Arieli A (1999) In-Vivo Optical Imaging of Cortical Architecture and Dynamics. In: U. Windhorst, H. Johansson, eds. Modern Techniques in Neuroscience research. Springer-Verlag, pp. 893–969.

Hansel D, Sompolinsky H (1998) Modeling Feature Selectivity in Local Cortical Circuits. In: C. Koch, I. Segev, eds. Methods in Neuronal Modeling: From Synapse to Networks. 2 edition MIT Press pp. 499–567.

Hertz J, Krogh A, Palmer R (1991) Introduction to the Theory of Neural Computation. Perseus Books.

Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc. Natl. Acad. Sci. USA 81: 3088–3092.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148: 574–591.

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. Nature 425: 954–956.

Kisvárday ZF, Crook JM, Buzás P, Eysel UT (2000) Combined physiological-anatomical approaches to study lateral inhibition. J. Neurosci. Methods 103: 91–106.

Kisvárday ZF, Ferecská AS, Kovács K, Buzás P, Budd JM, Eysel UT (2002) One axon-multiple functions: Specificity of lateral inhibitory connections by large basket cells. J. Neurocytol. 31: 255–264.

Kisvárday ZF, Tóth E, Rausch M, Eysel UT (1997) Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. Cereb. Cortex 7: 605–618.

Lampl I, Reichova I, Ferster D (1999) Synchronous membrane potential fluctuations in neurons of the cat visual cortex. Neuron 22: 361–374.

Löwel S, Schmidt KE, Kim DS, Wolf F, Hoffsümmer F, Singer W, Bonhoeffer T (1998) The layout of orientation and ocular dominance domains in area 17 of strabismic cats. Eur. J. Neurosci 10: 2629–2643.

Malach R, Amir Y, Harel M, Grinvald A (1993) Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. Proc. Natl. Acad. Sci. USA 90: 10469–10473.

Maldonado PE, Gödecke I, Gray CM, Bonhoeffer T (1997) Orientation selectivity in pinwheel centers in cat striate cortex. Science 276: 1551–1555.

Miyawaki Y, Okada M (2004) A network model of perceptual suppression induced by transcranial magnetic stimulation. Neural Comput 16: 309–331.

Müller T, Stetter M, Hübener M, Sengpiel F, Bonhoeffer T, Gödecke I, Chapman B, Lowel S, Obermayer K (2000) An analysis of orientation and ocular dominance patterns in the visual cortex of cats and ferrets. Neural Comput 12: 2573–2595.

Petersen CC, Grinvald A, Sakmann B (2003) Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. J. Neurosci. 23: 1298–1309.

Pouget A, Zhang K, Deneve S, Latham PE (1998) Statistically efficient estimation using population coding. Neural Comput. 10: 373–401.

Redish AD, Elga AN, Touretzky DS (1996) A coupled attractor model of the rodent head direction system. Network 7: 671–685.

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

Schummers J, Marino J, Sur M (2002) Synaptic integration by V1 neurons depends on location within the orientation map. Neuron 36: 969–978.

Sharma J, Angelucci A, Sur M (2000) Induction of visual orientation modules in auditory cortex. Nature 404: 841–847.

Sharon D, Grinvald A (2002) Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. Science 295: 512–515.

Skaggs WE, Knierim JJ, Kudrimoti HS, McNaughton BL (1995) A model of the neural basis of the rat's sense of direction. Adv. Neural. Inf. Process. Syst. 7: 173–180.

Somers DC, Nelson SB, Sur M (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neurosci. 15: 5448–5465.

Sompolinsky H, Shapley R (1997) New perspectives on the mechanisms for orientation selectivity. Curr. Opin. Neurobiol 7: 514–522.

Sterkin A, Lampl I, Ferster D, Glaser D, Grinvald A, Arieli A (1999) Exploring cortical synchronization by simultaneous intracellular recording and dye imaging in cat visual cortex. Soc. Neurosci. Abstr. 25.

Swindale NV, Grinvald A, Shmuel A (2003) The spatial pattern of response magnitude and selectivity for orientation and direction in cat visual cortex. Cereb. Cortex. 13: 225–238.

Ts'o DY, Gilbert CD, Wiesel TN (1986) Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. J. Neurosci. 6: 1160–1170.

Tsodyks M, Kenet T, Grinvald A, Arieli A (1999) Linking spontaneous activity of single cortical neurons and the underlying functional architecture. Science 286: 1943–1946.

Tsodyks M, Sejnowski T (1995) Associative memory and hippocampal place cells. Int. J. Neural. Syst. 6: 81–86.

Van Hooser SD, Heimel JA, Chung S, Nelson SB, Toth LJ (2005) Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. J. Neurosci 25: 19–28.

Volgushev M, Pernberg J, Eysel UT (2002) A novel mechanism of response selectivity of neurons in cat visual cortex. J. Physiol. 540: 307–320.

Wang G, Ding S, Yunokuchi K (2003) Difference in the representation of cardinal and oblique contours in cat visual cortex. Neurosci. Lett. 338: 77–81.

Weliky M, Kandler K, Fitzpatrick D, Katz LC (1995) Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. Neuron 15: 541–552.

Wilson HR, Cowan JD (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik 13: 55–80.

Xie X, Hahnloser RH, Seung HS (2002) Double-ring network model of the head-direction system. Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys. 66: 041902.

Yousef T, Tóth E, Rausch M, Eysel UT, Kisvárday ZF (2001) Topography of orientation centre connections in the primary visual cortex of the cat. Neuroreport 12: 1693–1699.

Yu HB, Shou TD (2000) The oblique effect revealed by optical imaging in primary visual cortex of cats. Sheng Li Xue Bao 52: 431–434.

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