SO(3) Symmetry Breaking Mechanism for Orientation and Spatial Frequency Tuning in the Visual Cortex

Paul C. Bressloff[1] and Jack D. Cowan[2]
[1]*Department of Mathematics, University of Utah, Salt Lake City, Utah 84112*
[2]*Mathematics Department, University of Chicago, Chicago, Illinois 60637*
(Received 12 October 2001; published 31 January 2002)

A dynamical model of orientation and spatial frequency tuning in a cortical hypercolumn is presented. The network topology is taken to be a sphere whose poles correspond to orientation pinwheels associated with high and low spatial frequency domains, respectively. Recurrent interactions within the sphere generate a tuned response via an SO(3) symmetry breaking mechanism.

DOI: 10.1103/PhysRevLett.88.078102 PACS numbers: 87.19.La, 05.45.−a, 11.30.Qc, 42.66.Lc

The discovery that most neurons in the visual cortex ($V1$) of cats and primates respond preferentially to locally oriented edges or bars [1] has led to many studies of the precise circuitry underlying this property. The classical model of Hubel and Wiesel proposes that the orientation preference of a cortical neuron arises primarily from the geometric alignment of the receptive fields of thalamic neurons in the lateral geniculate nucleus (LGN) projecting to it. This has been confirmed by a number of recent experiments [2,3]. However, there is also growing experimental evidence suggesting the importance of intracortical feedback. For example, the blockage of extracellular inhibition in cortex leads to considerably broader orientation tuning [4]. Moreover, intracellular measurements indicate that direct inputs from the LGN to $V1$ provide only a fraction of the total excitatory inputs relevant to orientation selectivity [5].

The possible role of recurrent cortical connections in generating orientation tuning has been analyzed using the *ring model* of a cortical hypercolumn [6–8]. This model consists of interacting neural populations labeled by their orientation preference $\phi \in [0, \pi)$. Through a combination of recurrent excitation and inhibition, an orientation tuning curve can be generated in the network via spontaneous symmetry breaking of an underlying $\mathbf{O}(2)$ symmetry around the ring. The peak of the tuning curve is then fixed by a weakly biased input from the LGN that explicitly breaks the hidden $\mathbf{O}(2)$ symmetry [6,8].

Although the ring model has been quite successful in accounting for some aspects of the response properties of cortical cells, it has a number of limitations. First, it does not take into account the two-dimensional distribution of orientation preferences within a hypercolumn as revealed by optical imaging data and microelectrode recordings [9–12]. This distribution has characteristic features illustrated in Fig. 1(a): (i) orientation preference appears to change continuously as a function of cortical location except at singularities or pinwheels, (ii) the pinwheels tend to align with the centers of ocular dominance (left/right eye) stripes, and (iii) within each pinwheel region there is a broad distribution of orientation preferences so that the (population) average orientation selectivity is weak [13]. These observations suggest an underlying spatial periodicity in the microstructure of $V1$ with a period of approximately 1 mm (in cats and primates). If we define a hypercolumn to be the fundamental domain of this periodic tiling of the cortical plane [14], then each hypercolumn contains two sets of orientation preferences $\phi \in [0, \pi)$, one for each eye, and four pinwheels. The ring model collapses this structure onto a circular set of orientation domains around a single pinwheel; see Fig. 1(a).

A second major limitation of the ring model is that it neglects the fact that $V1$ cells are also selective for spatial frequency. Indeed, there is considerable physiological and psychophysical evidence to suggest that cortical cells act like bandpass filters for both orientation and spatial frequency, so that a hypercolumn carries out a localized two-dimensional spatial frequency filtering of a stimulus rather than simply performing local edge detection (for a review see [15]). The distribution of spatial frequency preferences across cortex is less clear than that of orientation. Nevertheless, recent optical imaging studies [16,17] suggest that both orientation and spatial frequency are distributed almost continuously across cortex, spatial frequencies at the extremes of the continuum tend to be located at the orientation pinwheel singularities, and around the pinwheels iso-orientation and isofrequency contours are

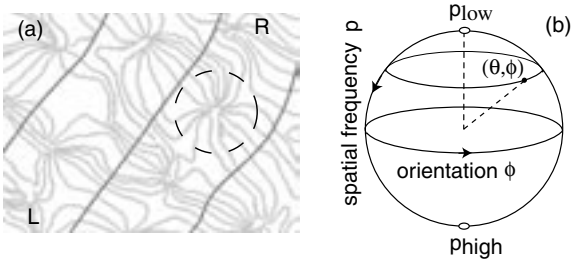

FIG. 1. (a) Iso-orientation (light) and ocular dominance (dark) contours in a small region of $V1$. Around each orientation pinwheel is a ring of orientation selective cells [13]. (b) Spherical model of orientation and spatial frequency tuning.

approximately orthogonal so that they generate a local curvilinear coordinate system. Such results provide functional evidence that hypercolumns implement localized two-dimensional spatial frequency analysis.

In this Letter we construct a minimal dynamical model of a hypercolumn that (a) includes both orientation and spatial frequency preferences, (b) incorporates the orientation pinwheels, and (c) exhibits sharply tuned responses in the presence of recurrent interactions and weakly biased LGN inputs via a symmetry breaking mechanism.

For simplicity, we restrict ourselves to a single ocular dominance column, and assume that the hypercolumn is parametrized by two cortical labels, which represent the orientation preference $\phi \in [0, \pi)$ and spatial frequency preference $p \in [0, \infty)$ of a local patch or column of cells. Given that p is not a periodic variable within a hypercolumn, one cannot extend the ring model of orientation tuning by including a second ring so that the network topology becomes a torus $S^1 \times S^1$. An important clue on how to proceed is provided by the fact that each hypercolumn (when restricted to a single ocular dominance column) contains two orientation singularities. These are assumed to correspond, respectively, to the two extremes of spatial frequency within the hypercolumn. This suggests the network topology of a sphere S^2, with the two singularities identified as the north and south poles, respectively [see Fig. 1(b)]. Introducing spherical polar coordinates $(r, \theta, 2\phi)$ with $r = 1$, $\theta \in [0, \pi)$, and $\phi \in [0, \pi)$, we set

$$\theta \equiv \mathcal{Q}(p) = \pi/[1 + (p_0/p)^{\beta}] \tag{1}$$

with p_0, β fixed. The compressive nonlinearity is required in order to represent the semi-infinite range of spatial frequencies within the bounded domain of a hypercolumn. The gain parameter β determines the effective spatial frequency bandwidth of the hypercolumn. The latter is typically around four octaves, that is, $p_{\max} \approx 2^n p_{\min}$ with $n = 4$, which corresponds to a gain of around $\beta = 1.5$.

Let $a(\theta, \phi, t)$ denote the activity of a population of cells on the sphere. Our *spherical model* of a hypercolumn is then defined according to the evolution equation

$$\frac{\partial a(\theta, \phi, t)}{\partial t} = -a(\theta, \phi, t) + h(\theta, \phi) + \int_{\mathbf{S}^2} w(\theta, \phi \mid \theta', \phi') \times g[a(\theta', \phi', t)]\, \mathcal{D}(\theta', \phi'), \tag{2}$$

where $\mathcal{D}(\theta, \phi) = 2 \sin\theta\, d\theta\, d\phi$. Here w represents the distribution of recurrent interactions within the hypercolumn, $h(\theta, \phi)$ is a weakly biased input from the LGN, and $g[a]$ is the smooth nonlinear function $g[a] = g_{\max}/(1 + e^{-\eta(a - a_{th})})$ for constant gain η and threshold a_{th}. Equation (2) is the natural extension of the ring model to the sphere. It is important to emphasize that the sphere describes the network topology of the local weight distribution expressed in terms of the internal labels for orientation and spatial frequency. It is not required to match the actual spatial arrangement of cells within a hypercolumn.

In order to extend the symmetry breaking mechanism of the ring model [6] to the spherical model (2), we first have to construct a weight distribution that is invariant with respect to the symmetry group of the sphere, namely, SO(3). This requires the use of spherical harmonics. Any sufficiently smooth function $f(\theta, \phi)$ on the sphere can be expanded in a uniformly convergent double series of spherical harmonics $f(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y_n^m(\theta, \phi)$. The functions $Y_n^m(\theta, \phi)$ constitute the angular part of the solutions of Laplace's equation in three dimensions, and thus form a complete orthonormal set. The orthogonality relation is

$$\int_{\mathbf{S}^2} Y_{n_1}^{m_1 *}(\theta, \phi) Y_{n_2}^{m_2}(\theta, \phi)\, \mathcal{D}(\theta, \phi) = \delta_{n_1, n_2} \delta_{m_1, m_2}. \tag{3}$$

The spherical harmonics are given explicitly by

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

for $n \geq 0$ and $-n \leq m \leq n$, where $P_n^m(\cos\theta)$ is an associated Legendre function. The action of SO(3) on $Y_n^m(\theta, \phi)$ involves $(2n + 1) \times (2n + 1)$ unitary matrices associated with irreducible representations of SU(2) [18]. From the unitarity of these representations, one can construct an SO(3) invariant weight distribution of the general form

$$w(\theta, \phi \mid \theta', \phi') = \mu \sum_{n=0}^{\infty} W_n \sum_{m=-n}^{n} Y_n^{m *}(\theta', \phi') Y_n^m(\theta, \phi) \tag{5}$$

with W_n real and μ a coupling parameter. As an illustration, consider the SO(3) distribution joining neurons with the same spatial frequency (same latitude on the sphere) for the particular case $W_0 < 0$, $W_1 > 0$, $W_n = 0$ for $n \geq 2$. One finds that away from the pinwheels (poles of the sphere), cells with similar orientation excite each other, whereas those with differing orientation inhibit each other. This is the standard interaction assumption of the ring model. On the other hand, around the pinwheels, all orientations uniformly excite, which is consistent with the experimental observation that local interactions depend on cortical separation [19]. That is, although the cells around a pinwheel can differ greatly in their orientation preference, they are physically close together within the hypercolumn.

We now show how sharp orientation and spatial frequency tuning can occur through spontaneous symmetry breaking of SO(3). First, substitute the distribution (5) into Eq. (2) and assume, for the moment, that there is constant external drive from the LGN [with $h(\theta, \phi) = \bar{h}$]

such that Eq. (2) has $a(\theta, \phi) = \bar{a}$ as a fixed point solution. Linearizing about the fixed point and setting $a(\theta, \phi, t) = \bar{a} + e^{\lambda t} u(\theta, \phi)$ leads to an eigenvalue equation for the linear eigenmodes $u(\theta, \phi)$:

$$\lambda u(\theta, \phi) = -u(\theta, \phi) + \mu \int_{\mathbf{S}^2} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} W_n \times Y_n^{m*}(\theta', \phi') Y_n^m(\theta, \phi) u(\theta', \phi') \mathcal{D}(\theta', \phi') \tag{6}$$

[with a factor $g'(\bar{a})$ absorbed into μ]. The orthogonality relation (3) shows that the linear eigenmodes are spherical harmonics with $\lambda = \lambda_n \equiv -1 + \mu W_n$ for $u(\theta, \phi) = Y_n^m(\theta, \phi)$, $-n \leq m \leq n$. Thus the nth eigenvalue is $(2n + 1)$-fold degenerate.

Now suppose that $W_1 > W_n$ for all $n \neq 1$. The fixed point solution $a = \bar{a}$ then destabilizes at a critical value of the coupling $\mu_c = 1/W_1$ due to excitation of the first-order spherical harmonics. Sufficiently close to the bifurcation point, the resulting activity profile can be written in the form

$$a(\theta, \phi) = \bar{a} + \sum_{m=0,\pm} c_m f_m(\theta, \phi) \tag{7}$$

for real coefficients $c_0, c_\pm$ and $f_0(\theta, \phi) = \cos\theta$, $f_+(\theta, \phi) = \sin\theta \cos(2\phi)$, $f_-(\theta, \phi) = \sin\theta \sin(2\phi)$. Amplitude equations for these coefficients can be obtained by carrying out a perturbation expansion of Eq. (2) with respect to the small parameter $\varepsilon = \mu - \mu_c$ using the method of multiple scales [20]. This leads to the Stuart-Landau equations [8,21]

$$\frac{dc_k}{dt} = c_k \left(\mu - \mu_c + \Lambda \sum_{m=0,\pm} c_m^2 \right), \qquad k = 0, \pm, \tag{8}$$

where

$$\Lambda = \frac{3g_3}{5} + \frac{2g_2^2}{3} \frac{\mu_c W_0}{1 - \mu_c W_0} + \frac{8g_2^2}{15} \frac{\mu_c W_2}{1 - \mu_c W_2}. \tag{9}$$

Here g_2, g_3 are coefficients in the Taylor expansion of the firing rate function, $[g(a) - g(\bar{a})]/g'(\bar{a}) = [(a - \bar{a}) + g_2(a - \bar{a})^2 + g_3(a - \bar{a})^3 + \ldots]$. It is clear that the amplitude equations are equivariant with respect to the action of the orthogonal group SO(3) on (c_0, c_+, c_-), which reflects the underlying spherical symmetry. Moreover, defining $R = \sum_{m=0,\pm} c_m^2$ we see that

$$\frac{dR}{dt} = 2R(\mu - \mu_c + \Lambda R), \tag{10}$$

which has a stable fixed point at $R_0 = (\mu - \mu_c)/|\Lambda|$, provided that $\Lambda < 0$. This corresponds to an SO(3)-invariant submanifold of marginally stable states.

Equation (7) represents a *tuning surface* for orientation and spatial frequency with a solitary peak whose location is determined by the values of the coefficients (c_0, c_+, c_-). Such a solution spontaneously breaks the underlying SO(3) symmetry. However, full spherical symmetry is recovered by noting that rotation of the solution corresponds to an orthogonal transformation of the coefficients $c_0, c_\pm$. Thus the action of SO(3) is to shift the location of the peak of the activity profile on the sphere, that is, to change the particular orientation and spatial frequency selected by the tuning surface.

Now suppose that there exists a weakly biased, time-independent input from the LGN of the form

$$h(\theta, \phi) = \bar{h} + \sum_{m=0,\pm} h_m f_m(\theta, \phi), \tag{11}$$

where $h_0 = \Gamma \cos\Theta$, $h_+ = \Gamma \sin\Theta \cos 2\Phi$, $h_- = \Gamma \times \sin\Theta \sin 2\Phi$ where Γ is the effective contrast of the input stimulus. Equation (11) describes a unimodal function on the sphere with a single peak at $\{\Theta, \Phi\}$, which corresponds to an input orientation $\phi_S = \Phi$ and an input spatial frequency $p_S = \mathcal{Q}^{-1}(\Theta)$ with $\mathcal{Q}$ defined according to Eq. (1). Here we ignore higher-order spherical harmonic contributions to the LGN input, since the pattern forming instability amplifies only the first-order harmonic components—it is these components that couple to the cubic amplitude equation. If $\Gamma = \mathcal{O}(\epsilon^{3/2})$ then the input (11) generates an additional contribution h_k to the right-hand side of the cubic amplitude equation that explicitly breaks the hidden SO(3) symmetry and fixes the peak of the tuning surface at $\{\Theta, \Phi\}$. Since $h_k = \mathcal{O}(\varepsilon^{3/2})$, whereas $c_k = \mathcal{O}(\varepsilon^{1/2})$, we see that the cortical model acts as an amplifier for the first spherical harmonic components of the weakly biased input from the LGN.

In order to confirm the above analytical results, we solved Eq. (2) numerically, using the discretization scheme considered by Varea *et al.* [22] in their study of pattern formation for a reaction-diffusion system on a sphere. In Fig. 2 we plot the relative firing rate $g(a)/g_{\max}$ in response to a weakly biased input from the LGN, Eq. (11), with $\Theta = \pi/2$ (corresponding to an intermediate input frequency $p_S = 2c/\text{deg}$) and $\Phi = \pi/2$. Figure 2(a) shows a surface plot in the $\{p, \phi\}$ plane for $\Gamma = 0.1$. It can be seen that the hypercolumn exhibits a tuning surface that is localized with respect to two-dimensional spatial frequency and its peak is locked to the LGN input at $p = 2c/\text{deg}$, $\phi = \pi/2$. In Fig. 2(b) we plot the response as a function of spatial frequency at the optimal orientation for various input amplitudes Γ. The height of the spatial frequency tuning curves increases with the input amplitude Γ, but the width at half-height is approximately the same (as can be checked by rescaling the tuning curves to the same height). Since Γ increases with the contrast of a stimulus, this shows that the network naturally exhibits contrast invariance. Corresponding orientation tuning curves are shown in Fig. 2(c), and are also found to exhibit contrast invariance.

Note that projecting the spherical tuning surface onto the $\{p, \phi\}$ plane breaks the underlying **SO**(3) symmetry of the sphere. Consequently, the shape of the planar tuning surface is not invariant under shifts in the location of

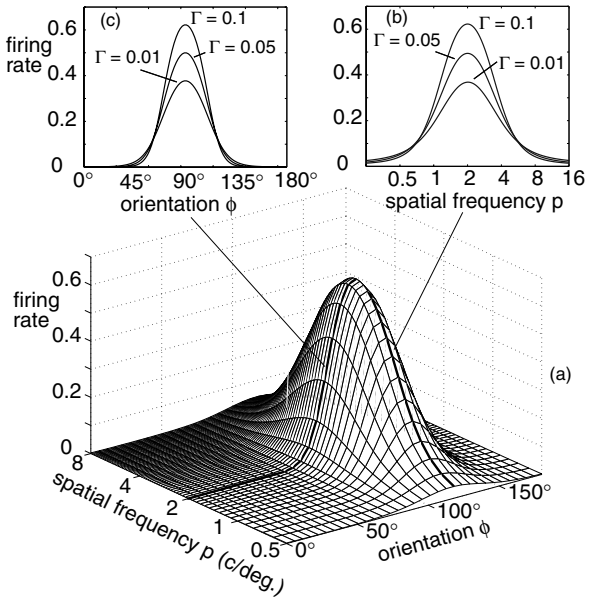

FIG. 2. Plot of normalized firing rate $g(a)/g_{\max}$ in response to a weakly biased input from the LGN with $\Theta = \pi/2$, $\Phi = \pi/2$, and $\Gamma \ll 1$. The firing rate function has a gain $\eta = 5$, a threshold $a_{th} = 0.6$, and we take $\bar{a} = 0$. The weight coefficients are $W_0 = -2$, $W_1 = 1$, $W_n = 0$, $n > 1$, and $g_{\max}\mu = 5$. (a) Tuning surface in the $\{p, \phi\}$ plane. (b) and (c) Spatial frequency and orientation tuning curves.

the peak of the tuning surface. Such distortions generate behavior that is consistent with recent experimental observations. First, at low and high spatial frequencies (i.e., towards the pinwheels) there is a broadening of the tuned response to orientation, as found in [13]. Second, there is a systematic shift in the peak of spatial frequency tuning curves at nonoptimal orientations that is towards the closest pinwheel. Interestingly, there is some suggestion of spatial frequency shifts in recent optical imaging data [17].

In conclusion, representing the topology of a hypercolumn (with respect to orientation and spatial frequency labels) as a sphere is a natural way to accommodate the orientation pinwheels while providing a recurrent mechanism for generating two-dimensional spatial frequency tuning. One issue that we have not addressed here is how the projection of the LGN input on to its first spherical harmonic components encodes information regarding properties of a visual stimulus. Intriguingly, examination of the response to a sinusoidal grating stimulus indicates that if such a stimulus is first filtered by the action of the feedforward pathway from retina to cortex, before the recurrent dynamics amplifies its first-order spherical harmonic components, then the representation of spatial frequency is not faithful. That is, there is a mismatch between the spatial frequency at the peak of the tuning surface and the stimulus frequency. We expect a similar conclusion to hold for any recurrent mechanism that amplifies harmonic components of two-dimensional stimuli. One possible mechanism for correcting this mismatch is via the massive feedback pathways from $V1$ back to LGN [21].

[1] D. Hubel and T. Wiesel, J. Neurosci. **3**, 1116 (1962).
[2] R. C. Reid and J. M. Alonso, Nature (London) **378**, 281 (1995).
[3] D. Ferster, S. Chung, and H. Wheat, Nature (London) **380**, 249 (1997).
[4] S. Nelson, L. Toth, B. Seth, and M. Sur, Science **265**, 774 (1994).
[5] R. J. Douglas, C. Koch, M. Mahowald, K. Martin, and H. Suarez, Science **269**, 981 (1995).
[6] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **92**, 3844 (1995).
[7] D. Somers, S. Nelson, and M. Sur, J. Neurosci. **15**, 5448 (1995).
[8] P. C. Bressloff, N. W. Bressloff, and J. D. Cowan, Neural Comput. **12**, 2473 (2000).
[9] G. G. Blasdel and G. Salama, Nature (London) **321**, 579 (1986).
[10] T. Bonhoeffer and A. Grinvald, Nature (London) **364**, 166 (1991).
[11] K. Obermayer and G. G. Blasdel, J. Neurosci. **13**, 4114 (1993).
[12] G. G. Blasdel (personal communication).
[13] P. E. Maldonado and C. M. Gray, Visual Neuroscience **13**, 509 (1996).
[14] D. H. Hubel and T. N. Wiesel, J. Comp. Neurol. **158**, 295 (1974).
[15] R. L. De Valois and K. K. De Valois, *Spatial Vision* (Oxford University Press, Oxford, 1988).
[16] M. Hubener, D. Shoham, A. Grinvald, and T. Bonhoeffer, J. Neurosci. **15**, 9270 (1997).
[17] N. P. Issa, C. Trepel, and M. P. Stryker, J. Neurosci. **20**, 8504 (2000).
[18] G. Arfken, *Mathematical Methods for Physicists* (Academic Press, San Diego, 1985), 3rd ed.
[19] A. Das and C. Gilbert, Nature (London) **399**, 655 (1999).
[20] Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence* (Springer, New York, 1984).
[21] P. C. Bressloff and J. D. Cowan (to be published).
[22] C. Varea, J. L. Aragon, and R. A. Barrio, Phys. Rev. E **60**, 4588 (1999).