Phil. Trans. Roy. Soc. B, In Press

A spherical model for orientation and spatial frequency tuning in a cortical hypercolumn

Paul C. Bressloff[1] and Jack D. Cowan[2]‡

[1]*Department of Mathematics, University of Utah*
Salt Lake City, Utah 84112
(bressloff@math.utah.edu)
[2]*Mathematics Department, University of Chicago, Chicago Il. 60637*
(cowan@math.uchicago.edu)

‡Correspondence should be sent to:
Professor Jack D. Cowan
Department of Mathematics, University of Chicago
5734 S. University Avenue, Chicago IL 60637
Tel: +1-(1)773-702-1076; Fax: +1-(1)773-702-9787
E-mail: cowan@math.uchicago.edu

[Key words: orientation, spatial frequency, hypercolumn, neural modelling, cortico–geniculate feedback]

Contents

Summary

A theory is presented of the way in which hypercolumns in primary visual cortex (V1) are organized to detect important features of visual images, namely local orientation and spatial frequency. Given the existence in V1 of dual maps for these features, both organized around orientation pinwheels, we construct a model of a hypercolumn in which orientation and spatial frequency preferences are represented by the two angular coordinates of a sphere. The two poles of this sphere are taken to correspond, respectively, to high and low spatial frequency preferences.

In Part I of the paper we use mean–field methods to derive exact solutions for localized activity states on the sphere. We show how cortical amplification through recurrent interactions generates a sharply tuned, contrast–invariant population response to both local orientation and local spatial frequency, even in the case of a weakly biased input from the lateral geniculate nucleus (LGN). A major prediction of our model is that this response is non–separable with respect to the local orientation and spatial frequency of a stimulus. That is, orientation tuning is weaker around the pinwheels, and there is shift in spatial frequency tuning towards that of the closest pinwheel at non–optimal orientations.

In Part II of the paper we show that a simple feedforward model of spatial frequency preference, unlike that for orientation preference, does not generate a faithful representation when amplified by recurrent interactions in V1. We then introduce the idea that cortico–geniculate feedback modulates LGN activity to generate a faithful representation, thus providing a new functional interpretation of the role of this feedback pathway. Using linear filter theory we show that if the feedback from a cortical cell is taken to be approximately equal to the reciprocal of the corresponding feedforward receptive field (in the two–dimensional Fourier domain), then the mismatch between the feedforward and cortical frequency representations is eliminated. We therefore predict that cortico–geniculate feedback connections innervate the LGN in a pattern determined by the orientation and spatial frequency biases of feedforward receptive fields. Finally, we show how recurrent cortical interactions can generate cross–orientation suppression.

1 Introduction

A prominent feature of the functional architecture of visual cortex (V1) is the existence of an orderly retinotopic mapping of the visual field onto its surface, with left and right halves of the visual field mapped onto left and right V1 respectively. Superimposed upon this are additional maps reflecting the fact that neurons respond preferentially to stimuli with particular features such as orientation and ocularity (Hubel & Wiesel, 1977; Obermayer & Blasdel, 1993; Swindale, 1996). Maps of both ocularity and orientation preference have been well characterized in cat and monkey, via microelectrode recording (Hubel & Wiesel, 1962, 1968, 1977), autoradiographic studies using proline (Wiesel, Hubel, & Lam, 1974) or 2–deoxyglucose (2–DG) (Hubel, Wiesel, & Stryker, 1978), and optical imaging (Blasdel & Salama, 1986; Bonhoeffer & Grinvald, 1991; Blasdel, 1992). The topography revealed

Figure 1: Iso–orientation (light) and ocular dominance (dark) contours in a small region of Macaque V1. [Redrawn from Blasdel (1992)].

by these methods has a number of characteristic features (Obermayer & Blasdel, 1993). (i) Orientation preference changes continuously as a function of cortical location, except at singularities or *pinwheels*. (ii) There exist *linear zones*, approximately 750×750 μm^2 in area (in Macaque), bounded by pinwheels, within which iso–orientation regions form parallel slabs. (iii) linear zones tend to cross the borders of ocular dominance stripes at right angles; pinwheels tend to align with the centers of ocular dominance stripes. All these features can be seen in the optical image shown in figure 1.

These observations suggest that the microstructure of V1 is spatially periodic with a period of approximately 1 mm (in primates). The fundamental domain of this tiling of the cortical plane is the *hypercolumn* (Hubel & Wiesel, 1974), which contains the full range of orientation preferences $\phi \in [0, \pi)$ organized around pinwheels, with one set of preferences for each ocular dominance column. The identification of the hypercolumn as a basic cortical module is still somewhat controversial (LeVay & Nelson, 1991). However, it has proved a very useful conceptual tool in the development of large scale dynamical models of cortical function. In its original form the hypercolumn was organized in terms of linear zones of orientation preference slabs and ocular dominance columns, as shown in figure 2a. This was later modified to include the cytochrome oxidase (CO) *blobs* observed in Macaque by Horton & Hubel (1981) (see figure 2b) and only later found in Cat (Murphy,

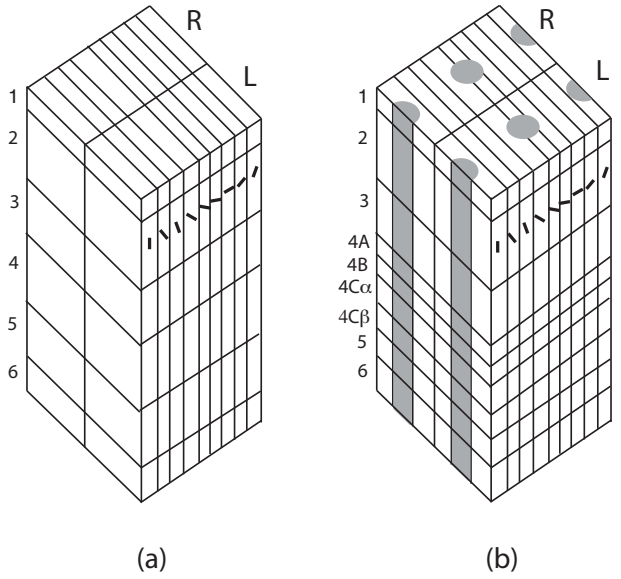

Figure 2: (a) Hubel and Wiesel's original *Icecube* model of a V1 hypercolumn, redrawn for Cat; (b) The Icecube model with CO blobs for Macaque V1.

Jones, & Sluyters, 1995). The blobs are regions of cells that are more metabolically active and hence richer in their levels of CO. They tend to be located in the centers of ocular dominance stripes and have a strong association with about half the orientation singularities.

The fact that orientation preference is a periodic quantity suggests that the internal structure of a hypercolumn can be idealized as a *ring* of orientation selective wedges or patches. In the last decade a number of network models have appeared based on such an idealization (Somers, Nelson, & Sur, 1995; Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Vidyasagar, Pei, & Volgushev, 1996; Mundel, Dimitrov, & Cowan, 1997; Ben-Yishai, Hansel, & Sompolinsky, 1997; Somers, Todorov, Siapas, Toth, Kim, & Sur, 1998; Li, 1999; Dragoi & Sur, 2000; Stetter, Bartsch, & Obermayer, 2000; Bressloff, Bressloff, & Cowan, 2000; Bressloff & Cowan, 2002a). These models have been used to investigate the role of intra-cortical interactions in orientation selectivity and tuning. The classical model of Hubel & Wiesel (1962) 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 (Reid & Alonso, 1995; Ferster, Chung, & Wheat, 1997). However,

there is also growing experimental evidence suggesting the importance of intra–cortical feedback for orientation tuning. For example, the blockage of extracellular inhibition in cortex leads to considerably broader tuning (Sillito, 1975; Nelson, Toth, Seth, & Sur, 1994). Moreover, intracellular measurements indicate that direct inputs from the LGN to neurons in layer 4 of the visual cortex provide only a fraction of the total excitatory inputs relevant to orientation selectivity (Douglas, Koch, Mahowald, Martin, & Suarez, 1995). A number of modeling studies have shown how local recurrent interactions within an isolated cortical hypercolumn (idealized as a ring network) can amplify certain Fourier components of network activity leading to sharp orientation tuning curves, even when the LGN inputs are weakly biased (Somers *et al.*, 1995; Ben-Yishai *et al.*, 1995, 1997; Bressloff *et al.*, 2000). Such an amplification mechanism provides one possible explanation for the approximate contrast invariance of the tuned response. More large-scale models of cortex based on a system of coupled ring networks have subsequently been used to investigate how orientation tuning is modulated by long-range interactions between hypercolumns (Mundel *et al.*, 1997; Somers *et al.*, 1998; Li, 1999; Dragoi & Sur, 2000; Stetter *et al.*, 2000; Bressloff & Cowan, 2002a)

Although ring models have been quite successful in accounting for some aspects of the response properties of hypercolumns, they have a number of limitations. For example, they do not take into account the two dimensional structure illustrated in figure 1, in which iso–orientation pinwheels alternate with linear zones, nor the presence of ocular dominance columns. More significantly for our interest, they also neglect the *spatial frequency* selectivity of V1 neurons. Such selectivity has been observed in many physiological experiments. Recordings from cat and monkey striate cortex have established that a large number of cells are narrowly tuned to spatial frequency. Figure 3, for example, shows the responses of a number of macaque monkey V1 cells to oriented gratings. The

Figure 3: Spatial frequency and orientation selectivity of cells in macaque V1. The thresholded response of a number of cells is plotted as a function of stimulus spatial frequency and orientation. The results are shown in log-polar coordinates with orientation given by the polar angle and spatial frequency by the radius (on a logarithmic scale). [Redrawn from De Valois *et al.* (1982)]

average bandwidth is between 1 and 2 octaves, which covers a small fraction of the total range of spatial frequencies (around 6 to 8 octaves in the fovea) to which the macaque is sensitive (De Valois & De Valois, 1988). As in the case of psychophysical studies (Kelly &

Magnuski, 1975), two dimensional stimuli such as checkerboards provide strong evidence that neurons are tuned to two dimensional spatial frequencies. In fact there is considerable physiological evidence to suggest that cortical neurons act like bandpass filters for both orientation and spatial frequency, so that a hypercolumn implements a localized or windowed two dimensional spatial frequency filtering of a stimulus rather than simply performing local edge detection (Webster & De Valois, 1985; Jones & Palmer, 1987).

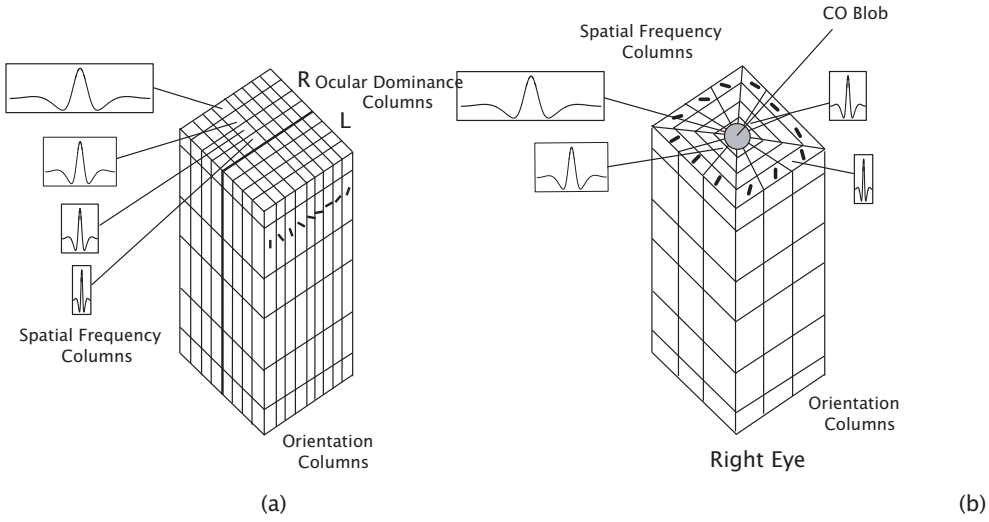

Figure 4: (a) De Valois and De Valois' modified icecube model of a Cat V1 hypercolumn. (b) The modified icecube model with CO blobs for Macaque V1. Redrawn from De Valois & De Valois (1988).

The distribution of spatial frequency preference across cortex is less clear than that of orientation preference. Nevertheless, based on the 2–DG studies available at the time [see Tootell, Silverman, & De Valois (1981)], De Valois & De Valois (1988) introduced the models of V1 hypercolumns shown in figure 4. In the macaque it was found that the CO blob regions were sites of cells that responded preferentially to low spatial frequencies, which suggested that spatial frequency increased radially away from the blobs. This picture has been extended by recent optical studies of the spatial frequency map in cat (Bonhoeffer, Kim, Malonek, Shoham, & Grinvald, 1995; Hübener, Shoham, Grinvald, & Bonhoeffer, 1997; Issa, Trepel, & Stryker, 2000) which indicate that (a) both orientation and spatial frequency preferences are distributed almost continuously across cortex, (b) spatial frequency preferences at the both extremes of the continuum tend to be located at

orientation pinwheels (i.e. the pinwheels that do not coincide with CO blobs correspond to regions of high spatial frequency), and (c) around the pinwheels iso–orientation and iso–frequency preference contours are approximately orthogonal (See figure 5). Note that in most local neighborhoods of the region of V1 shown in figure 5 one can identify a low and a high spatial frequency pinwheel connected by a linear zone. There are also a few cases in which high spatial frequency pinwheels are connected by linear zones. However they tend to be sited in different ocular dominance columns.

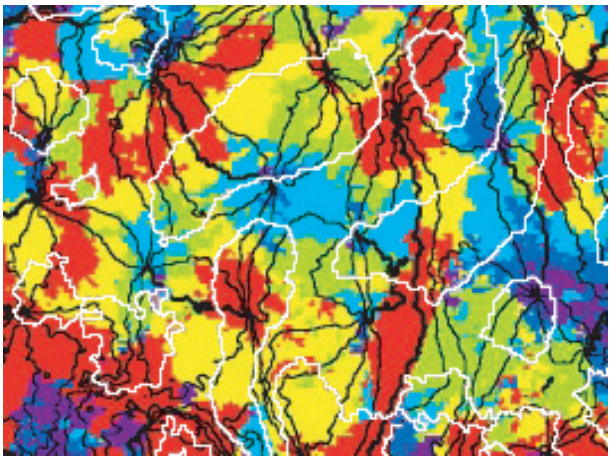

Figure 5: Map of iso–orientation preference contours (black lines), ocular dominance boundaries (white lines), and spatial frequency preferences of cells in cat V1. Redrawn from Issa *et al.* (2000). Red regions correspond to low spatial frequency preference, violet to high.

Motivated by such considerations, we introduce a minimal model of a hypercolumn that (i) includes both orientation and spatial frequency preferences, (ii) incorporates the orientation preference pinwheels, and (iii) exhibits sharply tuned responses in the presence of recurrent interactions and weakly biased LGN inputs. For simplicity we restrict ourselves to a singular ocular dominace column and a single cortical layer. In the ring model of orientation tuning the synaptic weights are taken to depend on the difference between the orientation preference of pre– and post–synaptic neurons, which naturally leads to a ring or circular network topology. Given that spatial frequency is not a periodic variable within a hypercolumn, we cannot extend the ring model by including a second ring so that the network topology becomes a torus. The simplest choice is to take the topology to be a cylinder as shown in figure 6. This leads to a network response that is separable with respect to the two stimulus features. However, recent experimental results

suggest that although separability appears to hold in the linear zones of the orientation map, there is significant non–separability close to the orientation pinwheels (Maldonado, Gödecke, Gray, & Bonhoeffer, 1997; Issa *et al.*, 2000; Mazer, Vinje, McDermott, Schiller, & Gallant, 2002). Combining this with the assumption that each hypercolumn typically contains two orientation pinwheels per ocular dominance column, and that these correspond respectively to the two extremes of spatial frequency within the hypercolumn, we introduce the network topology of a sphere to model a hypercolumn, with its two pinwheels identified as the north and south poles respectively, see figure 7.

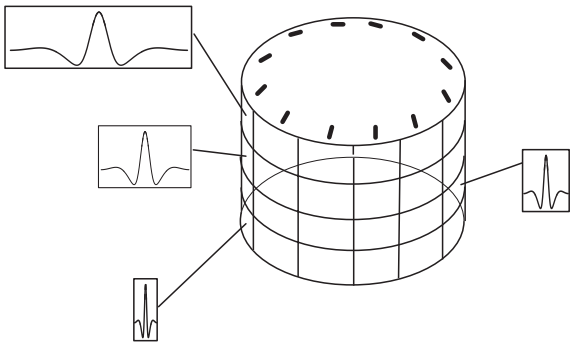

Figure 6: A cylindrical network topology. Spatial frequency preference decreases from top to bottom whereas orientation preference varies around the circumference of the cylinder.

Figure 7: A spherical network topology. High and low spatial frequency pinwheels are located at the poles of the sphere.

It is important to distinguish between the network topology shown in figures 6 or 7, which deal with synaptic weights as a function of orientation and spatial frequency preference labels, and the actual two dimensional spatial arrangement of neurons within a single cortical layer [see figure 4]. As in the ring model, the spherical model of a hypercolumn is an abstraction from a complicated set of experimental results such as those presented in figures 1 and 5. The model does not account for all of the details apparent in these figures. In fact it should also be noted that optical imaging data is inherently noisy so that some of the conclusions regarding the spatial frequency map and the nature of orientation pinwheels are still quite controversial. Nevertheless, we believe

that the analysis of conceptual models such as the one presented in this paper can lead to insights into the true nature of the action of V1.

Part I

Mean–field theory

In Part I of this paper, we present a dynamical theory of orientation and spatial frequency tuning in a cortical hypercolumn whose network topology is taken to be spherical. As we have already indicated in the introduction, this topology naturally accommodates the two orientation preference pinwheels (within a single ocular dominance column), which are located at the poles of the sphere, as well as the two dimensional curvilinear coordinate system we choose to represent orientation and spatial frequency preferences within a hypercolumn. Explicit solutions for localized activity states on the sphere are obtained using a mean–field approach (Ben-Yishai *et al.*, 1995; Hansel & Sompolinsky, 1997). We thus show how cortical amplification through recurrent interactions generates a sharply tuned, contrast invariant population response to both orientation and spatial frequency. A major prediction of our model is that this response is non–separable with respect to these stimulus features due to the presence of the pinwheels[1].

2 Details of the spherical model

We assume that a hypercolumn is parameterized by two cortical labels, which represent the orientation preference $\phi \in [0, \pi)$ and spatial frequency preference $p \in [p_{min}, p_{max}]$ of a local patch or column of cells. Typically, the bandwidth of a hypercolumn is between three and four octaves, that is, $p_{max} \approx 2^n p_{min}$ with $n = 4$. This is consistent with the observations of Hubel & Wiesel (1974), who found a two octave scatter of receptive field sizes at each cortical region they mapped. Motivated by the optical imaging data described in the introduction, we assume that the network topology is a sphere S^2 with the two pinwheels identified as the north and south poles respectively, see figure 7. If we take (θ, ϕ) to be the angular coordinates on the sphere with $\theta \in [0, \pi)$, $\phi \in [0, \pi)$ then θ

[1] A preliminary version of the spherical model has been reported briefly elsewhere (Bressloff & Cowan, 2002b). In particular, we used a perturbative amplitude equation approach to establish the basic principle of cortical amplification via spontaneous symmetry breaking. However, our analysis was restricted to the weakly nonlinear regime. Here we greatly extend the analysis using the mean–field approach

Figure 8: Spherical network topology. Orientation and spatial frequency labels are denoted by (ϕ, p) with $0 \leq \phi < \pi$ and $p_{min} \leq p \leq p_{max}$.

determines the spatial frequency preference p according to

$$\theta \equiv \mathcal{Q}(p) = \pi \frac{log(p/p_{min})}{log(p_{max}/p_{min})} \tag{2.1}$$

That is, θ varies linearly with $\log p$. This is consistent with experimental data that suggests a linear variation of $\log p$ with cortical separation (Issa *et al.*, 2000). This leads to the spherical coordinate system shown in figure 8.

Let $a(\theta, \phi, t)$ denote the activity of a local population of cells on the sphere with angular coordinates (θ, ϕ). The evolution equation for the state $a(\theta, \phi, t)$ is taken to be of the form

$$\frac{\partial a(\theta, \phi, t)}{\partial t} = -a(\theta, \phi, t) + [I(\theta, \phi, t) - \kappa]_+ \tag{2.2}$$

where κ is a threshold and $I(\theta, \phi, t)$ is the total synaptic current,

$$I(\theta, \phi, t) = \int_{\mathbf{S}^2} w(\theta, \phi | \theta', \phi') a(\theta', \phi', t) \mathcal{D}(\theta', \phi') + h(\theta, \phi) \tag{2.3}$$

with $\mathcal{D}(\theta, \phi) = \sin\theta d\theta d\phi / 2\pi$ the integration measure on the sphere. Here w represents the distribution of recurrent interactions within the hypercolumn and $h(\theta, \phi)$ is a weakly biased input from the LGN. Equation (2.2) is the natural extension of the activity based ring model of orientation tuning considered by Ben-Yishai *et al.* (1995, 1997). In order to generalize the amplification mechanism of the ring model to the spherical model, equation (2.2), we first construct a weight distribution that is invariant with respect to coordinate rotations of the sphere, that is, the symmetry group $SO(3)$. This rotational symmetry, which generalizes the $O(2)$ circular symmetry of the ring model, implies that the pattern of connections within the hypercolumn depends only on the relative distance of cells on the sphere as determined by their angular separation along geodesics or great circles. That is, given two points on the sphere (θ, ϕ) and (θ', ϕ') their angular separation α is (see figure 8)

$$\cos\alpha = \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(2[\phi - \phi']) \tag{2.4}$$

This suggest that the simplest non–trivial form for the weight distribution w is

$$w(\theta, \phi | \theta', \phi') = W_0 + W_1 \left(\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(2[\phi - \phi'])\right) \tag{2.5}$$

In figure 9 we plot w as a function of (θ, ϕ) for $\theta' = \theta$, $\phi' = 0$ and $W_1 > W_0$. It can be seen that away from the pinwheels (poles of the sphere at $\theta = 0, \pi$), 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 (Somers *et al.*, 1995; Ben-Yishai *et al.*, 1995), which has recently received experimental support (Roerig & Chen, 2002). On the other hand, around the pinwheels, all orientations uniformly excite, which is consistent with the fact that although the cells around a pinwheel can differ greatly in their orientation preference, they are physically close together within the hypercolumn.

Figure 9: Two-dimensional plot of $w(\theta, \phi|\theta', \phi')$ given by the $\mathbf{SO}(3)$ invariant weight distribution (2.9) with $W_0 = -1$, $W_1 = 1$ and $W_n = 0$ for $n \geq 2$. We set $\phi' = 0$, $\theta' = \theta$ and plot w as a function of θ and ϕ. (a) Contour plot of w on the sphere with light and dark regions correspond to excitation and inhibition respectively. (b) Surface plot of w in the (θ, ϕ)–plane.

It is possible to construct a more general form of $SO(3)$-invariant weight distribution using *spherical harmonics*. Any sufficiently smooth function $a(\theta, \phi)$ on the sphere can be expanded in a uniformly convergent double series of spherical harmonics

$$a(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm} Y_n^m(\theta, \phi) \tag{2.6}$$

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

$$\int_{\mathbf{S}^2} Y_{n_1}^{m_1 *}(\theta, \phi) Y_{n_2}^{m_2}(\theta, \phi) \mathcal{D}(\theta, \phi) = \frac{1}{4\pi} \delta_{n_1, n_2} \delta_{m_1, m_2} \tag{2.7}$$

The spherical harmonics are given explicitly by

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) \mathrm{e}^{2im\phi} \tag{2.8}$$

for $n \geq 0$ and $-n \leq m \leq n$, where $P_n^m(\cos\theta)$ is an associated Legendre function. (Note that we have adjusted the definition of the spherical harmonics to take into account the fact ϕ takes values between 0 and π). The action of $SO(3)$ on $Y_n^m(\theta, \phi)$ involves $(2n+1) \times (2n+1)$ unitary matrices associated with irreducible representations of $SU(2)$

(Arfken, 1985). From the unitarity of these representations, one can construct an $SO(3)$ invariant weight distribution of the general form

$$w(\theta,\phi|\theta',\phi') = 4\pi \sum_{n=0}^{\infty} W_n \sum_{m=-n}^{n} Y_n^{m*}(\theta',\phi') Y_n^m(\theta,\phi) \tag{2.9}$$

with W_n real. For simplicity, we shall neglect higher harmonic constributions to w by setting $W_n = 0$ for $n \geq 2$ so that equation (2.9) reduces to equation (2.5) on rescaling W_1.

Finally, the weakly biased LGN input $h(\theta,\phi)$ is assumed to be of the form

$$h(\theta,\phi) = C\left[1 - \epsilon + \epsilon\left(\cos\Theta\cos\theta + \sin\Theta\sin\theta\cos(2[\phi-\Phi])\right)\right] \tag{2.10}$$

This represents a unimodal function on the sphere with a single peak at (Θ,Φ). Here C is the effective contrast of the input and ϵ measures the degree of bias. In fact, equation (2.10) is the projection of the feedforward input from the LGN onto the zeroth and first order spherical harmonics. The *a posteriori* justification for this is based on the idea that recurrent interactions within the hypercolumn amplify these particular components of the feedforward input so that higher order harmonics can be neglected (Bressloff & Cowan, 2002b). We also note that recent optical imaging experiments provide strong support for the role of recurrent interactions in cortical amplification (Sharon & Grinvald, 2002). Rectification arising from the firing rate characteristics of cortical cells then leads to a sharply tuned, contrast invariant response to both orientation and spatial frequency (see §3). The peak response, which is located at (Θ,Φ), is assumed to faithfully encode the spatial frequency p_s and orientation ϕ_s of an external visual stimulus, that is, $\Theta = \mathcal{Q}(p_s)$ and $\Phi = \phi_s$. However, as we discuss in Part II, the relationship between Θ and p_s is far from straightforward. The transformation from visual stimulus to cortical input is typically described in terms of a convolution with respect to a feedforward receptive field modeled, for example, as a difference of Gaussians (Hawken & Parker, 1987). If the low order spherical harmonic components of the resulting feedforward input to a hypercolumn are now amplified, one finds that the cortical spatial frequency is shifted relative to the stimulus frequency—there is no corresponding shift in orientation. In other words, the network does not faithfully encode the stimulus spatial frequency unless an additional filtering operation is introduced. We suggest in Part II that feedback from V1 back to LGN (Murphy, Duckett, & Sillito, 1999) can modulate LGN activity to produce a faithful encoding of spatial frequency. However, we ignore these subtleties here and proceed with the form of LGN input given by equation (2.10).

3 Stationary localized states

It is convenient to introduce real versions of the first-order harmonics,

$$f_0(\theta,\phi) = \cos\theta,\ \ f_+(\theta,\phi) = \sin\theta\cos 2\phi,\ \ f_-(\theta,\phi) = \sin\theta\sin 2\phi \tag{3.1}$$

so that equations (2.5) and (2.10) can be rewritten in the form

$$w(\theta,\phi|\theta',\phi') = W_0 + W_1 \sum_{m=0,\pm} f_m(\theta,\phi) f_m(\theta',\phi') \tag{3.2}$$

and

$$h(\theta,\phi) = C\left[1-\epsilon+\epsilon\sum_{m=0,\pm} f_m(\Theta,\Phi) f_m(\theta,\phi)\right] \tag{3.3}$$

with $\sum_{m=0,\pm} f_m(\theta,\phi) f_m(\theta',\phi')$ equal to the angular separation of (θ,ϕ) from (θ',ϕ'). Substituting equations (2.3), (3.2), and (3.3) into the evolution equation (2.2) then gives

$$\frac{\partial a(\theta,\phi,t)}{\partial t} = -a(\theta,\phi,t) + \left[I_0(t) + \sum_{m=0,\pm} I_1^m(t) f_m(\theta,\phi)\right]_+ \tag{3.4}$$

where

$$I_0(t) = C(1-\epsilon) + W_0 R_0(t) - \kappa \tag{3.5}$$

$$I_1^m(t) = C\epsilon f_m(\Theta,\Phi) + W_1 R_1^m(t) \tag{3.6}$$

and R_0, R_1^m are the order parameters

$$R_0(t) = \int_{\mathbf{S}^2} a(\theta,\phi,t)\mathcal{D}(\theta,\phi) \tag{3.7}$$

$$R_1^m(t) = \int_{\mathbf{S}^2} a(\theta,\phi,t) f_m(\theta,\phi)\mathcal{D}(\theta,\phi) \tag{3.8}$$

Following along similar lines to the analysis of the ring model (Ben-Yishai *et al.*, 1995; Hansel & Sompolinsky, 1997), we study fixed point solutions of equation (3.4) in which the activity surface is centered at the peak of the LGN input (Θ,Φ). That is,

$$a(\theta,\phi) = \left[I_0 + \sum_{m=0,\pm} I_1^m f_m(\theta,\phi)\right]_+ \tag{3.9}$$

Such a solution is self–consistent provided that at the fixed point $R_1^m = R_1 f_m(\Theta, \Phi)$ for some R_1. Given such a fixed point solution, we define the network gain G to be the ratio between the maximal activity and the contrast relative to threshold

$$G = \frac{a(\Theta, \Phi)}{C - \kappa}. \tag{3.10}$$

It is useful to distinguish between broad and narrow activity profiles $a(\theta, \phi)$. We say that the profile is broad when all the cells are above threshold. That is, $I(\theta, \phi) \geq \kappa$ and hence $a(\theta, \phi) > 0$ for all $(\theta, \phi) \in \mathbf{S}^2$. On the other hand, a narrow profile is one for which $a(\theta, \phi)$ is only non–zero over a subdomain $\Sigma = \{\theta, \phi | 0 \leq \theta < \theta_0(\phi), 0 \leq \phi < \pi\} \subset \mathbf{S}^2$—this is what we mean by a *localized state*. The closed curve $\theta = \theta_0(\phi)$ determines the boundary of the localized state on the sphere. Note that although the two dimensional activity profile on the sphere is localized, it is not necessary that the resulting orientation tuning curves are themselves localized (see §4).

3.1 Broad activity profile

The analysis of a broad activity profile is relatively straightforward, since the fixed point equation (3.9) reduces to

$$a(\theta, \phi) = I_0 + \sum_{m=0,\pm} I_1^m f_m(\theta, \phi) \tag{3.11}$$

which can be substituted into equations (3.7) and (3.8) to give $R_0 = I_0$ and $R_1^m = I_1^m/3$. It follows from equations (3.5) and (3.6) that

$$R_0 = \frac{C(1-\epsilon) - \kappa}{1 - W_0}, \quad R_1^m = R_1 f_m(\Theta, \Phi), \quad R_1 = \frac{C\epsilon/3}{1 - W_1/3} \tag{3.12}$$

and

$$a(\theta, \phi) = R_0 + 3R_1 \sum_{m=0,\pm} f_m(\Theta, \Phi) f_m(\theta, \phi) \tag{3.13}$$

Since $\sum_m f_m(\Theta, \Phi)^2 = 1$, we deduce that the gain is

$$G = (C - \kappa)^{-1} \left[\frac{C(1-\epsilon) - \kappa}{1 - W_0} + \frac{C\epsilon}{1 - W_1/3} \right] \tag{3.14}$$

In terms of the effective stimulus tuning

$$\Gamma = \frac{\epsilon C}{C - \kappa} \tag{3.15}$$

we can re–express the gain as

$$G = \frac{1-\Gamma}{1-W_0} + \frac{\Gamma}{1-W_1/3} \tag{3.16}$$

Note that in the absence of any tuning or bias in the LGN input ($\epsilon = 0$), we have $\Gamma = 0$ and the broad activity profile reduces to the homogeneous state

$$a(\theta, \phi) = \frac{C-\kappa}{1-W_0} \tag{3.17}$$

with gain $G = 1/(1-W_0)$.

The existence and stability of a broad activity profile will depend on both Γ and the weights W_0, W_1. First, since $a_{min} = R_0 - 3R_1$ must be positive we require $\Gamma < \Gamma_c$ where

$$\frac{1}{\Gamma_c} = 1 + \frac{1-W_0}{1-W_1/3} \tag{3.18}$$

(When $\Gamma > \Gamma_c$ the state is narrowly tuned, see below). Second, a simple linear stability analysis shows that the broad activity profile is only asymptotically stable provided that

$$W_0 < 1, \qquad W_1 < 3 \tag{3.19}$$

At $W_0 = 1$ the system undergoes a bulk amplitude instability in which the activity across the network uniformly diverges. On the other hand, at $W_1 = 3$ there is a pattern forming instability associated with the bifurcation to a narrowly tuned or localized state. Indeed, as we establish below, when the spatial modulation of cortical recurrent interactions is sufficiently large, such a localized state can emerge spontaneously from the homogeneous state in the absence of any bias from the LGN input ($\epsilon = 0$).

3.2 Narrow activity profile

In order to simplify our analysis, we assume for the moment that the center of the activity profile is fixed at the low frequency pinwheel, that is, $\Theta = 0$. (The general solution can then be generated by carrying out an $SO(3)$ rotation on the sphere). In this particular case, a state is narrowly tuned if there exists $\theta_c < \pi$ such that $a(\theta, \phi) = 0$ for all $\theta_c \leq \theta \leq \pi$, $0 \leq \phi < \pi$. The cut–off angle θ_c satisfies the equation

$$I_0 + \sum_m I_1^m f_m(\theta_c, \phi) = 0, \quad 0 \leq \phi < \pi \tag{3.20}$$

Taking moments of the fixed point equation (3.9) with respect to the zeroth and first order spherical harmonics,

$$R_0 = I_0 \int_0^{\theta_c} \int_0^{\pi} \mathcal{D}(\theta,\phi) + \sum_{m=0,\pm 1} I_1^m \int_0^{\theta_c} \int_0^{\pi} f_m(\theta,\phi)\mathcal{D}(\theta,\phi) \tag{3.21}$$

and

$$R_1^n = I_0 \int_0^{\theta_c} \int_0^{\pi} f_n(\theta,\phi)\mathcal{D}(\theta,\phi) + \sum_{m=0,\pm} I_1^m \int_0^{\theta_c} \int_0^{\pi} f_n(\theta,\phi) f_m(\theta,\phi)\mathcal{D}(\theta,\phi) \tag{3.22}$$

and performing the integration over θ, ϕ then gives

$$R_0 = \frac{I_0[1-\cos\theta_c]}{2} + \frac{I_1^0[1-\cos 2\theta_c]}{8} \tag{3.23}$$

$$R_1^0 = \frac{I_0[1-\cos 2\theta_c]}{8} + \frac{I_1^0[1-\cos^3\theta_c]}{6} \tag{3.24}$$

and

$$R_1^{\pm} = \frac{I_1^{\pm}[2-3\cos\theta_c+\cos^3\theta_c]}{12} \tag{3.25}$$

It is useful to introduce the functions

$$\mathcal{A}_0(\theta_c) = \frac{1-2\cos\theta_c+\cos^2\theta_c}{4} \tag{3.26}$$

and

$$\mathcal{A}_1(\theta_c) = \frac{2-3\cos\theta_c+\cos^3\theta_c}{12} \tag{3.27}$$

Since $f_{\pm}(0,\Phi) = 0$ for all Φ, it follows from equations (3.6) and (3.25) that

$$R_1^{\pm}\left[1 - W_1\mathcal{A}_1(\theta_c)\right] = 0 \tag{3.28}$$

Provided that $W_1\mathcal{A}_1(\theta_c) \neq 1$, we deduce that $R_1^{\pm} = 0$ and hence $I_1^{\pm} = 0$. Setting $I_1^0 = I_1$ and $R_1^0 = R_1$, the condition for θ_c reduces to

$$I_0 + I_1\cos\theta_c = 0 \tag{3.29}$$

with (see equations (3.5) and (3.6))

$$I_0 = C(1-\epsilon) + W_0R_0 - \kappa, \quad I_1 = C\epsilon + W_1R_1 \tag{3.30}$$

Substituting into equations (3.23) and (3.24) gives $R_0 = \mathcal{A}_0(\theta_c)I_1$ and $R_1 = \mathcal{A}_1(\theta_c)I_1$ so that

$$R_0 = \frac{\epsilon C \mathcal{A}_0(\theta_c)}{1 - W_1 \mathcal{A}_1(\theta_c)} \tag{3.31}$$

and

$$R_1 = \frac{\epsilon C \mathcal{A}_1(\theta_c)}{1 - W_1 \mathcal{A}_1(\theta_c)} \tag{3.32}$$

Given the critical angle θ_c and the effective input I_1, the resulting localized state takes the form

$$a(\theta, \phi) = [I_1(\cos\theta - \cos\theta_c)]_+ \tag{3.33}$$

when centered about the $\Theta = 0$ pinwheel. The corresponding gain defined by equation (3.10) is

$$G = \frac{I_1(1 - \cos\theta_c)}{C - \kappa} \tag{3.34}$$

By performing an $SO(3)$ rotation, it immediately follows that a localized state centered at the point (Θ, Φ) on the sphere is

$$a(\theta, \phi) = \left[I_1 \left(\sum_{m=0,\pm} f_m(\Theta, \Phi) f_m(\theta, \phi) - \cos\theta_c \right) \right]_+ \tag{3.35}$$

Thus a is only non-zero if the angular separation of (θ, ϕ) from (Θ, Φ) is less than the critical angle θ_c. It follows that the boundary of the localized state $\theta = \theta_0(\phi)$ is given by the equation

$$\sum_{m=0,\pm} f_m(\Theta, \Phi) f_m(\theta_0(\phi), \phi) = \cos\theta_c. \tag{3.36}$$

We now determine properties of the localized state in different parameter regimes using a similar analysis to that of the ring model (Hansel & Sompolinsky, 1997).

Figure 10: Critical angle θ_c for the width of the localized state as a function of the stimulus tuning parameter Γ in the case of weak cortical modulation $W_1 \approx 0$.

Weak cortical modulation ($W_0 < 1, W_1 < 3$) For sufficiently weak cortical modulation as defined by the condition $W_1 < 3$, a non–trivial activity profile only exists in the presence of a biased LGN input ($\epsilon > 0$). Whether or not this state is broadly or narrowly tuned will depend on the stimulus parameter Γ. We have already established that the broadly tuned state exists only if $\Gamma < \Gamma_c$, see equation (3.18). On the other hand, when $\Gamma > \Gamma_c$ there exists a narrowly tuned state with critical angle θ_c determined self–consistently from equations (3.29) and (3.30),

$$-\cos\theta_c \equiv \frac{I_0}{I_1} = \frac{C(1-\epsilon)-\kappa}{C\epsilon}\left[1 - W_1\mathcal{A}_1(\theta_c)\right] + W_0\mathcal{A}_0(\theta_c)$$

which can be rearranged to give

$$\frac{1}{\Gamma} = 1 - \frac{W_0\mathcal{A}_0(\theta_c) + \cos\theta_c}{1 - W_1\mathcal{A}_1(\theta_c)} \tag{3.37}$$

Note that $\theta_c \leq \pi$ for $\Gamma \geq \Gamma_c$. In figure 10 we plot the critical angle θ_c as a function of Γ. The corresponding gain of the localized state is

$$G = \Gamma\left[\frac{1-\cos\theta_c}{1 - W_1\mathcal{A}_1(\theta_c)}\right] \tag{3.38}$$

where we have used equations (3.34) and (3.30).

It follows from equation (3.18) that if $W_0, W_1 \approx 0$ then $\Gamma_c \approx 1/2$ so that a stimulus with $\epsilon < 1/2$ and contrast $C \gg \kappa$ will necessarily generate a broad activity profile. Introducing global inhibition by taking $W_0 < 0$ and $W_1 \approx 0$ can sharpen the response by lowering Γ_c: $\Gamma_c \approx 1/(2 + |W_0|)$. However, the gain is also lowered when the level of inhibition is increased since $G \approx \Gamma(1 - \cos\theta_c)$ and the cortical inhibition reduces θ_c. Increasing the degree of cortical modulation W_1 for fixed W_0 also reduces Γ_c such that beyond the critical value $W_1 = 3$ we have $\Gamma_c = 0$ and a localized state can be generated even in the absence of a feedforward bias ϵ.

Figure 11: Phase diagram for spherical model in the case of a homogeneous input $\epsilon = 0$.

Marginal phase and strong cortical modulation ($\epsilon = 0$, $W_0 < W_c$, $W_1 > 3$) When $W_1 > 3$ the unique broadly tuned state (3.13) is unstable, so that any inhomogeneous

state must be narrowly tuned. In the absence of an LGN bias ($\epsilon = 0$) the former reduces to an unstable homogeneous state (3.17). Inspection of equation (3.32) shows that a localized state persists when $\epsilon = 0$ provided that

$$1 = W_1 \mathcal{A}_1(\theta_c) \tag{3.39}$$

Since $\mathcal{A}_1(\theta_c) < 1/3$ for $0 \leq \theta_c \leq \pi$, it follows that $W_1 > 3$ is a necessary condition for a narrowly tuned activity profile to occur when $\epsilon = 0$. The location (Θ, Φ) of the center of the localized state is now arbitrary since the LGN input is homogeneous. In other words, there is a continuum of localized states on the sphere which form a manifold of marginally stable fixed points, and the system is said to be in a *marginal phase*. In such a phase, a narrowly tuned state *spontaneously breaks* the underlying $SO(3)$ symmetry of the network, which is possible because the spatial modulation of the cortical interactions is sufficiently strong.

Figure 12: Variation of critical angle θ_c (dashed curve) and gain G (solid curve) as a function of cortical modulation W_1 in the case of a homogeneous input $\epsilon = 0$. The gain is shown for $W_0 = -10$.

In the marginal phase, the critical angle θ_c is determined by equation (3.39) and is thus independent of W_0. Equations (3.31) and (3.32) imply that

$$\frac{R_0}{R_1} = \frac{\mathcal{A}_0(\theta_c)}{\mathcal{A}_1(\theta_c)} = W_1 \mathcal{A}_0(\theta_c) \tag{3.40}$$

Combining this with equations (3.30) and (3.20) and setting $\epsilon = 0$ then gives

$$I_1 = -\frac{C - \kappa}{\cos\theta_c + W_0 \mathcal{A}_0(\theta_c)} \tag{3.41}$$

and $R_1 = I_1/W_1$. The corresponding gain (3.34) is

$$G = -\frac{1 - \cos\theta_c}{\cos\theta_c + W_0 \mathcal{A}_0(\theta_c)} \tag{3.42}$$

which can be rewritten as

$$G = -\frac{1 - \cos\theta_c}{\mathcal{A}_0(\theta_c)} \frac{1}{W_c - W_0} \tag{3.43}$$

where

$$W_c = -\frac{\cos\theta_c}{\mathcal{A}_0(\theta_c)} \tag{3.44}$$

Equation (3.43) imples that a second condition for the existence of a marginal localized state is that $W_0 < W_c$. Performing a stability analysis shows that as W_0 approaches W_c the system undergoes an amplitude instability analogous to that of the homogeneous state when $W_0 = 1$ and $W_1 < 3$ (see appendix A). The phase diagram for the stability of the various states in the presence of a homogeneous input is shown in figure 11. The variation of the critical angle θ_c and gain G as a function of W_1 is plotted in figure 12.

Figure 13: Two dimensional plot of the tuning surface on the sphere associated with the localized solution (3.35). The activity $a(\theta,\phi)$ is plotted as a function of (θ,ϕ) for fixed width $\theta_c = \pi/3$ and various optimal spatial frequencies Θ and orientations Φ: (a) $\Theta = \pi/4, \Phi = 90^o$. (b) $\Theta = \pi/2, \Phi = 135^o$. (c) $\Theta = 0, \Phi = 0^o$. Light and dark regions denote high and low activities respectively. The figures are related to each other by a rotation of the sphere.

In the case of strong cortical modulation, the presence of a weak input bias ($0 < \epsilon \ll 1$) will not affect the width of the activity profile but will explicitly break the hidden $SO(3)$ symmetry by locking the center of the response (Θ,Φ) to the peak of the LGN input. This establishes a recurrent mechanism for the joint contrast invariance of orientation and spatial frequency tuning curves (see §4). Particular examples of localized states on the sphere are illustrated in figure 13 for $\theta_c = \pi/3$ and various optimal spatial frequencies Θ and orientations Φ. It can be seen that the differing solutions are related by a rotation of the sphere, which reflects the underlying $SO(3)$ symmetry.

Finally note that in order to simplify our analysis of the spherical model, we have considered a one population model in which inhibitory and excitatory cell populations have been collapsed into a single equivalent population. Such a simplification greatly reduces the number of free parameters of the system. The basic insights gained from the one population model can be used to develop the mean field theory of a more realistic two population model. This is presented in appendix B.

Figure 14: Plot of localized tuning surface in the (p, ϕ)–plane in response to a weakly biased LGN input ($\varepsilon \ll 1$) with $\Phi = 90^o$ and (a) $\Theta = \pi/2$ (b) $\Theta = \pi/3$. The width of the localized state is taken to be $\theta_c = \pi/3$. The activity a is shown relative to its maximal value. We have assumed that θ is related to spatial frequency p according to equation (2.1) with $p_{min} = 0.5c/deg$ and $p_{max} = 8c/deg$.

4 Orientation and spatial frequency tuning curves

Our mean field analysis of the spherical model has generated exact solutions for two dimensional localized states on the sphere, which correspond to population tuning surfaces for orientation and spatial frequency preferences within a hypercolumn. A useful representation of the response is obtained by projecting the localized states onto the (p, ϕ)–plane. Surface plots of the resulting activity profiles in the marginal phase are shown in figure 14 for $\Phi = 90^o$ and either (a) $\Theta = \pi/2$ (corresponding to an intermediate spatial frequency $p = 2c/deg.$) or (b) $\Theta = \pi/3$ (corresponding to a lower spatial frequency $p \approx 1.2c/deg.$). Tuning curves for orientation and spatial frequency can then be extracted by taking vertical cross-sections through the tuning surface. Various examples are presented in figures 15–17. In particular, figure 15 illustrates the contrast invariance of the response with respect to both orientation and spatial frequency. In the marginal phase contrast invariance is exact since both the width θ_c and the gain G are independent of contrast, see equations (3.39) and (3.43). Interestingly, approximate contrast invariance also holds for weak cortical modulation (small W_1), since θ_c is a slowly varying function of the synaptic parameter Γ over a broad parameter regime (see figure 11).

Figure 15: Contrast invariance of (a) orientation and (b) spatial frequency tuning curves for $W_1 = 19.2$ and $W_0 = -10$ and a homogeneous input ($\epsilon = 0$). The critical angle $\theta_c = \pi/3$ and the gain $G = 4$. Curves correspond to contrasts (i) $C = 0.2$, (ii) $C = 0.1$ and (iii) $C = 0.05$ relative to threshold κ.

Figure 14 shows that projecting the spherical tuning surface onto the (θ, ϕ)–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 the peak of the tuning surface. This distortion is a direct consequence of the existence of pinwheels, which are

incorporated into our model using a spherical topology, and implies that the responses to orientation and spatial frequency are non-separable. That is, the activity profile cannot be written in the form $a(\theta, \phi) = U(\theta)V(\phi)$. However, we expect approximate separability to occur at intermediate spatial frequencies (away from the pinwheels). The non-separability of the response generates behavior that is consistent with some recent experimental observations:

Figure 16: (a) Orientation tuning curves showing broadening as the optimal spatial frequency Θ changes from intermediate to high or low spatial frequencies: $\Theta = \pi/6$ (thin dashed curve), $\Theta = \pi/3$ (thin solid curve), $\Theta = \pi/2$ (thick solid curve) and $\Theta = \pi/8$ (thick dashed curve). The optimal orientation is fixed at $\Phi = 90^o$ and $\theta_c = \pi/3$. The activity a is shown relative to its maximal value. (b) Spatial frequency tuning curves showing invariance of the degree of tuning with respect to Θ. Same parameter values as (a) except $\Theta = 2\pi/3$ (thick dashed curve). We have assumed that θ is related to p according to equation (2.1) with $p_{min} = 0.5c/deg$ and $p_{max} = 8c/deg$.

Figure 17: Spatial frequency tuning curves $a(\theta, \phi)$ as a function of θ for various orientations $\phi = \Phi + \delta\phi$: (i) $\delta\phi = 0^o$ (ii) $\delta\phi = 14^o$ (iii) $\delta\phi = 28^o$. In the case of a low optimal frequency $\Theta = \pi/3$, figure (a), there is a downward shift in the peak of the response, whereas there is an upward shift in the case of a high optimal frequency $\Theta = 2\pi/3$, figure (b).

(a) At high and low spatial frequencies (towards the pinwheels) there is a broadening of the tuned response to orientation. This is illustrated in figure 16(a) where we plot orientation tuning curves $a(\Theta, \phi)$ as a function of ϕ for various optimal spatial frequencies Θ. It can be seen that the width increases towards the low (and high) orientation pinwheel. No such broadening occurs for the corresponding spatial frequency tuning curves as shown in figure 16(b). In our model the reduction of orientation selectivity around the pinwheels is an aggregate property of a population of cells. Interestingly, it has been found experimentally that individual neurons close to pinwheels are actually orientation selective (O'Keefe, Levitt, Kiper, Shapley, & Movshon, 1998), but there is a broad distribution of orientation preferences within the pinwheel region so that the average response of the population is only weakly orientation selective. Note that our results differ from

those of (McLaughlin, Shapley, Shelley, & Wielaard, 2000) who find a sharpening of orientation tuning near pinwheels. We attribute this difference to the $SO(3)$ symmetry we impose on the weighting function $w(\theta, \phi|\theta', \phi')$.

(b) There is a systematic shift and narrowing of spatial frequency tuning curves at non–optimal orientations – the shift is towards the closest pinwheel, see figure 17. There is some suggestion of spatial frequency shifts in recent optical imaging data (Issa *et al.*, 2000). Note, however, that one difference between our model prediction and the data is that the latter appears to indicate a downward rather than an upward shift in response at high spatial frequencies. (A downward shift is also consistent with feedforward receptive field properties, see figure 21). We suggest in §7 that the downward shift could be reversed by cortico–geniculate feedback (after some delay).

Figure 18: Polar plots of localized activity state $a(\theta, \phi)$ for fixed width $\theta_c = \pi/3$, fixed optimal orientation $\Phi = 0^o$ and increasing optimal spatial frequency Θ: (a) $\Theta = \pi/6$ (b) $\Theta = \pi/3$ (c) $\Theta = \pi/2$ (d) $\Theta = 2\pi/3$. Here ϕ is taken to be the polar angle and θ the radius in the plane such that the origin represents the low frequency pinwheel at $\theta = 0$, whereas the outer circle represents the high frequency pinwheel at $\theta = \pi$. Darker regions correspond to higher levels of activity. In each figure, $\Delta\theta = 2\theta_c$ is indicated by the thick horizontal line and $\Delta\phi$ is indicated by the thick arc, reaching a minimum at $\Theta = \pi/2$.

Another useful representation of the response is to consider contour plots of the activity profile in the (θ, ϕ)–plane as shown in figure 18. Here we use polar coordinates with radius θ and polar angle ϕ. This figure further illustrates the non-separability of the response. Define $\Delta\theta$ as the width of the activity profile at the optimal orientation Φ and $\Delta\phi$ as the width of the activity profile at the optimal spatial frequency Θ. It follows from equation (3.35) that $\Delta\theta = 2\theta_c$ irrespective of the position of the center of the localized state. On the other hand, $\Delta\phi$ varies with the optimal frequency Θ, reaching a minimum at $\Theta = \pi/2$. Sufficiently close to the pinwheels, $\Theta < \theta_c/2$ or $\Theta > \pi - \theta_c/2$, we have $\Delta\phi = \pi$, which implies that although the response is localized on the sphere it is broadly tuned for orientation. Finally, in figure 19 we show a log-polar plot of various localized responses, which is at least suggestive of the single-cell data reproduced in figure 3. We select a narrow tuning width for ease of illustration and since the data in figure 3 is thresholded.

We emphasize that the results presented in this section describe the response of a cortical hypercolumn to a fixed visual stimulus (population tuning curves) rather than

Figure 19: Log-polar plot of various localized activity states for fixed width $\theta_c = \pi/6$ and various optimal orientations Φ and spatial frequencies $P = \mathcal{Q}^{-1}(\Theta)$ fixed optimal orientation $\Phi = 0^o$: (a) $P = 1c/deg, \Phi = 0^o$ (b) $P = 2c/deg, \Phi = 90^o$ (c) $P = 3c/deg, \Phi = 135^o$ (d) $P = 4c/deg, \Phi = 45^o$. Here ϕ is taken to be the polar angle and $\log_2 p$ the radius.

the response of a single cell to a range of stimuli (single cell tuning curves). The non–separability arising from the pinwheels is thus a population effect and may be reduced or even absent at the single cell response. Interestingly, recent single cell recordings suggest that there is approximate separability of orientation and spatial frequency tuning curves except at low and high spatial frequencies (Mazer *et al.*, 2002), which is consistent with our population results.

Part II

Receptive fields and cortico–geniculate feedback

In Part II of this paper we show that if the low order spherical harmonic components of the filtered feedforward input to a hypercolumn are amplified by recurrent interactions, then the spatial frequency at which the cortical response is optimal is shifted relative to the stimulus frequency—there is no corresponding shift in orientation. In other words, the network does not faithfully encode the stimulus spatial frequency. This shift in spatial frequency is not an artifact of the particular spherical network topology. A similar conclusion would obtain for any recurrent mechanism that amplifies both orientation and spatial frequency components of the LGN input. We propose that the feedback pathway from V1 back to LGN, recently investigated in cats (Murphy *et al.*, 1999), modulates LGN activity to produce a faithful encoding of spatial frequency. Using linear filter theory we show that if the feedback from a cortical cell is taken to be approximately equal to the reciprocal of the corresponding feedforward receptive field (in the two–dimensional Fourier domain), then the mismatch between the feedforward and cortical frequency representations is eliminated (at least at the linear level). We predict that for intermediate spatial frequencies, the cortico–geniculate innervation pattern is oriented in a direction related to the orientation bias of its V1 origin. However for high and low spatial frequencies, no direction of innervation should exist.

5 Feedforward receptive fields

One possible model of the two dimensional receptive field of a simple cell (in retinal co–ordinates $\mathbf{r} = (x, y)$) is the difference of Gaussians (Hawken & Parker, 1987)

$$u(\mathbf{r}) = \frac{\sqrt{\kappa}}{2\pi\sigma_+} \exp\left[-\frac{1}{2\sigma_+^2}\left(\kappa^2 x^2 + y^2\right)\right] - \frac{\alpha}{2\pi\sigma_-} \exp\left[-\frac{1}{2\sigma_-^2}\left(x^2 + y^2\right)\right] \tag{5.1}$$

This represents a center–surround profile in which the excitatory center is an ellipse with eccentricity $\kappa > 1$ whose major axis runs along the y–direction. The inhibitory surround is taken to be circular but with a larger half width, $\sigma_- > \sigma_+$. The parameter κ is a measure of the degree of feedforward orientation selectivity due to the alignment of LGN

circular receptive fields along the vertical direction ($\phi = 0$). Taking the two dimensional Fourier transform of u gives

$$U(\mathbf{k}) = \exp\left[-\frac{\sigma_+^2 k^2}{2}\left(\kappa^{-2}\cos^2\varphi + \sin^2\varphi\right)\right] - \alpha\exp\left[-\frac{\sigma_-^2 k^2}{2}\right] \tag{5.2}$$

for $\mathbf{k} = (k, \varphi)$ in polar coordinates. The function U has a maximum at $\mathbf{p} = (p, \phi)$ so that $U(\mathbf{p}) \geq U(\mathbf{k})$ for all $\mathbf{k}$, with $\phi = 0, \pi$ and

$$p = \sqrt{\frac{4}{\sigma_-^2 - \kappa^{-2}\sigma_+^2}\ln\left[\frac{\sqrt{\alpha}\kappa\sigma_-}{\sigma_+}\right]} \tag{5.3}$$

Setting $\sigma_+ = \sigma$ and $\sigma_- = \hat{\kappa}\sigma$ and taking $\hat{\kappa}, \kappa, \alpha$ to be fixed, it follows that the spatial frequency preference p is inversely proportional to the size σ of the receptive field,

$$p = \frac{A}{\sigma}, \quad A = \sqrt{\frac{4}{\hat{\kappa}^2 - \kappa^{-2}}\ln\left[\sqrt{\alpha}\kappa\hat{\kappa}\right]} \tag{5.4}$$

and we can rewrite u as

$$u(\mathbf{r}|p) = \frac{p\sqrt{\kappa}}{2\pi A}\exp\left[-\frac{p^2}{2A^2}\left(\kappa^2 x^2 + y^2\right)\right] - \frac{\alpha p}{2\pi\hat{\kappa}A}\exp\left[-\frac{p^2}{2A^2\hat{\kappa}^2}\left(x^2 + y^2\right)\right] \tag{5.5}$$

Now consider a cell with receptive field profile centered at the retinal coordinate $\widehat{\mathbf{r}}$ with (feedforward) orientation preference ϕ and spatial frequency preference p. Given a visual stimulus of intensity $i(\mathbf{r})$, the effective input from the LGN to the cell will be of the form

$$h_{LGN}(\widehat{\mathbf{r}}|\mathbf{p}) = \int i(\mathbf{r})u(\widehat{\mathbf{r}} - \mathbf{r}|\mathbf{p})d\mathbf{r} \tag{5.6}$$

where $u(\mathbf{r}|\mathbf{p}) = u(T_\phi\mathbf{r}|p)$ and

$$T_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{5.7}$$

Taking the Fourier transform of equation (5.6) gives

$$H_{LGN}(\mathbf{k}|\mathbf{p}) = I(\mathbf{k})U(\mathbf{k}|\mathbf{p}) \tag{5.8}$$

where

$$U(\mathbf{k}|\mathbf{p}) = \exp\left[-\frac{A^2k^2}{2p^2}\left(\kappa^{-2}\cos^2(\varphi - \phi) + \sin^2(\varphi - \phi)\right)\right] - \alpha\exp\left[-\frac{\hat{\kappa}^2A^2k^2}{2p^2}\right] \tag{5.9}$$

In the particular case of a sinusoidal grating of contrast C_s, spatial frequency p_s and orientation ϕ_s,

$$i(\mathbf{r}) = C_s \cos[\, p_s(x\cos\phi_s + y\sin\phi_s)]\,, \tag{5.10}$$

we have

$$h_{LGN}(\widehat{\mathbf{r}}|\mathbf{p}) = C_s\, U(\mathbf{p}_s|\mathbf{p})\cos(\mathbf{p}_s\cdot\widehat{\mathbf{r}}) \tag{5.11}$$

Thus, when the grating is centered in the receptive field of the neuron, so that $\widehat{\mathbf{r}} = 0$, $h_{LGN}(0|\mathbf{p}) = C_s\, U(\mathbf{p}_s|\mathbf{p})$, i.e., the resulting LGN input is given by the Fourier transform of the receptive field multiplied by the stimulus contrast C_s, as expected.

Figure 20: LGN input h_{LGN} to a single cell obtained by filtering a sinusoidal grating with the difference–of–Gaussian receptive field (5.5) for a range of stimulus spatial frequencies p_s and orientations ϕ_s (with zero spatial phase). Parameters of the LGN receptive field are $\kappa = 1.5$, $\hat{\kappa} = 3$ with a variable level of surround inhibition α. Figure (a) shows the input (in units of the stimulus contrast C_s) as a function of stimulus frequency p_s for a fixed spatial frequency preference $p = 1$ and $\phi = \phi_s$. The units of spatial frequency are taken to be cycles/deg. Figure (b) shows the corresponding input as a function of stimulus orientation ϕ_s for a fixed orientation preference $\phi = 90^0$ and $p = p_s$.

Figure 21: Shift in spatial frequency peak of the LGN input at non-optimal orientation $\phi \neq \phi_s$. Here $p = 4$, $\phi = 0^o$ and $\kappa = 2$. All other parameters as in figure 20.

In figure 20(a) we plot the resulting LGN input as a function of stimulus frequency p_s for $\phi = \phi_s$, $p = 1$ and various levels of surround inhibition α. It can be seen that for relatively low levels of inhibition, the LGN acts like a lowpass spatial frequency filter with a shallow maximum at $p_s = p$. When the inhibition is increased, however, the profile is sharpened and the LGN acts more like a bandpass filter. The corresponding input profile as a function of orientation preference ϕ is shown in figure 20(b) for $\phi_s = 90^o$. The response has a shallow maximum at $\phi = \phi_s$, with a relatively large DC component that decreases with increasing surround inhibition and increasing κ. There is also a spatial frequency shift in the LGN input at non–optimal orientations ($\phi \neq \phi_s$), which is always to lower frequencies. This follows from equations (5.9) and (5.11), since when $\phi \neq \phi_s$ there is an effective reduction in the anisotropy parameter κ of the form $\kappa^{-2} \rightarrow$

$\kappa^{-2}\left(\cos^2(\phi_s-\phi)+\sin^2(\phi_s-\phi)\right)$. Such a reduction reduces the spatial frequency at which the input reaches a maximum, see equation (5.3), and this is true for all spatial frequencies as shown in figure 21 for $p = 4c/deg$. This should be contrasted with the corresponding shift in the cortical response, which is to higher frequencies (see figure 17(b)).

6 Spherical harmonic projection of the LGN input

Now consider a cortical hypercolumn whose cells are parameterized by the orientation preference $\phi \in [0, \pi]$ and spatial frequency preference $p \in [p_{min}, p_{max}]$, with the pair (p, ϕ) determined by the feedforward receptive field properties of the cells (see §5). Following Part I, we assume that the network topology is a sphere with angular coordinates (θ, ϕ), where θ is related to the spatial frequency preference p according to equation (2.1). We have already shown how amplification and rectification of certain spherical harmonic components of a weakly biased LGN input can generate orientation and spatial frequency tuning. We are now interested in the consequences of selecting out these particular harmonic components without worrying about the additional rectification stage. Therefore, we restrict our analysis to linear theory and treat the cortex as a linear filter carrying out the transformation $h_{LGN} \to \mathbf{P}h_{LGN}$ where $\mathbf{P}$ denotes the projection onto the zeroth and first order spherical harmonic components and h_{LGN} is the total feedforward input from the LGN[2] (see figure 22).

Figure 22: Schematic diagram of feedforward pathways. A visual stimulus i is convolved with a feedforward receptive field u to generate a cortical input $h_{LGN} = u \circ i$. (The convolution operator $\circ$ is defined in equation (7.3)). Recurrent interactions within $V1$ amplify low order spherical harmonic components to generate a response $h = \mathbf{P}h_{LGN}$. The contour plot of a difference of Gaussians receptive field profile is shown in retinal coordinates. The length scale is in units of the range of feedforward excitation. Dark and light regions represent excitatory and inhibitory afferents respectively. Parameters of the LGN receptive field are $\kappa = 1.5$, $\hat{\kappa} = 3$ and $\alpha = 0.5$.

Suppose, for the moment, that the receptive field centers of all neurons within a given hypercolumn are located at the same retinal coordinate $\widehat{\mathbf{r}}$. Then $h_{LGN}(\widehat{\mathbf{r}}|\mathbf{p})$, for fixed

[2]At first sight, this may be confusing since we took $h = \mathbf{P}h_{LGN}$ to be the input to the cortex in Part I. We are essentially decomposing the operation of cortex into two distincts parts: (i) selection through amplification $h_{LGN} \to \mathbf{P}h_{LGN}$ and (ii) tuning through amplification and rectification $\mathbf{P}h_{LGN} \to a$.

$\widehat{\mathbf{r}}$, determines the LGN input distribution across the hypercolumn. Projecting onto the first-order harmonics, it follows that

$$\begin{aligned} h(\theta,\phi) &\equiv \mathbf{P}h_{LGN}(\widehat{\mathbf{r}}|\mathcal{Q}^{-1}(\theta),\phi) \\ &= h_0 + \sum_{m=0,\pm} h_1^m f_m(\theta,\phi) \end{aligned} \tag{6.1}$$

where

$$h_0 = \frac{1}{2\pi}\int_0^\pi \int_0^\pi h_{LGN}(\widehat{\mathbf{r}}|\mathcal{Q}^{-1}(\theta),\phi)\sin\theta d\theta d\phi \tag{6.2}$$

and

$$h_1^m = \frac{3}{2\pi}\int_0^\pi \int_0^\pi f_m(\theta,\phi)h_{LGN}(\widehat{\mathbf{r}}|\mathcal{Q}^{-1}(\theta),\phi)\sin\theta d\theta d\phi \tag{6.3}$$

Note that in order for the resulting distribution $h(\theta,\phi)$ to be a well defined function on the sphere, it must be independent of ϕ at $\theta = 0,\pi$. Equations (5.6) and (5.9) then require that $\kappa = 1$ at the pinwheels, in other words, the average orientation preference of receptive fields at the pinwheels must be zero. Hence, the existence of a non–zero preference away from the pinwheels, implies that the orientation selectivity parameter κ has to be spatial frequency dependent. For concreteness, we take

$$\kappa = \kappa(\theta) \equiv \kappa_0 \sin^2(\theta) + \cos^2(\theta) \tag{6.4}$$

with $\kappa_0 > 1$ so that the selectivity is maximal at intermediate spatial frequencies and zero at the pinwheels.

We now calculate $h(\theta,\phi)$ for a sinusoidal grating with stimulus frequency p_s, orientation ϕ_s and zero spatial phase ($\widehat{\mathbf{r}} = 0$). We use the identities $\cos(2\phi) = 2\cos^2\phi - 1 = 1 - 2\sin^2\phi$ and

$$\mathrm{e}^{x\cos 2\phi} = I_0(x) + 2\sum_{n\geq 1} I_n(x)\cos(2n\phi) \tag{6.5}$$

where $I_n(x)$ is the modified Bessel function of integer order n. Equation (5.9) can then be expanded as

$$U(\mathbf{k}|\mathbf{p}) = \sum_{n=0}^{\infty} U_n(k|p)\cos 2n(\varphi - \phi) \tag{6.6}$$

where

$$U_0(k|p) = \exp\left[-\frac{\kappa^+ A^2k^2}{2p^2}\right] I_0\left(\frac{\kappa^- A^2k^2}{2p^2}\right) - \alpha\exp\left[-\frac{\hat{\kappa}^2 A^2k^2}{2p^2}\right] \tag{6.7}$$

and

$$U_n(k|p) = 2\exp\left[-\frac{\kappa^+ A^2 k^2}{2p^2}\right] I_n\left(\frac{\kappa^- A^2 k^2}{2p^2}\right) \tag{6.8}$$

for $n > 0$ with $\kappa^\pm = (1 \pm \kappa^{-2})/2$. Setting $h_{LGN}(0|\mathbf{p}) = C_s U(\mathbf{p}_s|\mathbf{p})$ and using equations (6.1)–(6.3) and (6.6), we find that

$$h_0 = C_s h_0(p_s), \quad h_1^0 = C_s h_1(p_s) \tag{6.9}$$

and

$$h_1^+ = C_s h_2(p_s)\cos\phi_s, \; h_1^- = C_s h_2(p_s)\sin\phi_s \tag{6.10}$$

with

$$h_0(p_s) = \frac{1}{2}\int_{p_{min}}^{p_{max}} U_0(p_s|p)\sin(\mathcal{Q}(p))d\mathcal{Q}(p) \tag{6.11}$$

$$h_1(p_s) = \frac{3}{4}\int_{p_{min}}^{p_{max}} U_0(p_s|p)\sin(2\mathcal{Q}(p))d\mathcal{Q}(p) \tag{6.12}$$

and

$$h_2(p_s) = \frac{3}{4}\int_{p_{min}}^{p_{max}} U_1(p_s|p)\sin^2(\mathcal{Q}(p))d\mathcal{Q}(p) \tag{6.13}$$

Figure 23: Plot of cortical spatial frequency $\Theta = \widetilde{\mathcal{Q}}(p_s)$ as a function of stimulus spatial frequency p_s for the difference of Gaussians receptive field with $\kappa_0 = 1.5$, $\hat{\kappa} = 3$ and various levels of inhibition α. The dotted line $\Theta = \mathcal{Q}(p_s)$ corresponds to a faithful encoding of spatial frequency.

Substitution of equations (6.9) and (6.10) into equation (6.1) recovers the form assumed for h in equation (2.10) of Part I, namely,

$$\begin{aligned} h(\theta,\phi) &= C\left[(1-\epsilon) + \epsilon\sum_{m=0,\pm} f_m(\Theta,\Phi) f_m(\theta,\phi)\right] \\ &= C(1-\epsilon) + \epsilon C\left(\cos\Theta\cos\theta + \sin\Theta\sin\theta\cos(2[\phi-\Phi])\right) \end{aligned} \tag{6.14}$$

where

$$\Theta = \widetilde{\mathcal{Q}}(p_s), \quad \Phi = \phi_s \tag{6.15}$$

$$C(1-\epsilon) = C_s h_0(p_s), \quad C\epsilon = C_s \Gamma(p_s) \tag{6.16}$$

and

$$\widetilde{\mathcal{Q}}(p_s) = \tan^{-1}\frac{h_2(p_s)}{h_1(p_s)}, \quad \Gamma(p_s) = \sqrt{h_1(p_s)^2 + h_2(p_s)^2} \tag{6.17}$$

The phase $\widetilde{\mathcal{Q}}(p_s)$ is plotted as a function of stimulus frequency in figure 23 for various levels of feedforward inhibition α. This clearly shows that $\Theta = \widetilde{\mathcal{Q}}(p_s) \neq \mathcal{Q}(p_s)$—there is a strong magnification of the representation of spatial frequency in the intermediate range, with small changes in p_s inducing large changes in the location Θ of the peak of the tuned response. Thus, there is a mismatch between the spatial frequency encoded by the hypercolumn (given by Θ) and the input spatial frequency p_s of the stimulus. In figure 24 we plot the variation of $h_0(p_s)$ and $\Gamma(p_s)$ with stimulus frequency. These functions determine the effective contrast C and bias ϵ according to equation (6.16) so that, in particular, the contrast $C = h_0(p_s) + \Gamma(p_s)$. In the mean–field analysis of §3 we showed that under amplification and rectification a localized activity state is generated whose amplitude varies as ϵC (weak cortical modulation) or as C (strong cortical modulation, weak bias $\epsilon \ll 1$). We see from figure 24 that the projection onto spherical harmonics leads to a non–trivial dependence of the reponse amplitude on stimulus frequency. This appears to be inconsistent with physiological (Issa *et al.*, 2000) and psychophysical (De Valois & De Valois, 1988) data, which indicates that the response amplitude is a unimodal function that peaks at a single intermediate frequency. Another interesting observation regarding figure 24, is that the LGN bias ϵ cannot be assumed to be small across the whole spatial frequency range.

Figure 24: Plot of $h_0(p_s)$ (thin curve), $\Gamma(p_s)$ (dashed curve) and the contrast C (thick curve) as a function of stimulus spatial frequency p_s for (a) $\alpha = 0.5$ (b) $\alpha = 1.0$. Other parameter values as in figure 23

The origin of the mismatch $\Theta \neq \mathcal{Q}(p_s)$ is the assumption that recurrent cortical interactions amplify both orientation and spatial frequency components of the LGN input. Such a mismatch would not occur if Fourier modes with respect to the orientation label ϕ alone were amplified, as in the ring model of orientation tuning (Ben-Yishai *et al.*, 1995). In such a case one can represent the effective LGN input for a fixed spatial frequency preference p by

$$h(\phi) = C(1-\epsilon) + C\epsilon\cos(2[\phi - \phi_s)] \tag{6.18}$$

with, see equation (6.6),

$$C(1-\epsilon) = U_0(p_s|p), \quad C\epsilon = U_1(p_s|p) \tag{6.19}$$

Suppose that the stimulus frequency p_s is fixed and we plot $U_{0,1}(p_s|p)$ as a function of the spatial frequency preference p. The results are shown in figure 25. At the optimal orientation $\phi = \phi_s$ the spatial frequency dependence of the input is given by the effective contrast $C = U_0(p_s|p) + U_1(p_s|p)$. It can be seen from figure 25 that C peaks when the spatial frequency preference is approximately equal to the stimulus frequency, $p = p_s$, so that the network response now faithfully encodes the stimulus. However, the resulting spatial frequency tuning curves are neither sharply tuned nor contrast invariant. (These tuning curves are directly given by C since there is no amplification with respect to p). One way to achieve more realistic tuning curves is to posit that recurrent interactions also amplify spatial frequency components of the LGN input along the lines of the spherical model. One then has to tackle the resulting mismatch between stimulus frequency and response frequency.

Figure 25: Plot of $U_0(p_s|p)$ (dashed curve), $U_1(p_s|p)$ (thin curve) and contrast C (thick curve) as a function of spatial frequency preference p for fixed stimulus frequency p_s: (a) $p_s = 1c/deg$ and (b) $p_s = 4.0c/deg$. Other parameter values as in figure 23 for $\alpha = 0.5$.

7 Renormalizing the LGN input

It follows from the above analysis that if the cortex amplifies the first order spherical harmonic components of a stimulus, then in order to generate a faithful representation of spatial frequency, $\Theta = \mathcal{Q}(p_s)$, the LGN input cannot be determined only by the feedforward receptive field properties of single neurons. In other words, there must exist another filtering operation $\mathcal{P}$ that converts $\widetilde{\mathcal{Q}}(p_s)$ into $\mathcal{Q}(p_s)$. Of course, an alternative possibility is that the proposed amplification mechanism is itself invalid. However, we expect a similar conclusion to hold for any feedforward or recurrent mechanism that amplifies two dimensional Fourier components of the stimulus—the basic problem lies with the fact that the response is non–separable with respect to the orientation and spatial frequency labels. It therefore remains to discuss possible mechanisms for the filtering action $\mathcal{P}$ that effectively *renormalizes* the feedforward LGN input.

7.1 Feedforward mechanisms

One possible feedforward mechanism is patch averaging. For simplicity, we have assumed that every cortical cell within a local patch has the same receptive field profile u, equation (5.5), with identical parameters $\alpha, \kappa, \hat{\kappa}$ and receptive field centers $\widehat{\mathbf{r}}$. In reality there will be a distribution of receptive fields so that the filter action $\mathcal{P}$ could arise from some form of patch averaging. For example, figure 23 suggests that if there were some variation in the level of feedforward inhibition α, then this would smooth out the response. A more realistic source of variation is that of receptive field positions within each cortical column. Let the distribution of centers within a patch be $\rho(\widehat{\mathbf{r}}|p)$, where the degree of scatter may depend on p, the spatial frequency preference of the patch. Equation (6.3) is then modified according to

$$h_1^m = \frac{3}{2\pi}\int_0^\pi \int_0^\pi f_m(\theta,\phi)\left[\int d\widehat{\mathbf{r}}\rho(\widehat{\mathbf{r}}|\mathcal{Q}^{-1}(\theta))h_{LGN}(\widehat{\mathbf{r}}|\mathcal{Q}^{-1}(\theta),\phi)\right]\sin\theta d\theta d\phi \tag{7.1}$$

and similarly for equation (6.2). Such averaging may be expected to smooth cortical responses.

A second feedfoward mechanism is a nontrivial mapping of the cortical labels. The projection of the LGN input onto the low order spherical harmonics given by equations (6.2) and (6.3) assumes that the cortical labels for orientation and spatial frequency (p, ϕ) are determined completely by properties of the feedforward receptive fields (see §5). This is the classical Hubel–Wiesel mechanism for generating the feature preferences of a cell. We have shown that such an identification leads to a mismatch in the representation of spatial frequency within the cortex. One possible way to eliminate such a mismatch is to allow for a nontrivial mapping between properties of the receptive field and the cortical labels that regularizes the projection of the LGN input and, hence, generates a faithful representation of spatial frequency. This mapping reflects the fact that the actual spatial frequency and orientation preference of a cell is determined by a combination of feedforward and recurrent interactions. A renormalization scheme of this form would require the development of a pattern of innervation from LGN to cortex that involves some form of *feedback* from cortex to LGN in order to implement an error correcting procedure. But such feedback can itself provide a direct mechanism for renormalizing the LGN input, as we describe below.

7.2 Cortico–geniculate feedback

We construct a recurrent filter that converts the feedforward or bare receptive field u into an effective or renormalized one, namely u^*, such that the renormalized LGN input

$$h^*_{LGN}(\widehat{\mathbf{r}}|\mathbf{p}) = \int i(\mathbf{r}) u^*(\widehat{\mathbf{r}} - \mathbf{r}|\mathbf{p}) d\mathbf{r} \tag{7.2}$$

projects faithfully onto its spherical harmonic components. Inverse Fourier transforming the recurrent filter then determines the pattern of feedback connections from V1 to LGN. First, suppose that a cortical cell with receptive field center $\widehat{\mathbf{r}}$, spatial frequency preference p and orientation preference ϕ, has a distribution of feedback connections $v(\mathbf{r} - \widehat{\mathbf{r}}|\mathbf{p})$ to LGN cells that innervate cortical cells with the same feature preference and shifted center at $\mathbf{r}$, i.e. the LGN cells make cortical connections with a weighting function $u(\widehat{\mathbf{r}} - \mathbf{r}|\mathbf{p})$. In other words, we assume that localized patches in cortex and LGN are reciprocally related (Murphy *et al.*, 1999; Guillery, Feig, & van Lieshout, 2001), see figure 26. Such a principle also seems to hold with respect to feedback from extrastriate to striate areas (Angelucci, Levitt, & Lund, 2001).

Figure 26: Schematic diagram showing reciprocally related regions in V1 and LGN: (a) feedforward projections (b) feedback projections.

Within the framework of linear filter theory, we take the output activity of cortex to consist of the spherical harmonic components of the renormalized LGN input. (A more complete calculation would need to take into account amplification and rectification of $\mathbf{P}h_{LGN}$). We write this output activity in the form $\mathbf{P}u^* \circ i(\widehat{\mathbf{r}})$, where i is the input stimulus and $f \circ g$ for arbitrary functions f, g denotes the convolution

$$[f \circ g](\mathbf{r}) = \int_{\mathbf{R}^2} f(\mathbf{r}) g(\mathbf{r} - \mathbf{r}') d\mathbf{r}'. \tag{7.3}$$

We then assume that (see figure 27)

$$u^* \circ i = u \circ [i + v \circ \mathbf{P}\,(u^* \circ i)] \tag{7.4}$$

Taking the Fourier transform of this equation using the convolution theorem,

$$U^*(\mathbf{k}|\mathbf{p}) I(\mathbf{k}) = U(\mathbf{k}|\mathbf{p}) I(\mathbf{k})\,[1 + V(\mathbf{k}|\mathbf{p}) \mathbf{P} U^*(\mathbf{k}|\mathbf{p})] \tag{7.5}$$

for $\mathbf{k} = (k, \varphi)$ in polar co-ordinates. Rearranging this equation leads to the result

$$V(\mathbf{k}|\mathbf{p}) = \frac{U^*(\mathbf{k}|\mathbf{p}) - U(\mathbf{k}|\mathbf{p})}{U(\mathbf{k}|\mathbf{p})\mathbf{P}U^*(\mathbf{k}|\mathbf{p})} \tag{7.6}$$

As a further simplification, suppose $U^* - U = \mathbf{P}[U^* - U]$ so that

$$V(\mathbf{k}|\mathbf{p}) = \frac{1}{U(\mathbf{k}|\mathbf{p})}\left[1 - \frac{\mathbf{P}U(\mathbf{k}|\mathbf{p})}{\mathbf{P}U^*(\mathbf{k}|\mathbf{p})}\right] \tag{7.7}$$

Both $\mathbf{P}U(\mathbf{k}|\mathbf{p})$ and $\mathbf{P}U(\mathbf{k}|\mathbf{p})$ can be expressed in terms of zeroth and first-order spherical harmonics:

$$\begin{aligned}\frac{\mathbf{P}U(\mathbf{k}|\mathbf{p})}{\mathbf{P}U^*(\mathbf{k}|\mathbf{p})} &= \left[\frac{C(k)}{C^*(k)}\right] \\ &\times \frac{(1-\epsilon(k)) + \epsilon(k)\left(\cos\Theta(k)\cos\theta + \sin\Theta(k)\sin\theta\cos(2[\phi-\varphi])\right)}{(1-\epsilon^*(k)) + \epsilon^*(k)\left(\cos\Theta^*(k)\cos\theta + \sin\Theta^*(k)\sin\theta\cos(2[\phi-\varphi])\right)}\end{aligned} \tag{7.8}$$

The unrenormalized functions $C(k), \epsilon(k), \Theta(k)$ satisfy equations (6.15) and (6.16). In particular, $\Theta(k) = \widetilde{\mathcal{Q}}(k)$. Hence, the feedback distribution V is determined once we have specified the k–dependence of the renormalized functions $C^*(k), \epsilon^*(k)$ and $\Theta^*(k)$. It also follows that the mismatch in spatial frequencies highlighted in §6 is eliminated provided that $\Theta^*(k) = \mathcal{Q}(k)$.

Figure 27: Schematic diagram of feedforward, recurrent and feedback pathways that could be involved in the generation of a faithful representation of spatial frequency. The contour plot of a feedforward (difference of Gaussians) receptive field profile is shown in retinal coordinates, together with the corresponding pattern of feedback connections. The latter is calculated using linear filter theory. The length scale is in units of the range of feedforward excitation. Dark and light regions represent excitatory and inhibitory synapses respectively.

Further insight into the nature of the feedback connections can be obtained under the additional assumption that $C \ll C^*$ over the frequency bandwidth of the hypercolumn. The lowest order approximation is then

$$V(\mathbf{k}|\mathbf{p}) \approx \frac{\mathcal{A}}{U(\mathbf{k}|\mathbf{p})} \tag{7.9}$$

where $\mathcal{A} < 1$ is a constant. Using equation (6.6) and keeping only the lowest–order terms,

$$\begin{aligned} V(\mathbf{k}|\mathbf{p}) &\approx \frac{\mathcal{A}}{U_0(k|p) + U_1(k|p)\cos(2[\varphi - \phi])} \\ &\approx \frac{\mathcal{A}}{U_0(k|p)}\left[1 - \frac{U_1(k|p)}{U_0(k|p)}\cos(2[\varphi - \phi])\right] \end{aligned} \tag{7.10}$$

Taking the inverse Fourier transform of this equation shows that

$$v(\mathbf{r}|\mathbf{p}) = \mathrm{Re}\int_{p_{min}}^{p_{max}}\int_0^{2\pi} V(\mathbf{k}|\mathbf{p})\mathrm{e}^{i\mathbf{r}\cdot\mathbf{k}}\frac{kdkd\varphi}{4\pi^2} \tag{7.11}$$

where we have assumed that the feedback is restricted to lie within the frequency bandwidth $[p_{min}, p_{max}]$ of the hypercolumn. In order to evaluate this integral, introduce polar coordinates $x = r\cos\psi, y = r\sin\psi$ so that $\mathbf{k}\cdot\mathbf{r} = kr\cos(\varphi - \psi)$. The Bessel function expansion

$$\cos(x\cos\psi) = J_0(x) + 2\sum_{n=1}^{\infty}(-1)^n J_{2n}(x)\cos(2n\psi) \tag{7.12}$$

then gives

$$v(\mathbf{r}|\mathbf{p}) = v_0(r|p) + v_1(r|p)\cos(2[\phi - \psi]) \tag{7.13}$$

where

$$\begin{aligned} v_0(r|p) &= \int_{p_{min}}^{p_{max}} \frac{\mathcal{A}}{U_0(k|p)} J_0(kr)\frac{kdk}{2\pi} \\ v_1(r|p) &= \mathcal{A}\int_{p_{min}}^{p_{max}} \frac{U_1(k|p)}{U_0(k|p)^2} J_2(kr)\frac{kdk}{2\pi}. \end{aligned} \tag{7.14}$$

We have thus specified an approximate form of the recurrent filter that effectively renormalizes the LGN input. We see from equations (7.9) and (7.11) that this filter is approximately the inverse Fourier transform of the reciprocal of $U(\mathbf{k}|\mathbf{p})$, the Fourier transform of the feedforward receptive field $u(\mathbf{r}|p)$ located at $\mathbf{r}$ with spatial frequency bias $\mathbf{p} = (\phi, p)$. Figure 27 shows the form of such a filter in case the feedforward filter is tuned to $\phi = 0$. It will be seen that the patterned feedback found in the model appears to be consistent with that observed by Murphy *et al.* (1999) in that it depends on the orientation preference of its V1 origin.

8 Cross–orientation suppression

A consistent experimental finding is that when a hypercolumn is stimulated with a pair of orthogonal gratings or bars there is considerable suppression of the response to either stimulus. In particular, DeAngelis, Robson, Ohzawa, & Freeman (1992) show that this cross–orientation suppression originates within the receptive field of most cat neurons examined, and is a consistent finding in both simple and complex cells. Here we present a possible cortical mechanism for cross–orientation suppression, based on the idea that the local circuits of a hypercolumn amplify the first spherical harmonic components of a stimulus. Consider a stimulus consisting of the sum of two gratings with identical spatial frequency p_s and distinct orientations Φ and Φ' respectively:

$$i(\mathbf{r}) = \frac{C_s}{2}\cos\left(p_s[x\cos\Phi + y\sin\Phi]\right) + \frac{C_s}{2}\cos\left(p_s[x\cos\Phi' + y\sin\Phi']\right) \tag{8.15}$$

When this stimulus is filtered by the receptive field (5.5), the resulting LGN input is (neglecting spatial phase)

$$\begin{aligned} h_{LGN} &= \frac{C_s}{2}\exp\left[-\frac{A^2p_s^2}{2p^2}\left(\kappa^{-2}\cos^2(\Phi-\phi)+\sin^2(\Phi-\phi)\right)\right] \\ &+ \frac{C_s}{2}\exp\left[-\frac{A^2p_s^2}{2p^2}\left(\kappa^{-2}\cos^2(\Phi'-\phi)+\sin^2(\Phi'-\phi)\right)\right] - C_s\alpha\exp\left[-\frac{\hat{\kappa}^2A^2p_s^2}{2p^2}\right] \\ &= \frac{C_s}{2}\sum_{n=0}^{\infty}U_n(p_s|p)\left[\cos 2n(\Phi-\phi)+\cos 2n(\Phi'-\phi)\right] \end{aligned} \tag{8.16}$$

where we have used equations (6.7) and (6.8). If we now project out the first order spherical harmonic components we obtain an effective LGN input of the form

$$h(\theta,\phi) = C(1-\epsilon) + C\epsilon\left[\cos\Theta\cos\theta + \sin\Theta\sin\theta\frac{\cos(2[\phi-\Phi])+\cos(2[\phi-\Phi'])}{2}\right] \tag{8.17}$$

where $\Theta = \mathcal{Q}(p_s)$ (assuming some form of renormalization along the lines of §7).

Suppose that $\Theta \neq 0, \pi$. If $\Phi' = \Phi$ then we recover the case of a single grating with

$$h(\theta,\phi) = C(1-\epsilon) + C\epsilon\left[\cos\Theta\cos\theta + \sin\Theta\sin\theta\cos(2[\phi-\Phi])\right] \tag{8.18}$$

so that the peak cortical response is at $\theta = \Theta$ and $\phi = \Phi$. On the other hand, in the case of an orthogonal grating, $\Phi' = \Phi + \pi/2$, there is exact cancellation of cosines such that

$$h(\theta,\phi) = C(1-\epsilon) + C\epsilon\cos\Theta\cos\theta \tag{8.19}$$

and the maximal cortical response will occur at either the low frequency pinwheel $\theta = 0$ (when $\cos\Theta > 0$) or the high frequency pinwheel $\theta = \pi$ (when $\cos\Theta < 0$). More generally, we can rewrite equation (8.17) as

$$h(\theta,\phi) = \overline{C}(1-\overline{\epsilon}) + \overline{\epsilon}\overline{C}\left[\cos\overline{\Theta}\cos\theta + \sin\overline{\Theta}\sin\theta\cos(2[\phi-\overline{\Phi}])\right] \tag{8.20}$$

where

$$\overline{\Phi} = \frac{\Phi+\Phi'}{2} \tag{8.21}$$

$$\tan\overline{\Theta} = \cos(\Phi-\Phi')\tan\Theta \tag{8.22}$$

and

$$\overline{C} = C(1-\epsilon+\epsilon\kappa), \quad \overline{\epsilon} = \frac{\epsilon\kappa}{1-\epsilon+\epsilon\kappa} \tag{8.23}$$

with

$$\kappa = \sqrt{\cos^2\Theta + \sin^2\Theta\cos^2(\Phi-\Phi')} \tag{8.24}$$

Figure 28: Cross orientation suppression: shift in optimal spatial frequency of cortical response due to a pair of sinusoidal gratings with relative orientation $\Delta\Phi$ and the same spatial frequency $p_s = \mathcal{Q}^{-1}(\Theta)$. Here the shifted spatial frequency $\overline{\Theta}$ is plotted as a function of $\Delta\Phi$ for various values of Θ and $\epsilon = 0.2$. Note that the peak is shifted to lower spatial frequencies when $0 < \Theta < \pi/2$ and higher spatial frequencies when $\pi/2 < \Theta < \pi$.

Figure 29: Cross orientation suppression: reduction in LGN bias and contrast due to a pair of sinusoidal gratings with relative orientation $\Delta\Phi$ and the same spatial frequency $p_s = \mathcal{Q}^{-1}(\Theta)$. (a) Relative LGN contrast $\overline{C}/C$ is plotted as a function of $\Delta\Phi$ for various Θ. (b) Shifted LGN bias $\overline{\epsilon}$ is plotted as a function of $\Delta\Phi$ for various Θ. Here $\epsilon = 0.2$.

The variation of $\overline{\Theta}, \overline{C}$ and $\overline{\epsilon}$ with orientation difference $\Delta\Phi = \Phi - \Phi'$ is shown in figures 28 and 29. Two separate effects of cross orientation suppression can be identified. First, there is a shift in the peak response $\Theta \to \overline{\Theta}$ and $\Phi \to \overline{\Phi}$, which means that there will be *local* cortical suppression since cells that respond optimally at (Θ, Φ) will have their response suppressed. However, other cells will have their response enhanced. Second, there is a reduction in both the effective contrast and bias of the LGN input, which

implies that there is also *global* suppression due to a reduction in the cortical gain G defined by equation (3.10). Both effects increase with $\Delta\Phi$, reaching a maximum when the gratings are orthogonal. On the other hand, the degree of suppression decreases as Θ approaches one of the pinwheels. (The point $\Theta = \pi/2$ is a singular case, since there is no shift in spatial frequency with $\Delta\Phi$ but the LGN bias vanishes when $\Delta\Phi = \pi/2$).

Cross orientation suppression is also expected to occur for a checkerboard pattern, which is constructed by taking the product of two orthogonal sinusoidal gratings, one vertical and the other horizontal, say

$$\begin{aligned} i(\mathbf{r}) &= C_s \cos(p_s x)\cos(p_s y) \\ &= \frac{C_s}{2}\cos(p_s(x+y)) + \frac{C}{2}\cos(p_s(x-y)) \end{aligned} \tag{8.25}$$

Such a stimulus decomposes as a sum of two gratings at $\pm 45^0$ and effective spatial frequency $\sqrt{2}p_s$. Of course, it is possible to perceive checkerboards, crosses and other more complex stimuli. Hence, the occurrence of cross orientation suppression suggests that this is achieved at a more global level by combining the responses of many hypercolumns. It might also be possible that a hypercolumn amplifies higher order harmonic components of a stimulus, and uses this to resolve certain aspects of composite stimuli. However, as discussed by Carandini & Ringach (1997) within the context of the ring model, this could lead to the undesirable side effect of spurious peaks in the tuning curves.

9 Discussion

The main conclusion of this paper is that orientation and spatial frequency can be represented as the surface coordinates of a sphere in each region of V1 that corresponds to a Hubel–Wiesel hypercolumn. We re–emphasize that this proposed spherical or $SO(3)$ symmetry is an internal symmetry of the network topology, or equivalently, of the cortical labels for orientation and spatial frequency preferences, and is not a symmetry of the actual spatial arrangment of neurons in a hypercolumn. Such a spherical coordinate system naturally accommodates the existence of orientation preference pinwheels and their association with regions of both low and high spatial frequency preference. It follows that pattern formation on the sphere generated essentially by a symmetry breaking instability, in which the first few spherical harmonics are excited by incoming stimuli, can provide a mechanism for the existence of localized orientation and spatial frequency preferences and tuning, very much as suggested by De Valois & De Valois (1988). A major consequence of the spherical topology and its association with orientation preference pinwheels is that orientation and spatial frequency tuning curves are not separable.

The inclusion of spatial frequency preference and tuning as a property of V1 neurons is not as straightforward as in the case of orientation. If a local visual stimulus is filtered by the action of the geniculo–cortical pathway, as originally suggested by Hubel & Wiesel (1962) for the orientation preference label, then the representation of spatial frequency is not faithful. Thus figure 23 shows that the spatial frequency encoded by the hypercolumn in general differs from the actual spatial frequency of the stimulus. We conclude that in order to obtain a faithful representation of spatial frequency, we must insert another filtering operation. The most plausible and interesting possibility is that cortico–geniculate *feedback* generates such a filter. Thus we propose a role for the back–projection from V1 to the LGN: it exists, in part, to provide a means to continually update and modify the LGN input to V1 so that the representation of spatial frequency remains faithful when signalled by the projection onto the first order spherical harmonics. Interestingly, we find that the cortico–geniculate filter that results from our calculations is approximately the reciprocal of the feedforward filter. Thus it innervates the LGN in a pattern determined by the orientation and spatial frequency biases of the feedforward receptive field. We note that Murphy *et al.* (1999) found such patterns, at least in the case of orientation preferences. Recent observations by Sharon & Grinvald (2002) also appear to be consistent with the model. They found that orientation tuning is amplified during a cortical evoked

response. They also established that the time course of this amplification is not smooth but slows down about 50 msec after onset and then accelerates again. This is consistent with the postulated feedback process coming online after some 50 msec.

Our suggestion about the function of the cortico–geniculate feedback pathway differs considerably from many others, which have been concerned with such functions as gating retino–geniculate transmission, improving the precision of spike timing in LGN cells, enhancing the spatial frequency tuning of LGN cells, and synchronizing slow oscillations between V1 and LGN [see Funke, Kisvárdy, Volgushev, & Wörgötter (2001)]. Our model is closest in concept to that of Rao & Ballard (1999) who suggested that, in general, feedback connections carry *predictions* of lower level activities, whereas the corresponding feedforward connections carry the residual errors between the predictions and the actual lower level activities. It remains to determine what the connection is, if any, between our ideas and those concerning such predictive coding.

More specific results and predictions of our spherical model are as follows:

(a) Orientation preference and tuning should become weaker at low and high spatial frequencies, in part since, by hypothesis, such frequencies are located at the poles of the sphere. This is consistent with the early recordings of Hubel & Wiesel (1962), who found numerous cells with poor or no tuning for orientation, many of which they later located in CO–blob regions of V1 (Livingstone & Hubel, 1984), now known to be regions of low spatial frequency tuning (Hübener *et al.*, 1997).

(b) Spatial frequency preference shifts occur at both ends of the frequency spectrum. The direction of the shifts is always toward the high or low frequency poles. Thus low spatial frequencies tend to be signalled as even lower, high spatial frequencies as even higher. Issa *et al.* (2000) report such shifts at the low end of the spectrum. However, these authors also report high frequency shifts in the opposite direction to our predictions. We hypothesize that feedback could be responsible for the predicted change in direction of shifts in spatial frequency preferences. That is, downward shifts are consistent with the properties of the geniculo–cortical pathways before the effects of the cortico–geniculate feedback have time to act. We predict that the earliest responses of cortical neurons should all exhibit downward shifts in spatial frequency tuning, but for high spatial frequencies such shifts should eventually disappear or even reverse, when measured at later times. Some observations appear to be consistent with this prediction (Bredfeldt & Ringach, 2002).

(c) The contrast invariance of tuning curves for both orientation and spatial frequency is a natural property of our model, and indeed, of any model with an amplification mechanism. In this respect it differs somewhat from the quadratic threshold mechanism recently suggested by Miller & Troy (2002), in that even the linear rectifier defined in eqn.(2.2) will generate contrast invariant responses if there is an amplification process.

(d) Finally we remark that cross orientation suppression is also a natural property of our model that also follows from the amplification process. However the suppression mechanism is complicated. Firstly, there is a *local* suppression effect—cells that respond optimally at (Θ, Φ) will have their response suppressed by an orthogonal input. However, other cells will have their response enhanced. Secondly, there is also *global* suppression due to a reduction in the cortical gain G defined by equation (3.10). Both effects increase with $\Delta\Phi$, reaching a maximum when the gratings are orthogonal. On the other hand, the degree of suppression decreases as Θ approaches one of the pinwheels.

Appendix A

We analyze the stability of the localized state (centered at the $\Theta = 0$ pinwheel) by linearizing equation (3.4) about the fixed point solution (3.9). First, we set

$$\begin{aligned} I_0(t) &= I_0 + \gamma_0(t) \\ I_1^0(t) &= I_1 + \gamma_1(t) \\ I_0^{\pm}(t) &= \gamma_{\pm}(t) \end{aligned} \tag{A.1}$$

with I_0 and I_1 determined by the self-consistency conditions (3.30)–(3.32) and write

$$\frac{da}{dt} = -a + \left[I_0 + I_1 \cos\theta + \gamma_0(t) + \sum_{m=1,\pm} \gamma_m(t) f_m(\theta,\phi) \right]_+ \tag{A.2}$$

At a given time t, the boundary condition for the vanishing of the total synaptic drive is

$$J(\theta,\phi,t) \equiv I_1 \cos\theta + \gamma_0(t) + \sum_m \gamma_m(t) f_m(\theta,\phi) = 0 \tag{A.3}$$

This equation can be linearized by setting $\theta = \theta_c + \delta\theta(\phi,t)$ with

$$\delta\theta(\phi,t) = \frac{\gamma_0(t) + \sum_{m=1,\pm} \gamma_m(t) f_m(\theta_c,\phi)}{I_1 \sin\theta_c} \tag{A.4}$$

We have used the fact that $I_0 + I_1 \cos\theta_c = 0$. The next step is to take moments of equation (A.2) with respect to the zeroth and first order harmonics:

$$\frac{dR_0}{dt} = -R_0 + \int_0^{\pi} \int_0^{\theta_c+\delta\theta} J(\theta,\phi,t) \mathcal{D}(\theta,\phi) \tag{A.5}$$

and

$$\frac{dR_1^n}{dt} = -R_1^n \int_0^{\pi} \int_0^{\theta_c+\delta\theta} f_n(\theta,\phi) J(\theta,\phi,t) \mathcal{D}(\theta,\phi) \tag{A.6}$$

In order to linearize these equations set $R_0(t) = R_0 + r_0(t)$, $R_1^0(t) = R_1 + r_1(t)$ and $R_1^{\pm}(t) = r_{\pm}(t)$ with R_0, R_1 given by equations (3.31) and (3.32). It follows from equations (3.5), (3.6) and (3.30) that

$$\gamma_0(t) = W_0 r_0(t), \quad \gamma_m(t) = W_1 r_m(t) \tag{A.7}$$

for $m = 1, \pm$. Here $r_0(t)$ and $r_1(t)$ represent "longtitudinal" fluctuations of the localized state whereas $r_\pm(t)$ represent "transverse" fluctuations.

It turns out that the longtitudinal and transverse fluctuations decouple at the linear level. Expanding equations (A.5) and (A.6) to first order in $\delta\theta$ we find that the longtitudinal modes satisfy the pair of equations

$$\frac{dr_0}{dt} = -r_0 + \frac{1}{4}\,(2I_0 \sin\theta_c + I_1 \sin 2\theta_c)\overline{\delta\theta} + \frac{\gamma_0[1-\cos\theta_c]}{2} + \frac{\gamma_1[1-\cos 2\theta_c]}{8} \tag{A.8}$$

and

$$\frac{dr_1}{dt} = -r_1 + \frac{1}{4}(I_0 \sin 2\theta_c + 2I_1 \cos^2\theta_c \sin\theta_c)\overline{\delta\theta} + \frac{\gamma_0[1-\cos 2\theta_c]}{8} + \frac{\gamma_1[1-\cos^3\theta_c]}{6} \tag{A.9}$$

where

$$\overline{\delta\theta} \equiv \frac{1}{\pi}\int_0^\pi \delta\theta(\phi)d\phi = \frac{\gamma_0 + \gamma_1\cos\theta_c}{I_1 \sin\theta_c} \tag{A.10}$$

Equation (3.29) then implies that the coefficients multiplying $\overline{\delta\theta}$ actually vanish. Thus we have the simple matrix equation

$$\frac{d}{dt}\begin{pmatrix} r_0 \\ r_1 \end{pmatrix} = \mathbf{W}(\theta_c)\begin{pmatrix} r_0 \\ r_1 \end{pmatrix} \tag{A.11}$$

where

$$\mathbf{W}(\theta_c) = \begin{pmatrix} -1 + \dfrac{W_0[1-\cos\theta_c]}{2} & \dfrac{W_1[1-\cos 2\theta_c]}{8} \\ \dfrac{W_0[1-\cos 2\theta_c]}{8} & -1 + \dfrac{W_1[1-\cos^3\theta_c]}{6} \end{pmatrix} \tag{A.12}$$

Hence, we obtain an eigenvalue equation of the form

$$\lambda^2 - \lambda \mathrm{Tr}\mathbf{W}(\theta_c) + \det\mathbf{W}(\theta_c) = 0 \tag{A.13}$$

One can show that the localized state undergoes an amplitude instability as W_0 is increased for fixed W_1 due to a single real eigenvalue becoming positive. The condition for such an instability is $\det\mathbf{W}(\theta_c) = 0$. In the particular case of a homogeneous input ($\epsilon = 0$), one finds from equations (3.26) and (3.27) that

$$\det\mathbf{W}(\theta_c) = -\frac{W_1}{8}[1-\cos 2\theta_c][\cos\theta_c + W_0\mathcal{A}_0(\theta_c)] \tag{A.14}$$

so that the condition for an amplitude instability is

$$W_1 = W_c \equiv -\frac{\cos\theta_c}{\mathcal{A}_0(\theta_c)} \tag{A.15}$$

Finally, expanding equation (A.6) to first order in $\delta\theta$ for the transverse modes, we find that

$$\frac{dr_\pm}{dt} = -r_\pm + \frac{1}{2}(I_0 \sin^2\theta_c + I_1 \sin^2\theta_c \cos\theta_c)\overline{\delta\theta_\pm} + \gamma_\pm W_1 \mathcal{A}_1(\theta_c) r_\pm \tag{A.16}$$

where

$$\overline{\delta\theta_\pm} \equiv \frac{1}{\pi}\int_0^\pi \begin{pmatrix} \cos\theta_c \\ \sin\theta_c \end{pmatrix} \delta\theta(\phi) d\phi = \frac{W_1 r_\pm}{2I_1} \tag{A.17}$$

Equation (3.29) implies that the coefficient multiplying $\overline{\delta\theta_\pm}$ vanishes. Hence

$$\frac{dr_\pm}{dt} = \left[-1 + W_1 \mathcal{A}_1(\theta_c)\right] r_\pm \tag{A.18}$$

It immediately follows that in the case of a homogeneous input, the localized state is marginally stable with respect to excitation of the transverse modes since $1 = W_1 \mathcal{A}_1(\theta_c)$.

Appendix B

In order to simplify our analysis of the spherical model, we collapsed the inhibitory and excitatory cell populations into a single equivalent population. Such a simplification greatly reduces the number of free parameters of the system. However, the basic insights gained from the one population model can now be used to develop the mean field theory of a more realistic two population model.

Let us denote the activity of the excitatory (e) and inhinitory (i) populations by $a_r(\theta, \phi, t)$ with $r = e, i$. A two population version of equation (2.2) is then

$$\frac{\partial a_r(\theta,\phi,t)}{\partial t} = -a_r(\theta,\phi,t) + \beta_r \left[I_r(\theta,\phi,t) - \kappa_r\right]_+ \tag{B.1}$$

where κ_r is the threshold and $I_r(\theta, \phi, t)$ is the total synaptic current of the rth population,

$$I_r(\theta,\phi,t) = h_r(\theta,\phi) + \sum_{s=e,i}\int_{\mathbf{S}^2} w_{rs}(\theta,\phi|\theta',\phi') a_s(\theta',\phi',t)\mathcal{D}(\theta',\phi') \tag{B.2}$$

We have also introduced input gains β_r The weight distributions w_{rs} connecting the various cell populations are taken to be $SO(3)$ invariant, and are constructed out of zeroth and first-order spherical harmonics according to

$$w_{ee}(\theta,\phi|\theta',\phi') = w_{ie}(\theta,\phi|\theta',\phi') = W_{e,0} + W_{e,1}\sum_{m=0,\pm} f_m(\theta,\phi)f_m(\theta',\phi') \tag{B.3}$$

and

$$w_{ei}(\theta,\phi|\theta',\phi') = w_{ii}(\theta,\phi|\theta',\phi') = -W_{i,0} - W_{i,1}\sum_{m=0,\pm} f_m(\theta,\phi)f_m(\theta',\phi') \tag{B.4}$$

with $W_{r,1}, W_{r,0} \geq 0$. As in the one population model, the weakly biased LGN input $h_r(\theta,\phi)$ is assumed to be of the form

$$h_r(\theta,\phi) = \alpha_r h(\theta,\phi) \tag{B.5}$$

with h given by equation (2.10 and α_r determining the relative strength of the input to the two populations.

Introducing the order parameters

$$R_{r,0}(t) = \int_{\mathbf{S}^2} a_r(\theta,\phi,t)\mathcal{D}(\theta,\phi) \tag{B.6}$$

$$R^m_{r,1}(t) = \int_{\mathbf{S}^2} a_r(\theta,\phi,t)f_m(\theta,\phi)\mathcal{D}(\theta,\phi) \tag{B.7}$$

equation (B.1) can be rewritten in the form

$$\frac{\partial a_r(\theta,\phi,t)}{\partial t} = -a_r(\theta,\phi,t) + \left[I_{r,0}(t) + \sum_{m=0,\pm} I^m_{r,1}(t)f_m(\theta,\phi)\right]_+$$

where

$$I_{r,0} = \beta_r\left[\alpha_r C(1-\epsilon) - \kappa + W_{e,0}R_{e,0} - W_{i,0}R_{i,0}\right] \tag{B.8}$$

and

$$I^m_{r,1} = \beta_r\left[\alpha_r C\epsilon f_m(\Theta,\Phi) + W_{e,1}R^m_{e,1} - W_{i,1}R^m_{i,1}\right] \tag{B.9}$$

We focus on a fixed point solution of equation (B.8) in which both the excitatory and inhibitory populations are in a stationary localized state of the form

$$a_r(\theta,\phi) = I_{r,1}\left[\sum_{m=0,\pm} f_m(\Theta,\Phi)f_m(\theta,\phi) - \cos\theta_{c,r}\right]_+ \tag{B.10}$$

Taking moments of equation (B.8) with respect to the zeroth and first order spherical harmonics and proceeding along identical lines to the one population model (see §3), one finds that such a solution exists provided that

$$R^m_{r,1} = R_{r,1} f_m(\Theta, \Phi), \quad I^m_{r,1} = I_{r,1} f_m(\Theta, \Phi), \tag{B.11}$$

with

$$R_{r,0} = I_{r,1} \mathcal{A}_0(\theta_{c,r}), \quad R_{r,1} = I_{r,1} \mathcal{A}_1(\theta_{c,r}) \tag{B.12}$$

and

$$I_{r,0} + I_{r,1} \cos\theta_{c,r} = 0 \tag{B.13}$$

The functions $\mathcal{A}_0$, $\mathcal{A}_1$ are defined by equations (3.26) and (3.27). Combining equations (B.11) and (B.12) with equations (B.8) and (B.9) leads to the mean–field equations

$$R_{r,0} = \beta_r \left[\alpha_r C\epsilon + W_{e,1} R_{e,1} - W_{i,1} R_{i,1}\right] \mathcal{A}_0(\theta_{c,r}) \tag{B.14}$$

and

$$R_{r,1} = \beta_r \left[\alpha_r C\epsilon + W_{e,1} R_{e,1} - W_{i,1} R_{i,1}\right] \mathcal{A}_1(\theta_{c,r}) \tag{B.15}$$

Marginal phase ($\epsilon = 0$) In the case of a homogeneous input, equation (B.15) reduces to the matrix equation

$$\mathbf{W}(\theta_{c,e}, \theta_{c,i}) \begin{pmatrix} R_{e,1} \\ R_{i,1} \end{pmatrix} = 0 \tag{B.16}$$

where

$$\mathbf{W}(\theta_{c,e}, \theta_{c,i}) = \begin{pmatrix} -1 + \beta_e W_{e,1} \mathcal{A}_1(\theta_{c,e}) & -\beta_e W_{i,1} \mathcal{A}_1(\theta_{c,e}) \\ \beta_i W_{e,1} \mathcal{A}_1(\theta_{c,i}) & -1 - \beta_i W_{i,1} \mathcal{A}_1(\theta_{c,i}) \end{pmatrix} \tag{B.17}$$

A necessary condition for the existence of a non–trivial localized state can then be expressed as $\det \mathbf{W}(\theta_{c,e}, \theta_{c,i}) = 0$, that is,

$$\beta_e W_{e,1} \mathcal{A}_1(\theta_{c,e}) - \beta_i W_{i,1} \mathcal{A}_1(\theta_{c,i}) = 1 \tag{B.18}$$

Acknowledgments

We wish to thank Prof. Martin Golubitsky and Dr. Peter Thomas for useful comments, Mr. Michael Buice for mathematical proof reading and Miss Tanya Baker for help with figure 5. We also thank the referees for their thorough reading of the paper and their many helpful suggestions. This work was supported in part by grant 96–24 from the James S. McDonnell Foundation to JDC.

References

Angelucci, A., Levitt, J. B., & Lund, J. S., 2001. Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Progress in Brain Research (in press)* .

Arfken, G., 1985. *Mathematical methods for physicists, 3rd. edition*. Academic Press: San Diego.

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H., 1995. Theory of orientation tuning in visual cortex. *Proc. Nat. Acad. Sci.* **92**: 3844–3848.

Ben-Yishai, R., Hansel, D., & Sompolinsky, H., 1997. Traveling waves and the processing of weakly tuned inputs in a cortical network module. *J. Comput. Neurosci.* **4**: 57–77.

Blasdel, G. G., 1992. Orientation selectivity, preference, and continuity in monkey striate cortex. *J. Neurosci.* **12**: 3139–3161.

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

Bonhoeffer, T. & Grinvald, A., 1991. Orientation columns in cat are organized in pinwheel like patterns. *Nature* **364**: 166–146.

Bonhoeffer, T., Kim, D. S., Malonek, D., Shoham, D., & Grinvald, A., 1995. Optical Imaging of the layout of functional domains in area 17/18 border in cat visual cortex. *European J. Neurosci.* **7** (9): 1973–1988.

Bredfeldt, C. E. & Ringach, D. L., 2002. Dynamics of spatial frequency tuning in macaque V1. *J. Neurosci.* **22**: 1976–1984.

Bressloff, P. C., Bressloff, N. W., & Cowan, J. D., 2000. Dynamical mechanism for sharp orientation tuning in an integrate–and–fire model of a cortical hypercolumn. *Neural Comput.* **12**: 2473–2511.

Bressloff, P. C. & Cowan, J. D., 2002a. An amplitude equation approach to contextual effects in primary visual cortex. *Neural Comput.* **14**: 493–525.

Bressloff, P. C. & Cowan, J. D., 2002b. An SO(3) symmetry breaking mechanism for orientation and spatial frequency tuning in visual cortex. *Phys. Rev. Lett.* **88**: 078102.

Carandini, M. & Ringach, D., 1997. Predictions of a Recurrent Model of Orientation Selectivity. *Vision Res.* **37** (21): 3061–3071.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G., 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research* **22**: 545–559.

De Valois, R. L. & De Valois, K. K., 1988. *Spatial Vision*. Oxford: Oxford University Press.

DeAngelis, G. C., Robson, J. G., Ohzawa, I., & Freeman, R. D., 1992. Organization of Suppression in Receptive Fields of Neurons in Cat Visual Cortex. *J. Neurophysiol.* **68** (1): 144–163.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., & Suarez, H. H., 1995. Recurrent excitation in neocortical circuits. *Science* **269**: 981–985.

Dragoi, V. & Sur, M., 2000. Some properties of recurrent inhibition in primary visual cortex: contrast and orientation dependence on contextual effects. *J. Neurophysiol.* **83**: 1019–1030.

Ferster, D., Chung, S., & Wheat, H., 1997. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**: 249–281.

Funke, K., Kisvárdy, Z. F., Volgushev, M., & Wörgötter, F., 2001. Integrating Anatomy and Physiology of the Primary Visual Pathway: From LGL to Cortex. In *Models of Neural Networks IV, Early Vision and Attention*, (eds. J. L. van Hemmen, J. D. Cowan, & E. Domany), pp. 97–182. Springer.

Guillery, R. W., Feig, S. L., & van Lieshout, D. P., 2001. Connections of higher order visual relays in the thalamus: a study of corticothalamic pathways in cats. *J. Comp. Neurol.* **438**: 66–85.

Hansel, D. & Sompolinsky, H., 1997. Modeling feature selectivity in local cortical circuits. In *Methods of Neuronal Modeling*, (eds. C. Koch & I. Segev), pp. 499–567. Cambridge: MIT Press, 2nd edition.

Hawken, M. J. & Parker, A. J., 1987. Spatial properties of neurons in the monkey striate cortex. *Proc. Roy. Soc. B* **231**: 251–288.

Horton, J. C. & Hubel, D. H., 1981. Regular patchy distribution of cytochrome oxidase staining in primary visual cortex of macaque monkey. *Nature* **292**: 762–764.

Hubel, D. H. & Wiesel, T. N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Neurosci.* **3**: 1116–1133.

Hubel, D. H. & Wiesel, T. N., 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. Lond.* **195**: 215–243.

Hubel, D. H. & Wiesel, T. N., 1974. Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *J. Comp. Neurol.* **158**: 295–306.

Hubel, D. H. & Wiesel, T. N., 1977. Functional architecture of macaque monkey visual cortex. *Proc. Roy. Soc. B* **198**: 1–59.

Hubel, D. H., Wiesel, T. N., & Stryker, M. P., 1978. Anatomical demonstration of orientation columns in macaque monkey. *J. Comp. Neurol.* **177**: 361–380.

Hübener, M., Shoham, D., Grinvald, A., & Bonhoeffer, T., 1997. Spatial Relationships among Three Columnar Systems in Cat Area 17. *J. Neurosci.* **17** (23): 9270–9284.

Issa, N. P., Trepel, C., & Stryker, M. P., 2000. Spatial frequency maps in cat visual cortex. *J. Neurosci.* **20**: 8504–8514.

Jones, J. P. & Palmer, L. A., 1987. An evaluation of the two–dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58** (6): 1233–1258.

Kelly, D. H. & Magnuski, H. S., 1975. Pattern detection and the two–dimensional Fourier transform: Circular targets. *Vision Res.* **15**: 911–915.

LeVay, S. & Nelson, S. B., 1991. Columnar organization of the visual cortex. In *The neural basis of visual function*, (ed. A. G. Leventhal), pp. 266–315. Boca Raton : CRC Press.

Li, Z., 1999. Pre–attentive segmentation in the primary visual cortex. *Spatial Vision* **13**: 25–39.

Livingstone, M. S. & Hubel, D. H., 1984. Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* **4**: 309–356.

Maldonado, P. E., Gödecke, I., Gray, C. M., & Bonhoeffer, T., 1997. Orientation selectivity in pinwheel centers in cat striate cortex. *Science* **276**: 1551–1555.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L., 2002. Spatial frequency and orientation tuning dynamics in V1. *Proc. Nat. Acad. Sci.* **99**: 1645–1650.

McLaughlin, D., Shapley, R., Shelley, M., & Wielaard, D. J., 2000. A Neuronal Network Model of Macaque Primary Visual Cortex (V1): Orientation Tuning and Dynamics in the Input Layer 4Cα. *Proc. Natl. Acad. Sci.* **97**: 8087–8092.

Miller, K. D. & Troy, T. W., 2002. Neural Noise Can Explain Expansive, Power–law Nonlinearities in Neural Response Functions. *J. Neurophysiol.* **87**: 653–659.

Mundel, T., Dimitrov, A., & Cowan, J. D., 1997. Visual cortex circuitry and orientation tuning. In *Advances in Neural Information Processing Systems*, (eds. M. C. Mozer, M. I. Jordan, & T. Petsche), volume 9, pp. 886–893. MIT Press.

Murphy, K., Jones, D. G., & Sluyters, R. C. V., 1995. Cytochrome oxidase blobs in cat primary visual cortex. *J. Neurosci.* **15**: 4196–4208.

Murphy, P. C., Duckett, S. G., & Sillito, A. M., 1999. Feedback connections to the lateral geniculate nucleus and cortical response properties. *Science* **286**: 1552–1554.

Nelson, S., Toth, L., Seth, B., & Sur, M., 1994. Orientation selectivity of cortical neurons during extra-cellular blockade of inhibition. *Science* **265**: 774–777.

Obermayer, K. & Blasdel, G., 1993. Geometry of orientation and ocular dominance columns in monkey striate cortex. *J. Neurosci.* **13**: 4114–4129.

O'Keefe, L. P., Levitt, J. B., Kiper, D. C., Shapley, R. M., & Movshon, J. A., 1998. Functional Organization of Owl Monkey Lateral Geniculate Nucleus and Visual Cortex. *J. Neurophysiol.* **80**: 594–609.

Rao, R. P. N. & Ballard, D. H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra–classical receptive–field effects. *Nature Neurosci.* **2** (1): 79–87.

Reid, R. C. & Alonso, J. M., 1995. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **378**: 281–284.

Roerig, B. & Chen, B., 2002. Relationships of local inhibitory and excitatory circuits to orientation preference maps in ferret visual cortex. *Cerebral Cortex* **12**: 187–198.

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

Sillito, A. M., 1975. The contribution of inhibitory mechanisms to the receptive–field properties of neurones in the striate cortex of the cat. *J. Physiol. Lond.* **250**: 305–329.

Somers, D. C., Nelson, S., & Sur, M., 1995. An Emergent Model of Orientation Selectivity in Cat Visual Cortical Simple Cells. *J. Neurosci.* **15**: 5448–5465.

Somers, D. C., Todorov, E. V., Siapas, A. G., Toth, L. J., Kim, D.-S., & Sur, M., 1998. A local circuit approach to understanding integration of long–range inputs in primary visual cortex. *Cerebral Cortex* **8**: 204–217.

Stetter, M., Bartsch, H., & Obermayer, K., 2000. A mean field model for orientation tuning, contrast saturation, and contextual effects in the primary visual cortex. *Biol. Cybernetics* **87**: 291–304.

Swindale, N. V., 1996. The development of topography in visual cortex: a review of models. *Network* **7**: 161–247.

Tootell, R. B. H., Silverman, M. S., & De Valois, R. L., 1981. Spatial frequency columns in primary visual cortex. *Science* **214**: 813–815.

Vidyasagar, T. R., Pei, X., & Volgushev, M., 1996. Multiple mechanisms underlying the orientation selectivity of visual cortical neurons. *TINS* **19** (7): 272–277.

Webster, M. A. & De Valois, R. L., 1985. Relationship between spatial–frequency and orientation tuning of striate–cortex cells. *J. Opt. Soc. Am. A* **2**: 1124–1132.

Wiesel, T. N., Hubel, D. H., & Lam, D. M. K., 1974. Autoradiographic demonstration of ocular–dominance columns in the monkey striate cortex by means of transneuronal transport. *Brain Res.* **79**: 273–279.