The role of early retinal lateral inhibition: More than maximizing luminance information

ROSARIO M. BALBOA\(^1\) AND NORBERTO M. GRZYWACZ\(^2\)
Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco
(RECEIVED March 15 1999; ACCEPTED August 23, 1999)

Abstract
Lateral inhibition is one of the first and most important stages of visual processing. There are at least four theories related to information theory in the literature for the role of early retinal lateral inhibition. They are based on the spatial redundancy in natural images and the advantage of removing this redundancy from the visual code. Here, we contrast these theories with data from the retina’s outer plexiform layer. The horizontal cells’ lateral-inhibition extent displays a bell-shape behavior as function of background luminance, whereas all the theories show a fall as luminance increases. It is remarkable that different theories predict the same luminance behavior, explaining “half” of the biological data. We argue that the main reason is how these theories deal with photon-absorption noise. At dim light levels, for which this noise is relatively large, large receptive fields would increase the signal-to-noise ratio through averaging. Unfortunately, such an increase at low luminance levels may smooth out basic visual information of natural images. To explain the biological behavior, we describe an alternate hypothesis, which proposes that the role of early visual lateral inhibition is to deal with noise without missing relevant clues from the visual world, most prominently, the occlusion boundaries between objects.

Keywords: Information Theory, Natural images, Contrast estimation, Horizontal cells, Edge localization

Introduction
What does the early visual system extract from natural images? It is reasonable to assume that the early visual system optimally extracts information for the later visual stages (Atick & Redlich, 1990) to perform visual inferences and tasks that are important for the animal (behavior). Such an optimality must take into account the regularities of the natural scene and the neural limitations of the performed computations. Here we analyze these regularities and neural constraints in the context of the lateral inhibition mediated by the horizontal cells in the outer plexiform layer (OPL) of the vertebrate retina. Lateral inhibition is one of the first stages of visual processing (see for a review Dowling, 1987), and its presence in Limulus (Ratliff, 1965) shows that lateral inhibition has existed for at least 400 million years. In the OPL, lateral inhibition is mediated by synaptic action from horizontal cells onto photoreceptors and the bipolar cells, providing information about the response in distant photoreceptors.

Ratliff (1965) proposes that the main functional role of early lateral inhibition is aiding in the detection of sharp image transitions in space or time. Temporally transient responses are essential for vision (Barlow, 1965; Yarbus, 1967). And lateral-inhibition-mediated responses may deal with optical blur to enhance signals due to edges or contours, or compute contrast in the image. However, Ratliff does not attempt to prove that lateral inhibition is optimal for these tasks. To do so, he would have to show that it maximizes the type of information of interest to the animal.

There are theories in the literature related to information theory that could explain the goal of early lateral inhibition. Each of these theories can be thought to focus on a neural limitation and a useful property of natural images used by the visual system. We begin by highlighting four well-established theories: (1) The predictive coding theory (Srinivasan et al., 1982) proposes that the main neural limitation is the narrow dynamic range of interneurons (the bipolar cell in the OPL) and the relevant property of natural images is the autocorrelation function.\(^3\) (2) The maximal decorrelation theory (Atick & Redlich, 1992) proposes that the main neural limitation is the amount of information that can be handled at the later stages of visual processing. Thus, the goal of lateral inhibition would be to simplify the signal delivered to these stages. Following Barlow (Barlow, 1961a,b, 1989), the simplification that Atick and Redlich propose is a signal without second-order correlation at the outputs. (3) The neural limitation in the maximal signal-to-noise ratio theory (McCarthy & Owen, 1996) is the noisy phototransduction process.

\(^3\)The code in this theory would be the difference between the photoreceptor response and the response predicted by neighboring photoreceptors, resulting in light-intensity compression at the interneurons’ output. The size of the predicted neighborhood would depend on the image’s autocorrelation function.
In this theory, the goal of lateral inhibition is to emphasize the relevant signal, filtering out the irrelevant noise. In the authors' words, the goal is to extract discrepancies due to noise from expectations of the environment, which are useful for the animal behavior. Both, the maximal decorrelation and the maximal signal-to-noise ratio theories, use as the relevant property of natural images the spatial power spectrum's fall like $1/f^2$, where $f$ is spatial frequency (Field, 1987). (4) The maximal kurtosis theory is an adaptation of ideas by Field (1994) to account for cortical receptive fields. According to this adaptation, the main neural limitation is the small number of neurons carrying information to later visual-processing stages. The relevant property of natural images is that the distribution of their contrasts present large tails and thus large kurtosis. By maximizing kurtosis, the bipolar cell would carry only the most salient visual signal.

These theories were developed to gain insight into different visual-processing stages. The original application of the predictive coding theory (Srinivasan et al., 1982) was for insect vision. The maximal decorrelation theory (Attick & Redlich, 1992) was developed for retinal ganglion cells and provided good fits to human psychophysics. And as mentioned above, the maximal kurtosis theory (Field, 1994) was developed for cortex. Only the maximal signal-to-noise ratio theory was directly applicable to the OPL. Here, without belittling the original relevance of all these theories, we ask whether they are applicable to the OPL, the earliest stage of lateral visual processing. We do so since one cannot trivially reject their principles as the most basic ones for early vision.

In Balboa's doctoral thesis (1997) and in the Results section of this paper, we show that these theories are not consistent with adaptation in the retina's OPL. Hence, we have recently proposed an alternate theory to mitigate this problem. The neural limitation in this theory, which we called the minimal-local asperity hypothesis, is quantal noise. The relevant property of natural images for this hypothesis is the probability that an occluding edge passes through a point given the contrast there (Balboa & Grzywacz, 1999c). In this hypothesis, the goal of the OPL's adaptation is to minimize two terms of error: one that is due to quantal noise and which is reduced by increasing the lateral-inhibition extent, and another that is due to straddling of occlusion boundaries by lateral inhibition and which is reduced by reducing the lateral-inhibition extent.

To test the minimal-local asperity hypothesis and the other four theories described above, we use biological results on light adaptation of the lateral inhibition mediated by horizontal cells. We focus on these cells because they are the first stage of lateral processing and thus affect all later computations. At low background luminance, the spatial extent of the lateral inhibition mediated by horizontal cells appears to increase with luminance (Mangel & Dowling, 1985; Xin & Bloomfield, 1999). Then, this extent appears to stabilize at intermediate light levels and to fall with luminance at higher levels (Baldridge & Ball, 1991; Myhr et al., 1994; Lankheet et al., 1996). The fall bell-shape behavior of the inhibitory extent with luminance has been observed in both receptive-field mappings (Fig. 1) and tracer-coupling studies of horizontal cells in fish (Baldridge, 1993; Jamieson, 1994) and rabbit (Xin & Bloomfield, 1999). The key independent variable controlling this behavior appears to be the temporally averaged background luminance (Baldridge, 1993; Jamieson, 1994). To a large extent, this control operates through the modulation of gap junctions between horizontal cells (Mangel & Dowling, 1985; Weiler & Akopian, 1992; Myhr et al., 1994). Why did evolution create such a relatively complex control mechanism, that is, why have a bell-shape behavior? Is this behavior related to the maximization of luminance information as postulated by the theories above?

In this paper, we compare the predictions of the five theories described above and the bell-shape behavior of horizontal cells. Our analysis of the first four theories and our hypothesis appeared previously in abstract form (Balboa et al., 1997; Balboa & Grzywacz, 1998). The mathematical details of the minimal-local asperity hypothesis are presented elsewhere (Balboa & Grzywacz, 1999b). Here, we are not going to describe these details and the results from that work will be only summarized. Finally, we are not going to show the results for the maximal signal-to-noise ratio theory, since they are similar (though not identical) to those of the maximal decorrelation theory (Balboa, 1997).

**Formulation of the theories**

This section summarizes the predictive coding theory (Srinivasan et al., 1982), maximal decorrelation theory (Attick & Redlich, 1992),

![Graph](image)

**Fig. 1.** Electrophysiological measurements of receptive-field size for two types of horizontal cells in the rabbit retina. Type A horizontal cells (HA; empty circles). Type B horizontal cells (HB; filled circles). In darkness ($-\infty$), the receptive-field extension is small. It rises gradually with intensity until a maximum, after which the receptive-field size starts to fall, returning almost to the size at the dimmest intensities. Although the details are different for these two types of cells, the general intensity-dependent behavior is the same. Graph adapted from data from Xin and Bloomfield (1999).
maximal kurtosis theory (Field, 1994), and the minimal-local asperity hypothesis (Balboa & Grzywacz, 1999b) for the role of lateral inhibition. Appendix A presents the methods used for the computer simulations of the predictive coding, maximal decorrelation, and maximal kurtosis theories. In addition, this appendix presents the procedure to extract the lateral-inhibition extent from the simulated filters. The methods for the computer simulations of the maximal-local asperity hypothesis appear elsewhere (Balboa & Grzywacz, 1999b).

Predictive coding theory

The goal of this theory is to encode light intensities such that their wide natural range fits in the narrow dynamic range of the second-order neurons. The encoded signal sent to later stages of visual processing is the difference between the intensity in one point and the intensity predicted from neighbor points (Srinivasan et al., 1982). The theory uses a linear-regression model (Papoulis, 1965) to calculate the optimal weight and extent of the neighborhood (the lateral inhibition's shape). This neighborhood depends on the image's autocorrelation function. To compute this neighborhood, one begins with the photoreceptor matrix. If the matrix that models the photoreceptors layer has size $(2n + 1) \times (2n + 1)$, then the predicted signal in the central point is

$$\hat{x}_{0,0} = \sum_{i=-n}^{n} \sum_{j=-n}^{n} h_{i,j} V_{i,j},$$

where $h_{0,0} = 0$ and $h_{i,j}$ is the contribution of the response $V_{i,j}$ of the photoreceptor $(i,j)$ to the prediction. The linear system that must be solved to obtain the filter $h$ has $(2n + 1)^2 - 1$ equations, with the $(i + j)$th equation $(i \neq 0$ and $j \neq 0)$ being

$$\sum_{i=-n}^{n} \sum_{j=-n}^{n} R_{i,j;k,l} h_{k,l} = R_{0,0;i,j},$$

where $R_{i,j;k,l}$ is the correlation coefficient between the $(i,j)$ and $(k,l)$ positions. Srinivasan et al. (1982) define the correlation function as

$$R_{i,j;k,l} = \langle I_{i,j} I_{k,l} \rangle_{\text{Images}},$$

where $I_{i,j}$ is the intensity at the $(i,j)$ position and $\langle \rangle$ is the average across the population of natural images. They assume $I_{i,j}$ and $V_{i,j}$ to be proportional, and thus, it does not matter whether we use one or the other to obtain $h_{i,j}$. The reason is that multiplying $R_{i,j;k,l}$ and $R_{0,0;i,j}$ by the same proportionality constant does not change eqn. (2). Srinivasan et al. further assumed the correlation between two points to be translationally and rotationally invariant. Translational invariance is justified if one thinks that no point in the image is more important than others. Rotational invariance is assumed for the sake of simplicity and because it leads to isotropic receptive fields, which are not ruled out by existing evidence. With these assumptions,

$$R_{i,j;i,j} = I^2 + S^2 + N^2, \quad (i,j) = (k,l),$$

$$R_{i,j;k,l} = I^2 + S^2 F(r), \quad (i,j) \neq (k,l),$$

where $r = \sqrt{(i-k)^2 + (j-l)^2}$ is the distance between photoreceptors, $I$ is the mean intensity, $S$ is the standard deviation of the image intensities, $N$ is the photoreceptor noise (standard deviation), and $F$ is the model for the autocorrelation function, which falls with distance. Srinivasan et al. (1982) used $F(r) = e^{-(r/\lambda)}$, with $\lambda$ being a space constant.

After some algebra to eliminate the dependent parameters in the theory, the autocorrelation function becomes

$$R_{i,j;i,j} = 1 + \Omega^2 + \frac{\nu^2}{I}, \quad (i,j) = (k,l),$$

$$R_{i,j;k,l} = 1 + \Omega^2 F(r), \quad (i,j) \neq (k,l),$$

where we absorb $I^2$ into $R_{i,j;k,l}$ (this is possible because doing so does not change eqn. (2)), $\Omega = S/I$ is the statistical contrast of the image, and $\nu = N^2/I$ corresponds to the photoreceptor noisiness (the larger the $\nu$, the more noisy the photoreceptor). Consistently with physiology (Fuortes & Yeandle, 1964; Lillywhite, 1977; Baylor et al., 1979; Grzywacz et al., 1988), eqn. (6) assumes the photoreceptor noise to be Poisson and thus to have a variance proportional to the intensity.

To analyze the behavior of the theory with the shape of the autocorrelation function, one must consider several realistic forms for it. One can model the autocorrelation function starting with its Fourier transform, the power spectrum $(\hat{F}(f))$. The mean power spectra of natural images falls approximately as $f^{-p}$, where $f$ is the two-dimensional spatial frequency (Ruderman, 1997; Balboa & Grzywacz, 1999a). The mean of the exponent $p$ is approximately 2 (Field, 1987). However, this exponent has wide variability across natural images (Tolhurst et al., 1992; Balboa & Grzywacz, 1999a) and the function $f^{-p}$ has a singularity at zero frequency. To eliminate the singularity and explore the effect of this variability on the predicted lateral-inhibition profile, we use the following power-spectrum function:

$$\hat{F}(f) = \frac{B}{(C^2 + f^2)^{p/2}},$$

where $B$, $C$, and $p \geq 2$ are positive parameters.

Maximal decorrelation theory

The goal of this theory is to compute a neurally realistic filter whose output is maximally decorrelated in space. (A temporal version of the theory also exists (Dan et al., 1996) but it is outside the scope of this paper). Atick and Redlich (1992) decompose their theory into two stages of retinal computation. The first corresponds to noise filtering and the second to spatial-redundancy reduction. Before the first signal-processing stage, the quality of image encoding is compromised by the optic media and photoreceptor noise. The resulting code of the first stage is filtered to reduce the quantal noise. In the second stage, Atick and Redlich use Field's result on the log-log slope of the fall of the power spectrum from natural images (Field, 1987). They use this slope and calculate of variations to compute the filter that reduces maximally the correlation at the output. The filter's Fourier transform for the maximal decorrelation theory is

$$\hat{h}(f) = \frac{M(f) M}{(f^2 M^2 (f)^2 + N^2) + N_0}^{1/2},$$

where $M(f)$ is the power spectrum of the image, $\hat{h}(f)$ is the Fourier transform of the filter, and $N_0$ is the photoreceptor noise.
where $\kappa$ is a constant, $\hat{\Psi}(f)$ is the typical power spectrum of natural images defined as
\begin{equation}
\hat{\Psi}(f) = \frac{A\Omega^2 I^2}{\left(1 + \left(\frac{f}{a}\right)^2\right)^{p/2}}, \tag{10}
\end{equation}
and $M(f)$ is the part of the filter that accounts for the optical modulation transfer function (MTF) and noise reduction (the first stage of retinal computation):
\begin{equation}
M(f) = \frac{1}{N} \frac{1}{\left[\frac{1}{I} \frac{\hat{\Psi}(f)}{\hat{\Psi}(f) + N^2}\right]^{1/2}} e^{-20(f/f_r)^2}, \tag{11}
\end{equation}
The parameters in eqns. (9)–(11) are $N$ (where $N = \sqrt{I}$ is the noise due to quantal photon absorption—see eqn. (7)), $N_0$ (noise at the photoreceptor synapse), $A$ (constant), $\Omega$ (contrast), $a$ (spatial scale), $p$ (log-log slope of the power-spectrum slope), and $f_c$ and $a$ (that account for the eye's optics).

After eliminating the dependent parameters in the theory's eqns. (9)–(11), the filter remains
\begin{equation}
\tilde{h}(f') = \frac{(F_2(f') \Omega^2)^{1/2}}{[I'(\Omega^2 I' + \beta F_2(f'))(\Omega^2 + \beta F_2(f') e^{2(f/f_c')^2})]^{1/2}}, \tag{12}
\end{equation}
where
\begin{align*}
f' &= \frac{f}{a}, \\
P_2(f') &= (1 + f')^{p/2}, \\
\beta &= \frac{\nu N_0^2}{A}, \\
f_c' &= \frac{f_c}{a}, \text{ and } \\
I' &= I_0 N_0^2.
\end{align*}
The parameter $\beta$ synthesizes the photoreceptor noisiness and $f_c'$ is related to the optics of the eye (the larger the $f_c'$, the wider the eye's MTF, and the better the eye's resolution).

**Maximal kurtosis theory**

The goal of this theory is to compute a filter that when convolved with a natural image yields maximal kurtosis. This theory is an adaptation of ideas by Field (1994), who did not develop a mathematical formulation of the theory. (However, in his work, there are some demonstrations that receptive fields with central-excitation-and-peripheral-inhibition profiles produce outputs with high kurtosis.) We now present a mathematical formulation of this theory. The theory begins with the image $I(\vec{r})$ and the filter $h(\vec{r})$ that models lateral inhibition. What is transmitted after filtering is
\begin{equation}
B(\vec{r}) = \int_{\text{ retina}} d\vec{r}' I(\vec{r}') h(\vec{r} - \vec{r}'), \tag{13}
\end{equation}
In this equation for $B(\vec{r})$, we assume that lateral inhibition is linear, which is shared with some other theories. What we need to know is the kurtosis of $B(\vec{r})$ in the retina, that is,
\begin{equation}
K_{\text{ret}}(B(\vec{r})) = \frac{\langle (B(\vec{r}) - \langle B(\vec{r}) \rangle)^4 \rangle_r}{\langle (B(\vec{r}) - \langle B(\vec{r}) \rangle)^2 \rangle_r^2}, \tag{14}
\end{equation}
where $\langle \rangle_r$ indicates average over all the retina. The theory proposes to determine $h(\vec{r})$ such that $K_{\text{ret}}(B(\vec{r}))$ would be maximum (in Balboa, 1997 we demonstrate that this maximum exists). Because eqn. (14) is independent of the amplitude of $h(\vec{r})$, we set $h(0) = 1$. Furthermore, from eqns. (13) and (14) one sees that there are no free parameters in the calculation of $h$.

**Minimal-local asperity hypothesis**

The goal of this hypothesis is to help to encode edges, their contrast, and a compressed representation of intensities away from edges. Bipolar cells would carry the code for these variables as a transformation from the photoreceptor output mediated by lateral inhibition. This inhibition would produce Mach bands (Fiorentini & Ralitici, 1957; Ratliff, 1965), a code for edges. Moreover, if this inhibition is division-like (Merwine et al., 1995; Kamermans et al., 1996), then a compressed code for intensity and an accurate representation of local contrast emerges (Balboa & Grzywacz, 1999b). Mathematically, we modeled such a division-like inhibition as a feedback to photoreceptors, whose outputs are linearly related to their corresponding bipolar cells (Balboa & Grzywacz, 1999b):
\begin{equation}
B_{i,j} = \frac{T_{i,j}}{1 + (h_{i,j} * B_{i,j})^n}, \tag{15}
\end{equation}
where $B_{i,j}$ is the response of the bipolar cell at position $(i,j)$, $T_{i,j}$ is the transduction-dependent input to the photoreceptor’s synapse at position $(i,j)$, and "*" stands for convolution. The convolution means that the lateral inhibition pools the photoreceptors’ output linearly. We use $B_{i,j}$ in the convolution, setting up eqn. (15), since $B_{i,j}$ is proportional to this output. Raising the convolution to $n$ represents the cooperation of $n$ molecules of inhibitory transmitter or nonexclusively, nonlinearities in the calcium or chloride conductances of the inhibitory synapse (Kamermans et al., 1996).

To understand the errors that can occur in the codification of edges, contrast, and intensity when using a mechanism as in eqn. (15), we use a simple example: Consider an image with one black object on a white background illuminating the retina. How is the lateral processing affecting the response at each zone in the image? If one assumes that the lateral-inhibition extent is smaller than the object, then there are three zones in the image worth discussing. These zones are inside the object, outside the object, and over the borders. Inside and outside the object, each bipolar cell codes something like the local compressed intensity. Over the borders, each bipolar cell codes a mixture of the intensities inside and outside the object, which is similar to the local contrast. These codes for contrast and intensity are certainly strongly affected by quantal noise. To compensate for the quantal-noise error being large at low intensities, the system should average the signal over a wide sample of photoreceptors, increasing the $h_{i,j}$’s extent. However, more borders would be straddled in this case, affecting border detection and contrast estimation by mixing information from several objects. This causes
the local-asperity problem, and to reduce it, one has to shrink the $h_{i,j}$'s extent. Hence, the horizontal cells' receptive field can be seen as the minimization of two terms of error

$$E = E_Q + E_A,$$

(16)

where $E_Q$ and $E_A$ are the quantal-noise and local-asperity errors, respectively. Mathematical analyses (Grzywacz & Balboa, 1999) revealed that the quantal-noise error is to a good approximation

$$E_Q = \frac{\sum_i \sum_j \hat{h}_{i,j}^2}{\langle I_{i,j}\rangle^{1/2}},$$

(17)

where $\varphi_Q > 0$ indicates the photoreceptor noisiness, $\hat{h}_{i,j} = h_{i,j}/\Sigma_{i,j} h_{i,j}$ is the filter normalized, and $\langle I_{i,j}\rangle$ is the mean input to the photoreceptor's synapse from the photoreceptor's outer segment averaged over all positions $(i,j)$. And the local-asperity error is to a good approximation (Grzywacz & Balboa, 1999)

$$E_A = \varphi_A \sum_k \sum_I \left( \sum_j (i-k)^2 + (j-I)^2 \right)^m \left| \frac{\nabla I}{I} \right|_{i,j},$$

(18)

where $\varphi_A > 0$ is a constant, $\rho$ is the spatial standard deviation of $\hat{h}_{i,j}$, and $|\nabla I|/I$ is an estimate of the probability that there is a border at position $(i,j)$ given that $|\nabla I|/I$ is the contrast there.

Results

Predictive coding theory

The goal of the predictive coding theory is to encode light intensities such that their wide range fits in the narrow dynamic range of the second-order neuron. It is thus not surprising that this theory accounts for surround inhibition as a way to compress the center's response. Fig. 2A is an example of a filter with inhibitory surround obtained from our implementation of the predictive coding theory. After the inhibitory phase, the filter is excitatory again at large radii (this always occurs at sufficiently high intensities; Fig. 2B). From filters like this, we compute the lateral-inhibition extent.

The dependence of the lateral-inhibition extent on intensity is the same for all the parameters (Fig. 3). At low intensities, there is a plateau in the filter size, which falls at intermediate light intensities toward another plateau at high intensities. The effect of increasing the parameter $p$ (the power spectrum's log–log slope) is to reduce the fall of the lateral-inhibition extent (Fig. 3A). In the limit of $p = 2$, the lateral-inhibition extent is infinitesimally small. The reason is that for $p = 2$, the two-dimensional integral of the power spectrum is improper, leading to an infinite autocorrelation function at zero distances. Such an infinity occurs even if one avoids the zero-frequency singularity, as in eqn. (8) power spectrum, which leads to an autocorrelation proportional to $K_0$. Substituting $\sigma$ for $R_{i,j,i,j}$ in eqn. (2) gives $h_{i,j} = 0$. The effect of increasing the parameter $\Omega$ (statistical contrast) is to start the extent's fall at lower light intensities (Fig. 3B). As a result, the lateral-inhibition extent is smaller for larger statistical contrasts. Finally, the effect of reducing the parameter $\nu$ (photoreceptor noisiness) is to begin the extent's fall at lower intensities and therefore, the lateral-inhibition extent is smaller for small $\nu$ (Fig. 3C).

In summary, different from the apparent behavior of the lateral-inhibition extent in the OPL, the predictive coding theory predicts a fall of this extent with intensity. Appendix B shows that this conclusion is valid for all choices of parameters.

Maximal decorrelation theory

Fig. 4A is an example of a receptive field obtained with our implementation of the maximal decorrelation theory. This example shows the resulting lateral-inhibition filter, which therefore is an optimal mechanism to decorrelate the output spatially. From filters like this, we extract the lateral-inhibition extent. To do so, one has to contend with the filter showing oscillations in its negative zone, which makes the extraction more difficult. It is easier to see these oscillations in Fig. 4B, which is a zoomed portion of Fig. 4A. We estimate the lateral-inhibition's extent despite the oscillations by fitting the inhibition profile with a Bessel function (Fig. 4B).

For the maximal decorrelation theory, the lateral-inhibition extent falls as the mean intensity increases until reaching a plateau at high intensities (Fig. 5). Reducing $p$ causes a smaller lateral-inhibition extent (Fig. 5A). The effect of the parameter $\Omega$ is a reduction of the lateral-inhibition extent as $\Omega$ increases (Fig. 5B). The effect of the parameter $\beta$ is a narrower lateral-inhibition extent as $\beta$ (noisiness) decreases (Fig. 5C). Finally, there is not any effect of $f'_e$ (eye's optical resolution) at extreme intensities, and only at intermediate intensities, an increase in $f'_e$ causes a reduction of the lateral-inhibition extent (Fig. 5D).
Fig. 3. Result of the simulations varying all the parameters in the predictive coding theory. The solid line in each graph is the standard curve ($\Omega = \frac{1}{3}$, $\nu = 0.001$, and $p = 3$). The other lines in each graph correspond to modulations from the standard curve of only the indicated parameter. For all parameters, the lateral-inhibition extent falls with the mean intensity between two plateaus at low and high intensities. A: This fall is larger when $p$ is smaller and the extent becomes infinitesimally small for $p = 2$. B: The fall begins at lower intensities when $\Omega$ is larger. C: The fall begins at lower intensities when $\nu$ is smaller.

In summary, the fall of the lateral-inhibition extent with intensity also makes the maximal decorrelation theory incompatible with the OPL. Appendix C shows that this fall is independent of the choice of parameters at low intensities.

Maximal kurtosis theory

The receptive-field behavior with light intensity is more complex in the maximal kurtosis theory than in the other theories discussed in this paper. In particular, there is a discontinuity at low light intensities for all 89 images simulated. In Fig. 6, we show filters at four intensities from two images, one atmospheric and one underwater (labeled a107 and i241, respectively, in our catalog of images). The filters at the two lowest intensities show the discontinuity in the receptive-field behavior, with jumps from delta-function-like profiles to center-surround organization (Fig. 6, upper four panels). In the majority of the images (56.2%), the discontinuity occurs between mean intensities of 0.25 and 1.56 photons per integration time of the photoreceptors, with the discontinuity point varying from image to image. In the center-surround regime, there are not great differences in the filters' shape or amplitude as the intensity increases, though we observe a narrowing of the lateral-inhibition extent.

Fig. 4. A: An example of a filter obtained with our implementation of the maximal decorrelation theory ($t = 10^{-1}$, $\Omega = \frac{1}{3}$, $\beta = 0.001$, $p = 2$, and $f_c = 75$). An excitatory-center-inhibitory-surround organization is apparent. Observe also the oscillations in the inhibitory zone. B: Detail of the oscillations in the filter's inhibitory zone from A. The dashed line is the fit with the Bessel function used to define the lateral-inhibition extent of the filters.

Fig. 5. Result of the simulations varying all the parameters in the maximal decorrelation theory. The conventions are as in Fig. 3 with the standard curve being ($\Omega = \frac{1}{3}$, $\beta = 0.001$, $p = 2$, and $f_c = 50$). For all parameters, lateral-inhibition extent falls with the mean intensity. A: When the parameter $p$ is larger, the extent is larger. B: The extent is smaller when $\Omega$ is larger. C: The extent is larger when $\beta$ is larger. D: The parameter $f_c$ only affects at intermediate intensities such that there are larger extents when $f_c$ is smaller.
In conclusion, the maximal kurtosis theory does not account for the bell-shape behavior of the lateral-inhibition extent in the OPL as a function of intensity.

**Minimal-local asperity hypothesis**

For the minimal-local asperity hypothesis, there are four parameters, namely, $m$, $\varphi_Q$, $\varphi_A$, and $n$ [see eqns. (15)–(18)]. However, they are not all independent. Because the parameters $\varphi_Q$ and $\varphi_A$ weigh two terms whose sum is to be minimized, one can multiply them by a common factor without changing the filter. Hence, without loss of generality, we define $\varphi = \varphi_Q/\varphi_A$, which gives the relative weights of the two terms of error in eqn. (16). Thus, the only true free parameters are $m$, $\varphi$, and $n$. Fig. 8 shows the mean lateral-inhibition extents predicted by the hypothesis for 89 images. When the probability that a border passes through a point of a given contrast is linear with contrast ($m = 1$), the hypothesis behaves as the other theories. In this case, it predicts that the lateral-inhibition extent falls as a function of intensity. However, when this probability is supralinear ($m \geq 2$), the hypothesis behaves similarly to the experimental data (Fig. 1). The lateral-inhibition extent is small at low intensities, increases with intensity up to a maximum, and falls afterwards as the intensity increases further. It is fortunate that such bell-shape dependence on intensity requires a supralinear dependence of the probability of borders on contrast, because this is the case for natural images (Balboa & Gryzwozec, 1999c). Further comparison of the bell-shape behaviors in Figs. 1 and 8 shows that the HA and HB curves are similar to the $m = 4$ and $m = 2$ curves, respectively. These may complement each other by processing different types of images, at least with respect to the parameter $m$.

We do not illustrate the effects of the other two parameters of the hypothesis here, since we discuss them elsewhere (Balboa & Gryzwozec, 1999b). However, we can mention that increasing the parameter $\varphi$ increases the lateral-inhibition extent. In turn, the parameter $n$ affects the feedback's compression at the bipolar cells' output [eqn. (15)]. Larger $n$ implies that the OPL is coding a wider range of intensities.

In conclusion, different from the other theories analyzed in this paper, the minimal-local asperity hypothesis accounts for the bell-shape behavior of the lateral inhibition mediated by horizontal cells.

**Discussion**

The predictive coding, maximal decorrelation, and maximal kurtosis theories predict a fall of the lateral-inhibition extent as a function of intensity, while the OPL data show a bell-shape behavior. Consequently, these theories do not seem to be applicable to the OPL. The reason for the predicted fall is that at low intensities, noise is dominant and it is sensible to average the signal in a wide neighborhood of photoreceptors to enhance the signal-to-noise ratio. In contrast, the minimal-local asperity hypothesis predicts the intensity bell-shape behavior observed in the OPL data. The novelty in this hypothesis is the minimization of an error due to quantal noise, and another due to multiple straddling of borders by lateral inhibition. The shifting balance between these two errors as intensity changes produces the bell-shape behavior.

**Limitations**

The main limitation of our analysis is that it is based on particular instantiations of the three theories inadequate for the OPL. For
instance, Srinivasan et al. (1982) do not take into account the gap junctions known to exist between photoreceptors (for a review see Dowling, 1987), making their receptive-field center unreasonably punctual. Furthermore, Atick and Redlich (1992) make their receptive fields linear, not taking into account the various biophysical nonlinearities of the system. Could taking into account these various processes save the theories? We cannot rule out this possibility and the low-intensity discontinuity observed in the most nonlinear of the theories analyzed here (Fig. 6) suggests that this possibility is not far-fetched. However, the success of the minimal-local asperity hypothesis highlights that one must consider such nonlinearities and the information-processing needs of the visual system.

**General lessons from theories based on just maximizing luminance information**

Although three theories were not applicable to the OPL, manipulation of their parameters results in effects that could have general validity for vision. For instance, reducing the log-log slope of the power spectrum or the system’s noisiness, or increasing the statistical contrast or the optical resolution, always leads to narrower lateral-inhibition extent (Figs. 3 and 5). Another instructive behavior of lateral inhibition is the discontinuity at low intensities in the maximal kurtosis theory (Fig. 6). This theory also predicts filters with multiple-inhibitory zones (Fig. 7), which are not seen in the retina, but could be related to the receptive-field profiles in the visual cortex (for a review see DeValois & DeValois, 1988). Elsewhere, we discuss in detail how these predictions and other features of the curves come about (Balboa, 1997). For instance, we point out that high contrasts mean good signals, and thus there is little need for spatial averaging, because only extremely low intensities affect images with high contrasts. And if one considers the signal (contrast), one must also consider the noise. Photoreceptor noisiness increases when the period of integration of the photoreceptor decreases or when the amplitude of its response to an individual photon increases (Dodge et al., 1968; Wong et al., 1982; Grzywacz et al., 1988). In the predictive coding theory, the parameter \( \nu \) represents this noisiness and the larger the \( \nu \) is, the wider the lateral-inhibition extent. In the maximal decorrelation theory, the noisiness parameter \( \beta \) is related to \( \nu \) plus the noise at the

---

**Fig. 7.** Four upper panels: Lateral-inhibition extent behavior for four images (two atmospheric and two underwater, which in our catalog were i241, i83, a88, and a107, starting on the top-right panel and going clockwise) at intensities above the discontinuity. All these images show a fall as the mean intensity increases, and the shape of this fall can be steep or shallow depending on the particular image, but not its habitat. Bottom panel: Results for one image (a121 in our catalog) that yielded exceptional behavior. This image produced two kinds of filters: filters with only one inhibitory zone (filled circles) and filters with two inhibitory zones (filled triangles). The image also produced an exceptional peak in the lateral-inhibition extent at one intermediate intensity. This was due to the emergence of the two-inhibitory-zone receptive fields (pointed up by arrow).
The minimal-local asperity hypothesis: More than maximizing luminance information

From the discussion above, one can conclude that the OPL is doing something more than just coding luminance information optimally under biological constraints. All the theories that tried to code only general luminance information were inadequate for the OPL. Only the minimal-local asperity hypothesis, which dealt with specific visual attributes, behaved as the OPL's lateral-inhibition extent data. Moreover, the hypothesis appeared to be applicable to all types of horizontal cells (Fig. 1—Xin & Bloomfield, 1999). This hypothesis was also consistent with other properties of the OPL's lateral inhibition (Balboa & Grzywacz, 1999b). For instance, one of the hypothesis' predictions was that changes in the spatial properties of the stimulus would change horizontal-cell receptive-field sizes even if the mean luminance were constant (Balboa & Grzywacz, 1999b). Such changes were also predicted by Kamermans et al. (1996) with a realistic biophysical model of the OPL.

In conclusion, the function of the OPL's lateral-inhibition adaptation may not just be taking out the irrelevant noise so that as much "pure" signal as possible can go to the rest of the brain. Instead, besides dealing with noise, this adaptation may optimize the extraction of selective kinds of information, which we propose are edge detection, contrast, and intensity. We believe that this approach can be extended to explain lateral processing in other parts of the visual system. And, as for the OPL, the necessary ingredients for such an extension are clear specifications of the desired visual attributes, the relevant statistics of natural scenes, and the biological constraints.

Acknowledgments

We thank Dr. Stewart Bloomfield for sending us his data on horizontal-cell adaptation before publication and Dr. Alexander Ball for loaning us his only copy of Jamieson's Ph.D. thesis. We also thank Dr. Joaquín De Juan and Dr. Russell Hamer for critical comments, Drs. Alan Yuille, Pierre-Yves Burgi, and Davi Geiger for writing letters of support for R. Balboa's work to be defended as a Thesis in Computer Science, and Dr. José Oncina for financial support during R. Balboa's Ph.D. work at the University of Alicante, Alicante, Spain. This work was supported by National Eye Institute Grants EY08921 and EY11170, and the William A. Kettlewell Chair to N.M. Grzywacz, by Grant TIC97-0841 to Dr. José Oncina, and by a National Eye Institute Core Grant to Smith-Kettlewell.

References

Balboa, R.M. & Grzywacz, N.M. (1999a). Occlusions may explain why natural-image power spectra fall with frequency squared (submitted for publication).


Appendix A

This appendix explains the computational methods used for the predictive coding (Srinivasan et al., 1982), maximal decorrelation (Atick & Redlich, 1992), and maximal kurtosis (Field, 1994) theories. The computational methods for the minimal-local asperity hypothesis appear elsewhere (Balboa & Grzywacz, 1999b).

Predictive coding theory

Autocorrelation function

To explore different autocorrelation functions, one must vary the parameter $p$ in eqn. (8). The inverse Fourier transform (autocorrelation function) of eqn. (8) is (Gradshteyn & Ryzhik, 1994)

$$F(r) = \frac{B\pi^{(p-4)/2}}{2\Gamma(p/2)} \frac{r^{(p-2)/2}}{(r^{p/2})_{(p-2)/2}(2\pi r^{p/2})}.$$  

where $G$ is the gamma function and $K_{(p-2)/2}$ is the Bessel function of the second kind of order $(p-2)/2$. We fix $p$ to values close to those obtained for the power spectra from natural images (Balboa & Grzywacz, 1999a):
Role of early retinal lateral inhibition

\[ p = 2: F(r) = A_2 K_0(r/\lambda), \text{ where } A_2 \text{ and } \lambda \text{ are positive constants.} \]

\[ p = 3: F(r) = A_3 e^{-r/\lambda}, \text{ where } A_3 \text{ is a positive constant.} \]

\[ p = 5: F(r) = A_5 (1 + (r/\lambda)) e^{-r/\lambda}, \text{ where } A_5 \text{ is a positive constant.} \]

We want to test the theory for the various autocorrelation functions in the same spatial range. To do so, we force them to intersect at \( r = 5 \) by making \( A_2 = 1/0.421 \cdot e, A_3 = 1, A_5 = 1, \) and \( \lambda = 5. \) From these values, we estimated the largest size of the matrix used to model the filter \( h, \) that is, the matrix yielding largest lateral-inhibition extent that we can have in our results. We chose 29 \( \times \) 29 points, corresponding to when the lowest autocorrelation (that for \( p = 5 \)) falls by 95% of its maximum.

**Implementation**

Only one quadrant of the autocorrelation function was calculated, because this function is approximately isotropic (eqn. (5)). The size of the quadrant of the autocorrelation function was 57 \( \times \) 57 points. With this function, we solved the linear system that defined the lateral inhibition (eqn. (2)). For all solutions of the system, we used \( \lambda = 5. \) [Eqn. (5) demonstrates that variations of \( \lambda \) only change the spatial scale of the filter, not its general shape. Thus, our conclusions are independent of \( \lambda. \)] The solution of the system of equations obtained the filter as a function of one variable, the mean intensity (\( I; \) in a range of \([10^{-2}, 10^2]\) with 0.5 log-unit increments), parametric on the statistical contrast \((\Omega; \Omega = 1/2, \Omega = 1/2, \) and \( \Omega = 1)\), photoreceptor noisiness \((\nu; \nu = 10^{-4}, \nu = 10^{-2}, \) and \( \nu = 10^{-2})\), and log-log slope of the power spectrum \((p; p = 2, p = 3, \) and \( p = 5)\). While the first and third parameters are in the range of values obtained for natural images (van der Schaaf & van Hateren, 1996; Vu et al., 1997; Balboa & Grzywacz, 1999a), the range of the second was chosen empirically.

**Extraction of the lateral-inhibition extent**

The filter's center is set to 1 and the solution of eqn. 2 multiplied by \(-1). Consequently, the first change of sign is from positive to negative (appearance of lateral inhibition) and the second is the opposite (spatial end of lateral inhibition). We define the extent of lateral inhibition (an isotropic interaction) as the radius from the center to the second change of sign in the filter. To avoid discrete extents, we determine the position of this second change of sign by linear interpolation. If there is not a second change of sign, then we set the lateral-inhibition extent to its maximal possible value (15 points). Finally, we smooth the lateral-inhibition extent as a function of mean intensity with a spline.

**Maximal decorrelation theory**

**Implementation**

The size of the filter that models lateral inhibition is 513 \( \times \) 513. The filter is computed as a function of one variable, the mean intensity (\( I; \) in the range \([10^{-2}, 10^2]\) with 0.5 log-unit increments), and four parameters, the statistical contrast \((\Omega; \Omega = 1/2, \Omega = 1/2, \) and \( \Omega = 1)\), the photoreceptor noisiness \((\beta; \beta = 10^{-2}, \beta = 10^{-1}, \) and \( \beta = 10^{-4})\), the log-log slope of the power spectrum \((p; p = 1.5, p = 2, \) and \( p = 2.5)\), and the eye's optics \((f_1; f_1 = 50, f_1 = 75, \) and \( f_1 = 100)\). As with the predictive coding theory, the first and third parameters are in the range of values obtained for natural images (van der Schaaf & van Hateren, 1996; Vu et al., 1997; Balboa & Grzywacz, 1999a). The second parameter was chosen empirically, and for the fourth, we used values close to those used by Atick and Redlich (1992). (We set the parameter \( \alpha = 1.4 \) as did Atick and Redlich (1992) based on a survey of the literature).

**Extraction of the lateral-inhibition extent**

We show that the filter exhibits oscillations due to the Fourier-transform truncation (Fig. 4B). (Truncation causes oscillations, because the real filter is convolved with the inverse Fourier transform of a square wave. This is related with the Gibb's phenomenon—James, 1995). These oscillations presented problems for the method used to measure the extent of lateral inhibition in the predictive coding theory. The reason for this was that similar filters with different oscillations can have different lateral-inhibition extent, since they affect the transition from the negative to the positive regime. The oscillations could be reduced but not eliminated by increasing the filter's size. However, this alternative costs in computation time. We developed an alternate algorithm to extract the lateral-inhibition extent robustly in the presence of oscillations. This algorithm used the one-dimensional central section of the filter. From this vector, we extracted the negative zone of the filter between the minimum value and the first change of sign to a positive value. After that, we fitted the values of the filter \((y_k)\) in this zone to a Bessel function of the second kind of order zero. The reason to use this function was that when applying a current to a horizontal cell, the voltage fall in a two-dimensional grid that models the horizontal-cell gap-junction connectivity is proportional to \(K_0(r/\lambda)\) (Jack et al., 1983). Another reason is that this function gave excellent fits to our calculated filters (Fig. 4). The parameter \( \lambda \) of the Bessel function was used as a representative of the receptive-field extent. To obtain \( \lambda, \) we minimized

\[
\sum_{i=1}^{d} \left( \frac{K_0 \left( \frac{x_i}{\lambda} \right)}{K_0 \left( \frac{y_i}{\lambda} \right)} - \frac{y_i}{y_k} \right)^2,
\]

where \( d \) is the number of points used from the filter and \( k \) is a central index \((=d/2).\) To avoid negative values for \( \lambda \) in the minimization process, we limited this parameter to be larger than 1. Finally, we smoothed \( \lambda \) as a function of mean intensity with a spline.

**Maximal kurtosis theory**

In this theory, different from the two previous theories analyzed here, we work directly with natural images. We use the same sample of images as in Balboa and Grzywacz (1999a,c). To simulate different mean intensity levels, we add noise to the image. This noise-addition process has two steps: First, we multiply the image by a constant to bring the image to the desired mean intensity. We take the resulting gray levels in the image to be the mean numbers of absorbed photons \((A_p)\) during the photoreceptors' integration time. Second, we add noise to each pixel of the image. We add Poisson noise if the pixel's gray level is less than 20 and Gaussian noise if the pixel's gray level is greater than 20. When \( A_p > 20, \) the Poisson distribution of the numbers of absorbed photons can be approximated with a Gaussian distribution with mean and variance equal to \( A_p \) (Taylor & Karlin, 1994). After cursory exploration with a few images, we decided to begin the exploration of all the images with four mean intensities, corresponding to 0.25, 1.56, 9, and \( \infty \) photons per integration time of
the photoreceptor. (The latter means no additional noise included in the image.) A surprising discontinuity was found such that the optimal filter had no lateral interaction below a critical intensity, but displayed lateral inhibition above this intensity (Fig. 6). To determine more precisely the critical intensity, we used a linear grid of intensities between the tested intensities just below and above the threshold (typically 0.25 and 1.56).

After the noise-addition process, the kurtosis-maximization process begins by convolving the noise-corrupted image with the filter. We assume the filter to be isotropic as in the previous two theories and it is defined as follows: The filter’s central value is 1. Outside the center, the filter is arranged in rings such that the inner radius of the 1st ring is equal to the outer radius of the (i − 1)th ring. The ratio between the inner and outer radii of a ring is r0, where r0 is a number just smaller than 2. The inner radius of the first ring is r0 pixels and excludes only the central point. The outer radius of the last ring is larger than the filter’s size and thus we truncate the last ring. All values of the filter inside a ring are equal such that we have a stepwise approximation of the real filter. The program uses the Matlab’s (MathWorks, Natick, MA) function “const” to change the values of the filter’s rings to maximize kurtosis. The “const” function uses Sequential Quadratic Programming (Powell, 1978a,b; Gill et al., 1981) to find iteratively the minimum (in our case, we maximize kurtosis by minimizing kurtosis multiplied by −1). The function’s termination criteria were set to 10−4 for both the change in two consecutive iterations of parameters (rings’ heights) and function (kurtosis). Experimentally, we observe that the last ring of the filter effectively converges to zero with 5 rings, implying sizes of filters of 65 × 65. (We accepted the convergence of the maximization process if the value at the last ring of the filter was less or equal than 0.2, that is, 20% of the highest amplitude. However, the value at the last ring was typically much smaller than 0.01.) For the first few images, several random initial conditions were used for the values of the rings in the maximization process. However, after realizing that the final filters were similar, we used the solutions of one particular image at intensity ∞ ([ring1 = 1, ring2 = −0.3209, ring3 = 0.0051, ring4 = 0.0056, ring5 = −0.0003]) as the initial condition for the other images at the same intensity. After that, the initial condition for any particular intensity was the final solution for the intensity just above. The initial condition ([ring1 = 0, ring2 = 0, ring3 = 0, ring4 = 0, ring5 = 0]) was used at the lowest intensities to test further whether they were above the discontinuity.

**Cropping of images**

The implementation used natural images directly to obtain the filter. With the full images, the simulation time was slow and thus we tested for a few images whether there were differences between using the entire image and a cropped version of it. We performed tests with three sizes of cropping: 256 × 256, 128 × 128, and 64 × 64. For the first size, the simulation time was still long. The third one occasionally yielded results that did not reflect those of the full image. Thus, we used randomly placed cropped images of 128 × 128 pixels to reduce the simulation time.

**Appendix B**

This appendix demonstrates that, at low intensities, the spatial profile of lateral inhibition predicted by the predictive coding theory is proportional to the autocorrelation function. Thus, at low intensities, the lateral-inhibition filter should never become positive, since natural autocorrelation functions are normally nonnegative. This would cause an infinite filter extent in Fig. 3 and, as a consequence, a fall of the filter extent as the light intensity increases.

The only reason that one sees a plateau and not infinty at low light intensity is the finite size of the matrix used in the implementation.

The system of equations used to obtain the filter modeling lateral inhibition is [eqn. (2)]

\[ R_{i,j,k,l} h_{k,l} = R_{0,0,i,j}, \] (B1)

where we are using summation convention of tensorial calculus (when a suffix appears repeated in a term, we understand that there is a sum over all possible values of this suffix). Let us call the tensor \( R_{i,j,k,l} \) like a matrix such that the rows and columns are ordered by the first and second pair of indices, respectively. In other words, the order of the rows (and columns) is

\[
\begin{align*}
&i = -n, j = -n \\
&i = -n + 1, j = -n \\
&\ldots \\
&i = -n + 1, j = -n + 1 \\
&i = -n + 1, j = -n + 1 \\
&\ldots \\
&i = 1, j = -n \\
&i = 1, j = -n + 1 \\
&\ldots \\
&i = 1, j = 1
\end{align*}
\]

As a result, the values \( R_{i,j,k,l} \) are in the diagonal [eqn. (4)]. If we perform a Gaussian elimination of this matrix, then each pivot has the highest power of \( R_{0} \) among the elements of its row. From eqns. (6) and (7), while \( R_{0} \) depends on light intensity, \( R_{i,j,k,l} \) does not. As light intensity falls, the ratio \( R_{0,i,j,k,l} \) increases inversely proportional to intensity. Consequently, at low light levels, the pivots of the system of equations after the Gaussian elimination dominate. One can thus express the system of equations as

\[
R_{0} h_{n,n-n} + e_{n,n-n,n-n+1} h_{n,n-n+1} \\
+ \ldots + e_{n,n-n,n-n+1} h_{n,n-n+1} = R_{0,0,n,n}
\]

\[
R_{0} h_{n,n-n+1} + e_{n,n-n,n-n+1} h_{n,n} \\
= R_{0} R_{0,0,n,n+1}
\]

\[
\ldots \\
R_{0}^{(2n+1)^{2-1}} h_{n,n} = R_{0}^{(2n+1)^{2-2}} R_{0,0,n,n}
\]

**Extraction of the lateral-inhibition extent**

Neither of the previous two methods to extract the lateral-inhibition extent worked well for the filter implementation chosen for this theory. The reason is that here we had only a few points to define the filter. Consequently, we modified the predictive coding method to extract the lateral-inhibition extent from the filter. We first passed the filter through a threshold, setting a ring’s value to zero if the pre-threshold value was less than 0.05 (5% of the highest amplitude). We then fitted a parabola to the minimum value of the filter and the two neighbors of that point. The independent variable for this fit was the logarithm of the geometric-mean radius in each ring (we used logarithm because the filter radius grows geometrically). After obtaining the parameters of the parabola, we calculated its roots and the lateral-inhibition extent was the highest root. As in the other theories, we smoothed the lateral-inhibition extent as a function of mean intensity with a spline.
where the \( e_{i,j,k,l} \) are much smaller than \( R_0 \) raised to its corresponding powers at the pivots. The approximate solution for this system of equations is then

\[
h_{i,j} = \frac{R_{0,0; i,j}}{R_0}.
\]  

(B2)

Thus, at low light levels, the filter values fall proportionally to the autocorrelation function, independently of the theory's parameters.

\[\tilde{H}(f') = \frac{1}{\beta (\Omega^2 + \beta F_2(f') e^{2(f' l'' c')}}]^{1/2} \times \left[ \frac{1}{F_2^{1/2}(f')} - \frac{\Omega^2 I'}{4\beta F_2^{3/2}(f')} + \frac{\Omega^2 I'^2}{8\beta^2 F_2^{5/2}(f')} \right]. \]  

(C2)

The term before the square parenthesis has a low-pass filter profile, that is, it falls with frequency. Therefore, the absolute value of each of the three terms inside the parenthesis also has a low-pass filter profile, whether multiplied by the term outside the parenthesis or not. In the space domain, the two first terms correspond to an excitatory center (plus sign) and an inhibitory periphery (minus sign). The second term falls faster than the first in the frequency domain, and thus, the second term is wider in space than the first one. Furthermore, the amplitude of the second term is lower than the first one at sufficiently low intensities. The third term is wider still and the amplitude smaller, and thus, it corrects the inhibitory periphery, making it narrower. (Terms higher than the third one are proportional to high powers of \( I' \), being thus negligible at low intensities). In the asymptotic behavior near zero intensity, there is only an excitatory zone (the first term) without inhibition. By increasing the intensity, the second term rises proportionally to \( I' \), becoming significant, and causing inhibition to appear. At these low intensities, the third term is negligible because it is proportional to \( I'^2 \). At even higher intensities, the third term of the Taylor's expansion finally becomes significant, reducing the lateral-inhibition extent.

In conclusion, after inhibition appears at low intensities, its extent falls as a function of light intensity independently of the theory's parameters.

Appendix C

This appendix demonstrates that at low intensities, the maximal decorrelation theory predicts a fall of the lateral-inhibition extent, independently of the theory's parameters. This confirms that our computational results on the lateral-inhibition extent are real and not a numeric artifact.

Any multiplicative factor independent of spatial frequency in the filter's Fourier transform does not affect the profile and extent of the filter. Such a factor can only scale the filter's amplitude. Eliminating such factors from eqn. (12), the Fourier transform of the filter becomes

\[\tilde{H}(f') = \frac{F_2(f')^{1/2}}{\left[ (I' \Omega^2 + \beta F_2(f')) (\Omega^2 + \beta F_2(f') e^{2(f' l'' c'')}) \right]^{1/2}}. \]  

(C1)

At sufficiently low intensities, the three first terms of the Taylor's expansion in \( I' \) approximate this equation well and are