

To appear as a chapter in: **Statistical Theories of the Brain**
eds. R Rao, B Olshausen and M Lewicki
MIT Press, 2001.

Natural Image Statistics and Divisive Normalization: Modeling Nonlinearities and Adaptation in Cortical Neurons

Martin J. Wainwright
Stochastic Systems Group
Lab. for Info. & Decision Systems
MIT, Building 35-425
Cambridge, MA 02139

Odelia Schwartz
Center for Neural Science
New York University
New York, NY 10003

Eero P. Simoncelli
Center for Neural Science, and
Courant Inst. of Math. Sciences
New York University
New York, NY 10003

Understanding the functional role of neurons and neural systems is a primary goal of systems neuroscience. A longstanding hypothesis states that sensory systems are matched to the statistical properties of the signals to which they are exposed [e.g., 4, 6]. In particular, Barlow has proposed that the role of early sensory systems is to remove redundancy in the sensory input, resulting in a set of neural responses that are statistically independent. Variants of this hypothesis have been formulated by a number of other authors [e.g., 2, 52] (see [47] for a review). The basic version assumes a fixed environmental model, but Barlow and Foldiak later augmented the theory by suggesting that adaptation in neural systems might be thought of as an adjustment to remove redundancies in the responses to recently presented stimuli [8, 7].

There are two basic methodologies for testing such hypotheses. The most direct approach is to examine the statistical properties of neural responses under natural stimulation conditions [e.g., 25, 41, 17, 5, 40] or the statistical dependency of pairs (or groups) of neural responses. Due to their technical difficulty, such multi-cellular experiments are only recently becoming possible, and the earliest reports appear consistent with the hypothesis [e.g., 54]. An alternative approach is to “derive” a model for early sensory processing [e.g., 36, 43, 20, 2, 37, 9, 49, 53]. In such an approach, one examines the statistical properties of environmental signals and shows that a transformation derived according to some statistical optimization criterion provides a good description of the response properties of a set of sensory neurons. We follow this latter approach in this chapter.

A number of researchers [e.g., 36, 43, 3] have used the covariance properties of natural images to derive linear basis functions that are similar to receptive fields found physiologically in primary visual cortex (i.e., localized in spatial position, orientation and scale). But these early attempts required additional constraints, such as spatial locality and/or symmetry.

Covariance properties are adequate to characterize Gaussian probability models. But when the higher-order statistical properties of natural images are examined, they are found to be strikingly non-Gaussian [19, 61, 18]. More recent work has shown that these non-Gaussian characteristics may be captured using fairly simple parametric models [e.g., 29, 48, 16, 58]. Several authors have used higher-order statistical measures to derive linear basis functions that are similar to cortical receptive fields [e.g., 37, 9, 53].

We have empirically examined the responses of such linear basis functions to natural images, and found that these responses exhibit striking statistical dependencies, even when the basis functions are chosen to optimize independence [e.g., 48, 11]. Such dependencies cannot be removed through further linear processing. Rather, a nonlinear form of cortical processing is required, in which the linear response of each basis function is rectified (and typically squared) and then divided by a weighted sum of the rectified responses of neighboring neurons. Similar “divisive normalization” models have been used by a number of authors to account for nonlinear behaviors in neurons [39, 10, 21, 22, 13]. Our approach shows that natural image statistics, in conjunction with Barlow’s hypothesis, lead to divisive normalization as the appropriate nonlinearity for removing dependency. That is, the type of nonlinearity found in cortical processing is well-matched to the non-Gaussian statistics of natural images.

In earlier work, we have shown that our model, with all parameters determined from the statistics of a set of natural images, can account qualitatively for recent physiological observations of suppression of V1 responses by stimuli presented outside the classical receptive field [51]. Here, we show that the model can account for responses to non-optimal stimuli. In addition, we show that adjusting the model parameters according to the statistics of *recent* visual input can account for physiologically observed adaptation effects [57].

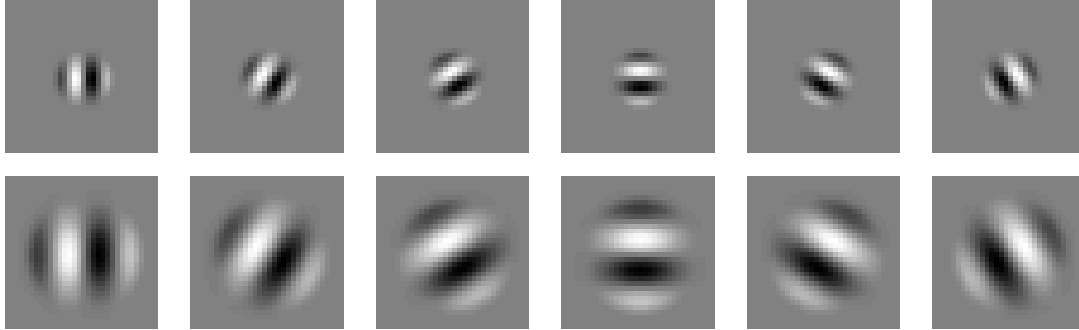


Figure 1: Example basis functions at two different scales, taken from a 6-orientation steerable pyramid.

1 Statistical Properties of Natural Images

As mentioned above, a number of authors have derived linear basis sets by optimizing higher-order statistical measures. For our purposes here, we find it convenient to work with a fixed linear basis that is an approximation to these optimal bases. In particular, we use a “steerable pyramid” [50], whose basis functions are translations, rotations, and dilations of a common filter kernel. This kernel has a spatial frequency bandwidth of roughly one octave, and an orientation tuning bandwidth of roughly 30 degrees. Example basis functions are shown in figure 1.

The marginal statistics of the coefficients obtained by projecting natural images onto such basis functions are known to be highly non-Gaussian and kurtotic [19, 29]. In addition, the joint statistics of these coefficients exhibit striking non-linear dependencies [59, 48]. We have modeled these dependencies, and used the models in a variety of image processing applications, such as compression [11] and denoising [49].

Figure 2 shows the joint statistics of neural responses from a typical pair of adjacent basis functions. Statistics were gathered by collecting response pairs at all spatial positions within a single natural image. A joint histogram is constructed by counting the number of response pairs that fall into each bin in the two-dimensional grid. The resulting array is displayed as a grayscale image, in which the pixel intensity is proportional to the bin counts, except that each column is independently re-scaled, thus forming a *conditional* histogram. The conditional histogram in the upper left panel of Figure 2 shows that this pair of neural responses is well decorrelated, since the expected value of the ordinate is approximately zero, independent of the abscissa. However, the “bowtie” shape of the histogram

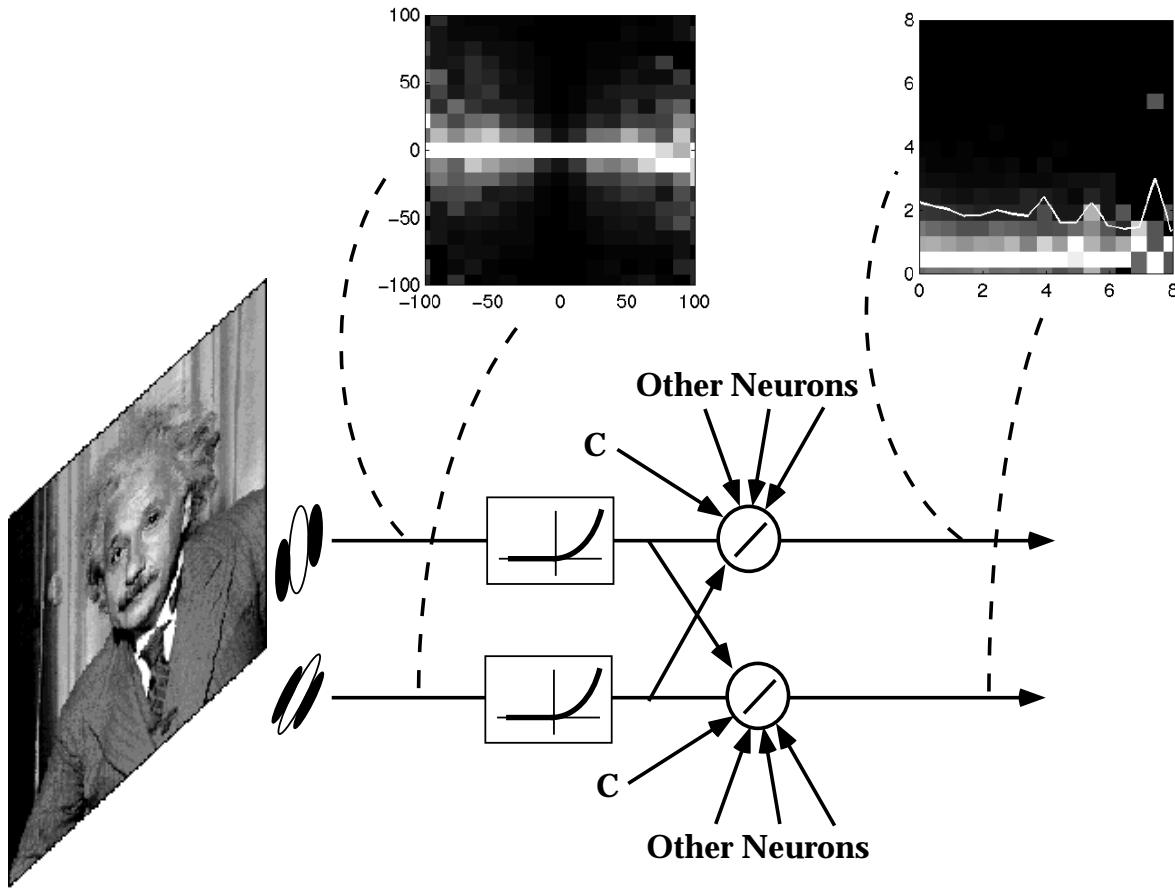


Figure 2. Illustration of image statistics as seen through two neighboring linear receptive fields. Left image: Joint conditional histogram of two coefficients. Pixel intensity corresponds to frequency of occurrence of a given pair of values, except that each column has been independently rescaled to fill the full intensity range. Right image: Joint histogram of divisively normalized coefficients (see text).

reveals that these coefficients are not statistically independent. Rather, the variance of the ordinate scales with the absolute value of the abscissa. We emphasize that this form of dependency cannot be captured using a traditional jointly Gaussian model, and cannot be eliminated with linear processing!

This type of statistical dependency appears in pairs of coefficients at nearby spatial positions, orientations, and scales. Figure 3 shows conditionalized joint histograms for a number of different basis function pairs. The strength of the dependency varies depending on the specific pair chosen. Loosely speaking, we find that the dependency is strongest for basis functions that are close in spatial position, orientation, and scale (spatial frequency), and decreases for pairs that differ markedly in one or more of these attributes.

We have observed this form of dependency in a wide variety of natural images, even when we use a basis set that is optimized to produce maximally independent responses [e.g., 9]. For a given pair

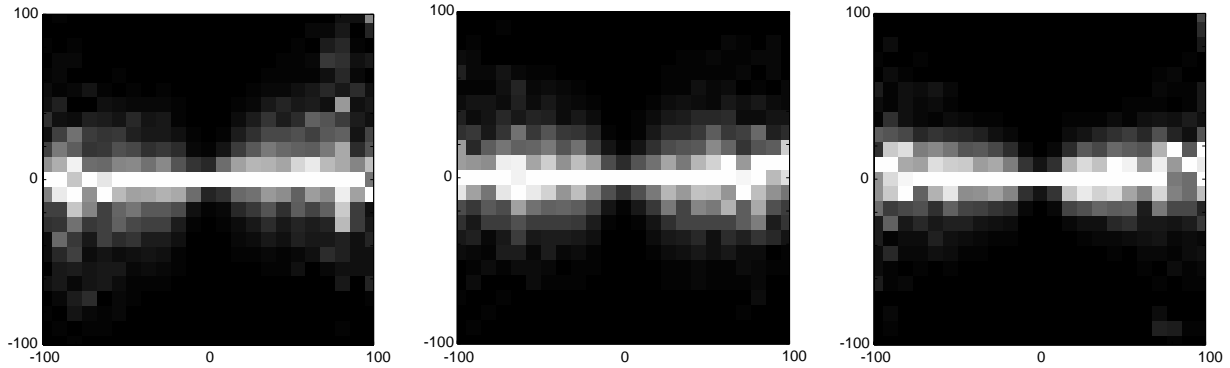


Figure 3. Conditional joint histograms for several different pairs of basis functions. Left: spatially adjacent pair (same orientation and scale); Middle: orthogonal orientations (same position and scale); Right: differing in scale by a factor of two.

of coefficients, the strength of dependency does vary slightly from image to image. Nevertheless, we emphasize that this is a property of the images themselves, and is *not* due purely to our choice of linear basis function. For example, no such dependency is observed when the input image consists of white noise (each pixel independent).

It should not be a surprise that one cannot decompose an image into independent components using a linear basis, since images are not formed from linear superpositions of independent patterns. Even if one assumes that the objects that constitute a scene are drawn independently, the most common combination rule for image formation is *occlusion*, which is nonlinear. The intuitive explanation for the magnitude dependency is that typical localized image structures (e.g., edges, corners) tend to have substantial power across several scales and orientations at the same spatial location. Such a structure will be represented in the wavelet domain via a superposition of the corresponding basis functions. The signs and relative magnitudes of the coefficients associated with these basis functions will depend on the precise location, orientation and scale of the structure. But all of the magnitudes will scale with the contrast of the structure. Thus, measurement of a large coefficient at one scale means that large values in nearby coefficients are more likely.

2 Redundancy Reduction through Divisive Normalization

In previous work [48, 51, 58], we have shown that these statistical dependencies may be reduced using a normalization operation, in which each linear coefficient, L_j , is squared (or halfwave rectified and squared) and then divided by a weighted sum of the squares of its neighbors $\{L_k\}$ and an additive constant:

$$R_j \equiv \frac{L_j^2}{\sigma_j^2 + \sum_k w_{jk} L_k^2} \quad (1)$$

The neighbors, $\{L_k\}$, correspond to linear responses of basis functions at nearby positions, orientations and spatial scales. The right-hand panel in figure 2 shows the conditional histogram of these normalized responses. Note that the probability mass is of approximately constant cross section (apart from statistical sampling errors), and the conditional variance (solid line) is approximately constant. Our current work aims to establish a more formal class of image model for which this divisive operation is optimal [58, 55].

The parameters $\{w_{jk}, \sigma_j^2\}$ used in computing the normalization signal are directly determined by statistical measurements of natural images. Intuitively, larger weights w_{jk} are associated with neighboring neurons whose responses are more predictive of the squared response of the given neuron. The constant σ_j corresponds to the residual variance that cannot be predicted from neighboring coefficients.

In order to compute explicit values for the weights and constant, we minimize the quantity

$$M(w_{jk}, \sigma_j^2) = \mathbb{E} \{ [\log R_j]^2 \}, \quad (2)$$

where \mathbb{E} denotes expected value, which is computed in practice by averaging over all spatial positions of a set of natural images. This procedure corresponds to maximum-likelihood (ML) estimation of the parameters, assuming a lognormal distribution for the conditional density:

$$p(L_j | \{L_k\}; \sigma_j, w_{j,k}) = \frac{1}{|L_j| \sqrt{2\pi\gamma^2}} \exp \left\{ -\frac{1}{2\gamma^2} \left[\log L_j^2 - \log[\sigma_j^2 + \sum_k w_{jk} L_k^2] \right]^2 \right\} \quad (3)$$

The conditional mean and standard deviation of this lognormal distribution are both proportional to $\{\sigma_j^2 + \sum_k w_{jk} L_k^2\}^{\frac{1}{2}}$. This choice of probability density is therefore consistent with the empirical

observation that the standard deviation of L_j roughly scales with the absolute value of the neighbor L_k .

Related choices of error function [48, 51] yield qualitatively similar results. For our the simulations of the following sections, statistical measurements were taken over an image ensemble of four natural images. In all cases, a group of eleven neighbors was used: five orientation neighbors, four spatial neighbors, and two neighbors in scale (coarse and fine).

3 Physiological Modeling: Divisive Suppression

In previous work [51], we have shown that a statistically-derived weighted normalization model can account for various suppressive effects that occur when stimuli are presented outside of the classical receptive field of neurons in primary visual cortex. In particular, a drifting grating presented in an annular region surrounding the receptive field reduces responses to stimuli presented within the receptive field. This reduction appears to be primarily divisive (i.e., a change in gain), and is usually strongest when the parameters (eg., orientation, spatial frequency) of the annular stimulus are matched to those of the preferred excitatory stimulus of the neuron.

Here we demonstrate examples of non-linear behavior inside the classical receptive field. In each example we compare model simulations to electrophysiological recordings a from a typical V1 simple cell in an anesthetized Macaque monkey [13]. Figure 4 shows responses to a sinusoidal grating at the optimal orientation, along with responses to a grating at an orientation 30 degrees away from optimal. Figure 5 shows responses to to a sinusoidal grating at the optimal spatial frequency, along with responses to a grating at frequency 0.78 times the optimal. For the model neuron we have scaled the non-optimal frequency by the ratio of bandwidths between neurons in our model (1 octave) and V1 cells (roughly 1.5 octaves). Note that in both model and cell, the contrast response curves for the non-optimal stimulus are shifted downward from those for the optimal stimulus. Compared with an optimal stimulus, the non-optimal stimulus induces stronger responses in other neurons, increasing the relative strength of the normalization signal. This changes the saturation level of the curve. In addition, the amount of shift in cell and model response curves is qualitatively matched.

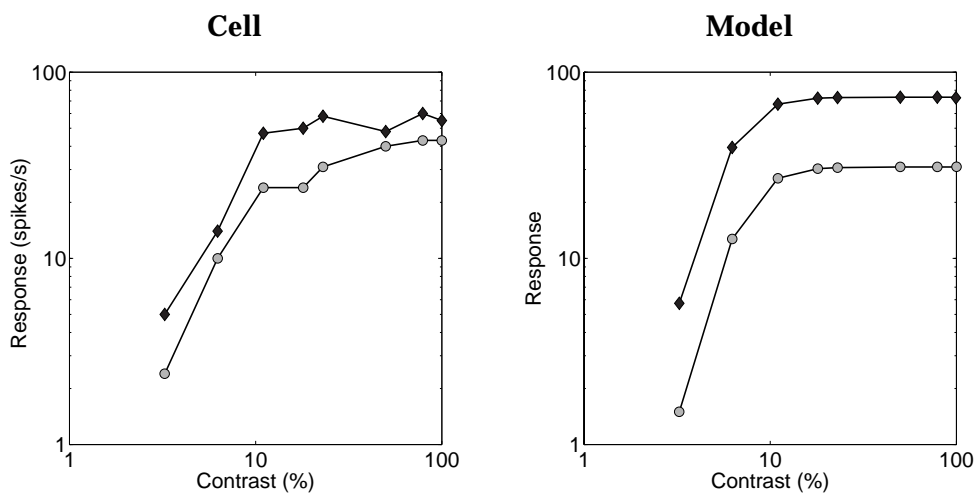


Figure 4. Responses to drifting sinusoidal stimuli at optimal (solid) and non-optimal (hollow) orientations. Left: Cell, taken from [13]. Right: Model.

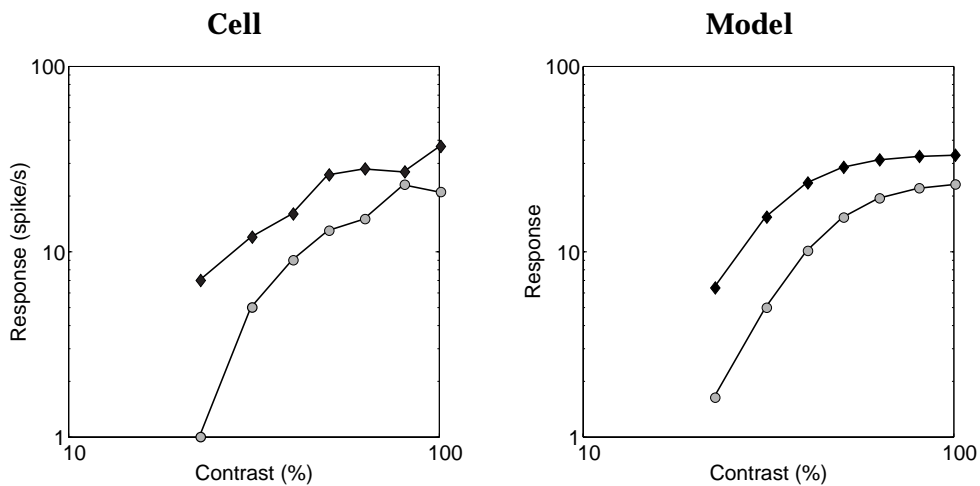


Figure 5. Responses to drifting sinusoidal stimuli at optimal (solid) and non-optimal (hollow) spatial frequencies. Left: Cell, taken from [13]. Right: Model.

4 Physiological Modeling: Adaptation

In the previous sections, we assumed a model with normalization parameters optimized for a generic ensemble of images. Biologically, such an optimal solution would arise through some combination of evolutionary (i.e. genetic) and developmental processes. But given that visual input is constantly changing, it is natural to ask whether the system attempts to adapt itself to the statistics of visual input over shorter time scales. Visual adaptation over relatively short periods is a well-known phenomenon. For example, very rapid adaptation occurs in the retina, where the photoreceptor response continually adjusts to the mean illumination [e.g., 26]. Adaptation effects are also well-documented in the cortex [e.g., 34, 1], but neither their functional purpose nor their mechanism are fully understood. A number of researchers have proposed mechanisms by which cortical adaptation could take place [e.g., 22, 60, 15]. Barlow and Foldiak [8, 7] proposed that linear receptive field properties are adjusted in order to decorrelate responses to recently presented visual stimuli. In related work, we have recently found that optimal changes to linear responses based on signal and noise power spectra predict various psychophysical results following adaptation [56].

In this section, we examine the hypothesis that normalization parameters are adjusted based on the statistics of recent visual input. In contrast to previous approaches to cortical adaptation [e.g., 8, 7, 15, 56], we assume that the linear receptive fields are held fixed, but that the normalization parameters are updated. We begin by showing that two canonical types of cortical adaptation can be distinguished on the basis of image statistics. We then test our adaptation hypothesis by computing a set of “adapted” normalization parameters based on a modified visual environment, and comparing the “adapted” neural responses to physiological recordings.

4.1 Types of adaptation

By considering specific types of changes to the visual environment, we find that image statistics serve to distinguish between two canonical types of adaptation, which we refer to as *contrast* and *pattern* adaptation. Each type of adaptation arises from a specific change in image statistics. In response to the statistical change, normalization parameters are altered in predictable ways, which then leads to a

change in the contrast response function.

For simplicity, consider a single pair of neurons: a linear receptive field and a “neighbor” tuned for a nearby scale. First imagine that we have some “default” visual world, which we call environment A. Typically, environment A corresponds to samples of natural images. The joint conditional statistics of the pair of basis functions in this world are shown in the top left panel of Figure 6, where we see the familiar type of “bowtie” dependency between the two responses. Now suppose that we form a new visual world, which we call environment B, by uniformly re-scaling the contrast of environment A by a factor β . The statistical consequence of this change is a simple re-scaling of the linear responses. This change is illustrated in the bottom left panel of Figure 6, where the axes of the conditional histograms corresponding environment B have been scaled up by a factor of $\beta = 2$. Note that apart from this re-scaling, there is no change in the shape of the conditional histogram. From the cost function in equation (2), it can be seen that rescaling the linear responses by a factor β will change the optimal constant σ but not affect the weights $w_{j,k}$. In particular, the optimal constant for environment B will now be $\sigma_B = \beta \sigma_A$, which compensates for the factor of β in front of all the linear responses L_k . This change in the normalization constant leads to a very specific change in the contrast response function, as shown in the graph on the left side of Figure 6. The constant σ affects the contrast at which saturation sets in, but not the maximal response level. Therefore, the increase in σ produces a rightward shift of the contrast response. In other work, rightward shifts of these curves have been modeled by an increase in the normalization constant [e.g., 22]. Our work establishes that such a change is in fact the optimal adjustment under a uniform re-scaling of contrast. In summary, re-scaling the contrast changes the normalization constant, which shifts the contrast response function to the right; this sequence of events will be called *contrast adaptation*.

A quite different type of change in image statistics involves changing the dependency between a model neuron and its neighbor, as shown in the right side of Figure 6. Such increased dependency could arise in a variety of ways. For illustration here, we constructed environment B by combining environment A with samples from a contrast-varying sinusoidal grating at a spatial frequency lying between the optimal frequencies of the two model neurons. In such an environment, the responses of

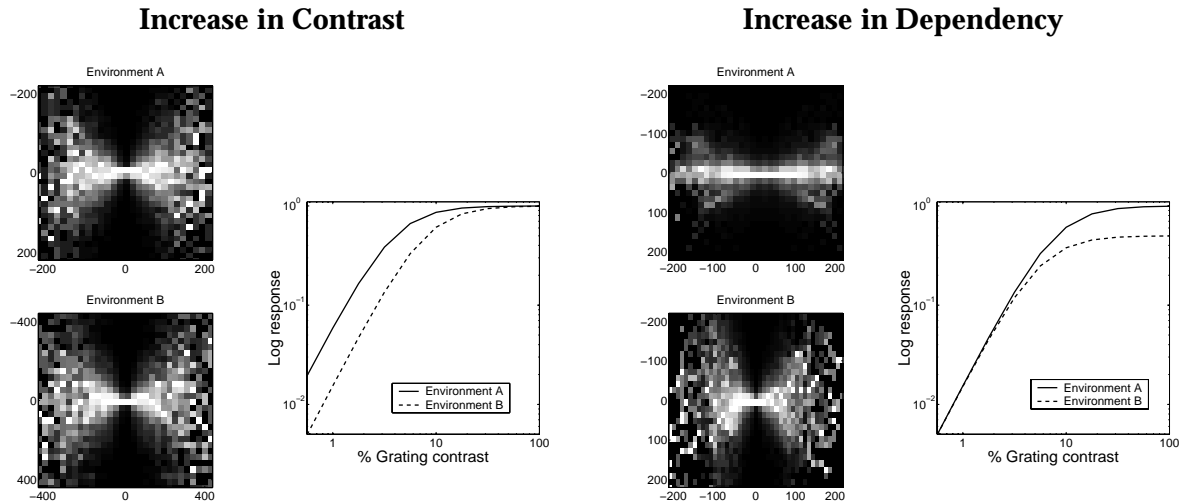


Figure 6. Canonical environmental changes, and the resulting adaptation effect. Left: Increase of contrast. Conditional histograms of linear receptive field responses at adjacent spatial scales (same orientation and spatial position) for two image ensembles (environments). Ensemble B is identical to Ensemble A, but with twice the contrast. The change in environmental contrast produces a change in the normalization constant σ , causing a horizontal shift in the contrast response function. Right: Increase of dependency. Ensemble B contains different spatial patterns than environment A, such that the dependency between the responses of the two linear receptive fields is increased. This is evident in the wider “bowtie” shape of the histograms. Right: the increase in dependency leads to an increase in the normalization weight w for this neighbor, causing a decrease in the saturation level of the contrast response.

the two neurons exhibit increased dependency, as reflected by the joint conditional histograms plotted in Figure 6. The increased dependency is shown by the greater thickness of the “bowtie” in the lower histogram (for environment B), as compared to the upper histogram (environment A). As a result, in order to eliminate the greater dependency between the neuron and its neighbor, more divisive normalization is required. Thus, the model predicts an increase in the normalization weight w , which represents the divisive influence of the neighbor on the neuron. The effect of increasing the weight is to decrease the saturation level of the response curve, as shown in the graph on the right side of Figure 6. In summary, increased dependency between a neuron and its neighbor increases the corresponding normalization weight, which in turn lowers the saturation level of the curve. This sequence of events will be called *pattern adaptation*.

4.2 Simulation

Contrast and pattern adaptation, as defined here, correspond to pure forms of adaptation in response to specific changes in the visual environment. As we will see in our comparisons to physiological data, standard adaptation protocols typically lead to a mixture of contrast and pattern adaptation. To simulate a physiological adaptation experiment, we begin by computing a set of *generic* normalization parameters using equation (2), where the expectation is taken over a fixed ensemble of four natural images. The generic parameters are used to compute responses of neurons under normal (i.e., “unadapted”) conditions, and form a baseline for comparison to the “adapted” responses.

The protocol typically used in neurophysiological experiments consists of presenting an adapting stimulus for an initial adaptation period that ranges from seconds to minutes. Brief probe stimuli are then interleaved with periods of top-up adaptation. We simulated such an adaptation protocol by creating an adapting image ensemble, consisting of the natural image ensemble augmented by the adapting stimulus. The adapting stimulus is chosen in accordance with the experimental protocol: sinusoidal grating images were used for the results reported here. The specific attributes of this grating (e.g., contrast, orientation, spatial frequency) were determined by the experiment under consideration. Note that the adapting ensemble is *not* formed by adding (pixel by pixel) the grating image to each member of the set of natural images. Rather, we collect sample responses from natural images, and mix them with sample responses from the sinusoidal grating image. This collection of responses is then used to compute an *adapted* set of normalization parameters via equation (2). These adapted parameters are used to simulate neural responses following adaptation.

4.3 Effect on contrast response: Different adapting contrasts

It is well-documented that prolonged stimulation of neurons in striate cortex changes their contrast-response function. When the response is measured over a limited range of contrasts, the primary effect of adapting to a high contrast grating can be described as a rightward shift on log-log axes [35]. However, measurements over a wider range of contrasts reveal that adaptation to high contrast gratings not only shifts the function to the right, but also changes its saturation level [1, 44].

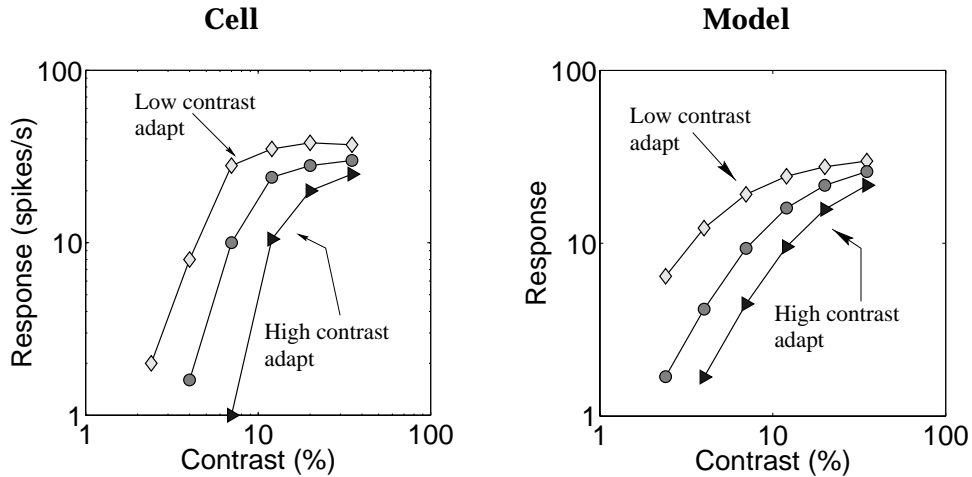


Figure 7. Left: Experimental data from Albrecht et al. [1], showing measured contrast response functions following adaptation to three different contrasts. Adapting stimuli were sinusoidal gratings at the optimal frequency. Right: Corresponding model predictions of changes in the response using an adapting ensemble with optimal frequency sinusoid at three different contrasts.

Shown on the left side of Figure 7 are experimental measurements [1] of contrast response functions following adaptation to optimal frequency sinusoids at three different contrasts. The right side of Figure 7 plots the corresponding model predictions. The drop-off of the model responses are not as steep as those of the cell, because we used a fixed exponent of two, whereas typical fits to cell data yield exponents closer to three. The response curves of both the cell and model data undergo a rightward shift (contrast adaptation), as well as a change in the saturation level (pattern adaptation). In the context of our normalization model, such a mixture of effects is to be expected, because adapting to a high contrast sinusoid not only changes the contrast level experienced by the cell, but also biases the pattern statistics of recent visual input.

4.4 Effect on contrast response: Different test frequencies

An interesting feature of adaptation is that its strength depends on the relationship between the adapting and test stimulus. Albrecht et. al. studied the contingent nature of adaptation by exposing a cell to a grating of optimal spatial frequency, and then testing it with either the optimal or a non-optimal frequency [1]. Plotted in Figure 8 are contrast response curves, tested with either the optimal spatial frequency (left panel), or with a non-optimal spatial frequency (right panel). The corresponding

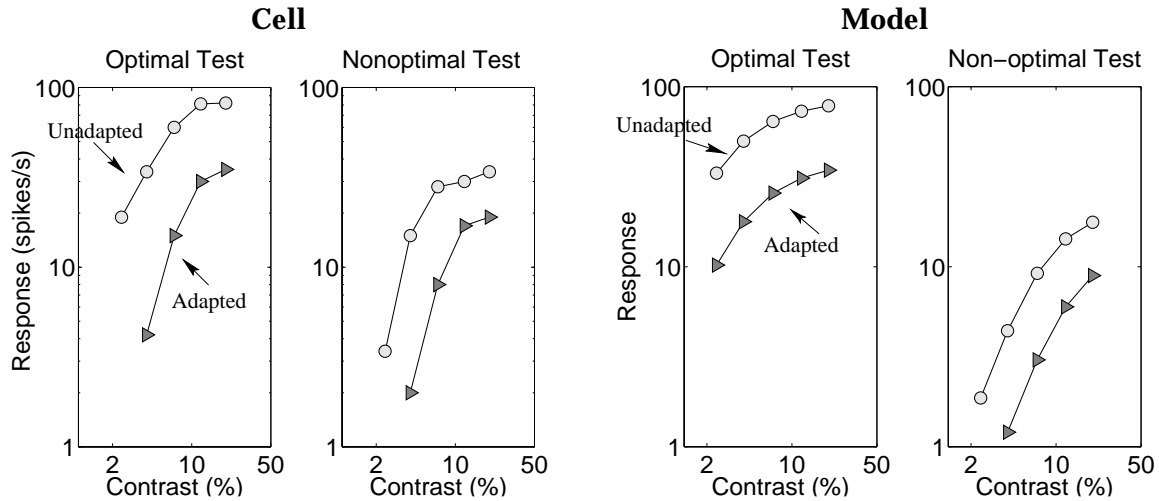


Figure 8. Left: Experimental data from Albrecht et al. [1]. Adaptation shifts the curves right, and changes the saturation level. The effect weakens as the test and adapting frequencies are separated (right panel). Right: Model predictions of changes in the response following adaptation to the optimal orientation and frequency. Each curve corresponds to a different spatial frequency under unadapted (top curves) and adapted (bottom curves) conditions.

model predictions under the same conditions are also shown in Figure 8. In each panel, the upper curve corresponds to the unadapted responses, whereas the lower curve shows the responses following adaptation to a sinusoid at the optimal spatial frequency.

As before, adaptation shifts the curves to the right, and changes the saturation level. However, note that the size of the shift is smaller when the test spatial frequency differs from the adapting frequency. The model also makes analogous predictions about changes in the responses following adaptation to different orientations, for which (to the best of our knowledge) published data is currently unavailable. That is, a close match between the test and adapting stimuli (whether in terms of orientation or frequency) leads to stronger effects of adaptation. This specificity illustrates that adaptation depends on a contingency between test and adapting stimuli.

4.5 Effect on tuning curves

A more dramatic illustration of the contingent nature of adaptation is provided by its effect on tuning curves. A number of researchers have reported that adaptation can alter the shape of tuning curves [e.g., 31, 1, 44, 45]. That is, adapting to a non-optimal stimulus (either in orientation or fre-

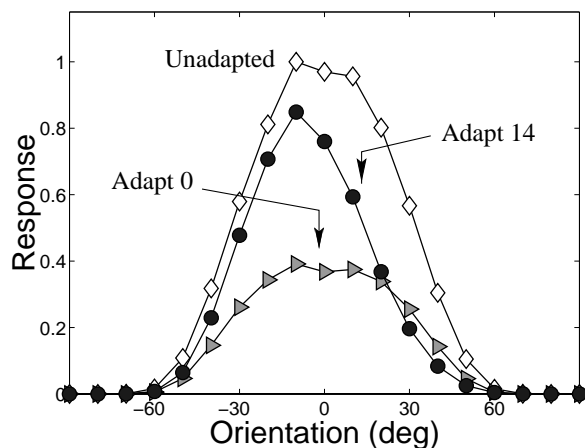


Figure 9. Orientation tuning curves under three conditions: unadapted, adapted to the optimal 0° stimulus, and adapted to the non-optimal $+14^\circ$ stimulus. The effects reflect a mixture of contrast and pattern adaptation, and illustrate the contingent nature of adaptation.

quency) suppresses one flank of the tuning curve, with little or no effect on the other flank. The overall effect is to skew the shape of the tuning curve. Model predictions of orientation tuning curves under such an adapting protocol are shown in Figure 9. The diamonds show the unadapted tuning curve; here the mild asymmetry in the tuning curve is due to mild asymmetries in the natural image statistics of our ensemble. The grey triangles show the tuning curve after adapting to the optimal stimulus at 0° . Here the main effect of adaptation is an overall suppression of the tuning curve, corresponding to contrast adaptation. Superimposed on this main effect is the effect of pattern adaptation, which causes increased suppression at and around the optimal orientation. Adapting to the non-optimal $+14^\circ$ stimulus causes an even more dramatic effect. Here, although the overall suppression of the tuning curve is weak, adaptation causes strong suppression in the right flank of the curve coupled with little effect on the opposite flank. The resultant skewing of the tuning curve is a nice illustration of the contingent nature of adaptation.

These types of differential effects on tuning curves were first reported by Movshon and Lennie [31]; later, Saul and Cynader [44, 45] explicitly plotted frequency tuning curves that were skewed following adaptation. More recently, Müller et al. [32] have reported such effects in some cells, but at time scales significantly shorter than those used in the previous studies. Further experimental investigation is needed to determine if these effects are due to the same adaptation mechanisms. Finally, Carandini et al. [14] have shown that adaptation to a plaid containing both the optimally oriented grating and

a perpendicular grating could increase the amount of cross-orientation suppression in V1 neurons. Such a result is completely consistent with our model, in which the increased statistical dependency in the responses of orthogonally oriented neurons would lead to an increase in the suppressive weights between them.

5 Discussion

Work reported in this chapter was motivated by the early proposal of Barlow [6]: namely, that a desirable goal in early sensory processing is to obtain neural responses that are independent. As emphasized by Barlow, independent neural responses lead to a more efficient representation of sensory input, and facilitate ecologically relevant tasks such as novelty detection. We began by demonstrating that the responses of linear receptive fields, when exposed to natural images, exhibit strikingly non-Gaussian dependencies. Interestingly, the type of nonlinearity suitable for removing these dependencies, per Barlow’s hypothesis, also reproduces the type of nonlinear behavior observed in cortical neurons. On this basis, we developed a divisive normalization model for cortical neurons, whose parameters are entirely determined by the statistical properties of visual images. We find it remarkable that this model can account qualitatively for the suppressive effects of non-optimal stimuli, as well as for basic adaptation effects. Moreover, by considering how normalization parameters should change in response to specific changes in image statistics, we were able to distinguish between two canonical types of adaptation.

Nonetheless, there are many issues that need to be resolved. On the theoretical side, we are working to understand the class of models for which divisive normalization provides an optimal “nonlinear whitening” operation [58, 55]. Moreover, there are also fundamental problems in attempting to state the independence hypothesis more formally. In particular, quantitative testing of the hypothesis requires specification of an image ensemble from which the statistics are taken. We have chosen a small set of generic scenes, but it could be the case that neurons are specialized for particular subclasses of images. Also lacking is a clear statement of which groups of neurons are meant to produce an ensemble of statistically independent responses. Even within cortical area V1, response properties vary consid-

erably from layer to layer, and even within layers. A sensible assumption might be that each stage of processing in the system takes the responses of the previous stage and attempts to eliminate as much statistical redundancy as possible, within the limits of its computational capabilities. This “successive whitening” version of the theory naturally leads one to ask: How far can this bottom-up theory go to explain higher levels of sensory processing? Surely, at some level, the tasks that the organism performs must have some influence on the design of the system.

There are a number of open questions concerning both temporal aspects of the model, as well as the mechanistic implementation of normalization. The current model addresses only steady-state behavior, and the normalization signal is computed and applied instantaneously. Clearly, a more realistic implementation must perform the normalization of a population of neurons using recursive lateral or feedback connections, thereby introducing temporal dynamics. Recent experiments on V1 neurons reveal temporal dynamics that seem consistent with such delayed normalization signals [e.g., 42, 32]. Such recurrent connections might directly provide shunting inhibition [e.g., 23, 12], but the divisive form of the steady-state model does not necessarily require this. We have recently begun to explore models that incorporate delayed feedback mechanisms, but produce steady-state responses consistent with the weighted normalization model described in this paper [30].

Regarding adaptation, our current work does not address the time course or mechanism of such processes. It is possible that several forms of adaptation occur simultaneously on different time scales. Moreover, we have not yet specified a local update (or “learning”) rule to instantiate the optimization specified in equation (2).

Our model is based on a mechanism which is fundamentally suppressive. But a number of authors have reported facilitatory influences from beyond the classical receptive field [e.g., 28, 33, 24]. Facilitation can arise in our model in several ways. First, facilitation of low-contrast optimal stimuli can occur when the secondary (non-optimal) stimulus slightly excites the linear kernel of a model neuron (e.g., when the facilitating stimulus overlaps the classical receptive field). This could explain a number of published results, in which interactions have been shown to be facilitatory at low contrast and suppressive at high contrast [e.g., 27, 38]. Second, a recurrent implementation of our model might

lead to dis-inhibition effects (those neurons that usually suppress responses in the recorded neuron are themselves being suppressed by the facilitating stimulus, leading to an increase in response). As our implementation is currently not recurrent, we have not examined this possibility in detail.

We believe the statistically derived decomposition we have described here may be applicable to other sensory modalities. In particular, we have recently found interesting parallels in the statistics of natural sounds, and the response properties of auditory nerve fibers [46]. Finally, it is our hope that elaboration of the independence hypothesis might benefit future experimental studies. In particular, we are working to design specialized stochastic stimuli for characterizing normalization models, as well as their behavior under various adaptation conditions.

References

- [1] D G Albrecht, S B Farrar, and D B Hamilton. Spatial contrast adaptation characteristics of neurons recorded in the cat's visual cortex. *J. Physiology*, 347:713–739, 1984.
- [2] J J Atick. Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*, 3:213–251, 1992.
- [3] J J Atick and A N Redlich. Convergent algorithm for sensory receptive field development. *Neural Computation*, 5:45–60, 1993.
- [4] F Attneave. Some informational aspects of visual perception. *Psych. Rev.*, 61:183–193, 1954.
- [5] R Baddeley, L F Abbott, M C Booth, F Sengpiel, T Freeman, E A Wakeman, and E T Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Roy. Soc. (Lond.)*, B264:1775–1783, 1998.
- [6] H B Barlow. Possible principles underlying the transformation of sensory messages. In W A Rosenblith, editor, *Sensory Communication*, page 217. MIT Press, Cambridge, MA, 1961.
- [7] Horace B Barlow. A theory about the functional role and synaptic mechanism of visual aftereffects. In C Blakemore, editor, *Vision: Coding and Efficiency*. Cambridge University Press, 1990.
- [8] Horace B Barlow and P Foldiak. Adaptation and decorrelation in the cortex. In R Durbin, C Miall, and G Mitchinson, editors, *The Computing Neuron*, chapter 4, pages 54–72. Addison-Wellesley, New York, 1989.
- [9] A J Bell and T J Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [10] A B Bonds. Role of inhibition in the specification of orientation of cells in the cat striate cortex. *Visual Neuroscience*, 2:41–55, 1989.
- [11] R W Buccigrossi and E P Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans Image Proc*, 8(12):1688–1701, December 1999.
- [12] M Carandini and D J Heeger. Summation and division by neurons in primate visual cortex. *Science*, 264:1333–1336, 1994.
- [13] M Carandini, D J Heeger, and J A Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17:8621–8644, 1997.
- [14] M Carandini, J A Movshon, and D Ferster. Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology*, 37:501–511, 1998.
- [15] F S Chance, S B Nelson, and L F Abbott. Synaptic depression and the temporal response characteristics of V1 cells. *Journal of Neuroscience*, 18:4785–4799, 1998.
- [16] M S Crouse, R D Nowak, and R G Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Proc.*, 46:886–902, April 1998.

- [17] Y Dan, J J Atick, and R C Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *J. Neuroscience*, 16:3351–3362, 1996.
- [18] John G. Daugman. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Trans. Biomedical Engineering*, 36(1):107–114, 1989.
- [19] D J Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.
- [20] P Foldiak. Forming sparse representations by local anti-hebbian learning. *Biol. Cybernetics*, 64:165–170, 1990.
- [21] W S Geisler and D G Albrecht. Cortical neurons: Isolation of contrast gain control. *Vision Research*, 8:1409–1410, 1992.
- [22] D J Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.
- [23] D J Heeger. Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70(5):1885–1898, 1993.
- [24] M K Kapadia, M Ito, C D Gilbert, and G Westheimer. Improvement in visual sensitivity. *Neuron*, 15(4), Oct 1995.
- [25] S B Laughlin. A simple coding procedure enhances a neuron’s information capacity. *Z. Naturforsch.*, 36c:910–912, 1981.
- [26] S.B. Laughlin. The role of sensory adaptation in the retina. *Journal of Experimental Biology*, 146:39–62, 1989.
- [27] J B Levitt and J S Lund. Contrast dependence of contextual effects in primate visual cortex. *Nature*, 387:73–76, 1997.
- [28] L Maffei and A Fiorentini. The unresponsive regions of visual cortical receptive fields. *Vision Research*, 16:1131–1139, 1976.
- [29] S G Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pat. Anal. Mach. Intell.*, 11:674–693, July 1989.
- [30] S Mikaelian and E Simoncelli. Modeling temporal response characteristics of V1 neurons with a dynamic normalization model. In *Proc. Computational Neuroscience*, Brugge, Belgium, 2000. To appear.
- [31] J A Movshon and P Lennie. Pattern-selective adaptation in visual cortical neurones. *Nature*, 278:850–852, 1979.
- [32] J R Müller, A B Metha, J Krauskopf, and P Lennie. Rapid adaptation in visual cortex to the structure of images. *Science*, 285:1405–1408, Aug 1999.
- [33] J I Nelson and B J Frost. Intracortical facilitation among co-oriented, coaxially aligned simple cells in cat striate cortex. *Exp. Brain Res.*, 61:54–61, 1985.

- [34] I Ohzawa, G Sclar, and R D Freeman. Contrast gain control in the cat visual cortex. *Nature*, 298:266–268, 1982.
- [35] I Ohzawa, G Sclar, and R D Freeman. Contrast gain control in the cat’s visual system. *J. Neurophysiology*, 54:651–667, 1985.
- [36] E Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982.
- [37] B A Olshausen and D J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [38] U Polat, K Mizobew, M W Pettet, T Kasamatsu, and A J Norcia. Collinear stimuli regulate visual responses depending on cell’s contrast threshold. *Nature*, 391:580–584, February 1998.
- [39] W Reichardt and T Poggio. Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.*, 35:81–100, 1979.
- [40] P Reinagel and R C Reid. Temporal coding of visual information in the thalamus. *J Neuroscience*, 2000. In Press.
- [41] F Rieke, D A Bodnar, and W Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B*, 262:259–265, 1995.
- [42] D L Ringach, M J Hawken, and R Shapley. The dynamics of orientation tuning in macaque primary visual cortex. *Nature*, 387:281–284, May 1997.
- [43] T D Sanger. Optimal unsupervised learning in a single-layer network. *Neural Networks*, 2:459–473, 1989.
- [44] A B Saul and M S Cynader. Adaptation in single units in the visual cortex: The tuning of aftereffects in the spatial domain. *Visual Neuroscience*, 2:593–607, 1989.
- [45] A B Saul and M S Cynader. Adaptation in single units in the visual cortex: The tuning of aftereffects in the temporal domain. *Visual Neuroscience*, 2:609–620, 1989.
- [46] O Schwartz and E Simoncelli. Natural sound statistics and divisive normalization in the auditory system. In *Adv. Neural Information Processing Systems*, volume 13, Cambridge, MA, 2001. MIT Press. Presented at Neural Information Processing Systems, Dec 2000.
- [47] E Simoncelli and B Olshausen. Statistical properties of natural images. *Annual Review of Neuroscience*, 24, May 2001. To Appear.
- [48] E P Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf on Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, November 1997. IEEE Computer Society. Available from <http://www.cns.nyu.edu/~eero/publications.html>.
- [49] E P Simoncelli. Bayesian denoising of visual images in the wavelet domain. In P Müller and B Vidakovic, editors, *Bayesian Inference in Wavelet Based Models*, chapter 18, pages 291–308. Springer-Verlag, New York, Spring 1999. Lecture Notes in Statistics, vol. 141.

- [50] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. *IEEE Trans Information Theory*, 38(2):587–607, March 1992. Special Issue on Wavelets.
- [51] E P Simoncelli and O Schwartz. Image statistics and cortical normalization models. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Adv. Neural Information Processing Systems*, volume 11, pages 153–159, Cambridge, MA, 1999. MIT Press. Presented at Neural Information Processing Systems, 1-3 Dec 1998.
- [52] J H van Hateren. A theory of maximizing sensory information. *Biol. Cybern.*, 68:23–29, 1992.
- [53] J H van Hateren and A van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B*, pages 359–366, 1998.
- [54] W E Vinje and J L Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287, Feb 2000.
- [55] M Wainwright, E Simoncelli, and A Willsky. Random cascades on wavelet trees and their use in modeling and analyzing natural imagery. *Applied and Computational Harmonic Analysis*, 2001. Special issue on wavelet applications. To appear.
- [56] M J Wainwright. Visual adaptation as optimal information transmission. *Vision Research*, 39:3960–3974, 1999.
- [57] M J Wainwright and E P Simoncelli. Explaining adaptation in V1 neurons with a statistically optimized normalization model. In *Investigative Ophthalmology and Visual Science Supplement (ARVO)*, volume 40, pages S–573, May 1999.
- [58] M J Wainwright and E P Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Adv. Neural Information Processing Systems*, volume 12, pages 855–861, Cambridge, MA, May 2000. MIT Press. Presented at Neural Information Processing Systems, Dec 1999.
- [59] B Wegmann and C Zetsche. Statistical dependence between orientation filter outputs used in a human vision based image code. In *Proc SPIE Visual Comm. and Image Processing*, volume 1360, pages 909–922, Lausanne, Switzerland, 1990.
- [60] H R Wilson and R Humanski. Spatial frequency adaptation and contrast gain control. *Vis. Res.*, 33(8):1133–1149, 1993.
- [61] C Zetsche and W Schönecker. Orientation selective filters lead to entropy reduction in the processing of natural images. *Perception*, 16:229, 1987.