Testing quantitative models of
binocular disparity selectivity in primary visual cortex

Jenny C. A. Read & Bruce G. Cumming

Authors’ affiliation:
Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health
Building 49 / Room 2A50, 49 Convent Drive, Bethesda, MD 20892-4435

Running head: Modeling disparity selectivity in primary visual cortex

Address for correspondence
Jenny Read,
Building 49 / Room 2A50,
49 Convent Drive,
Bethesda,
MD 20892-4435
Email: jcr@lsr.nei.nih.gov
Tel: (301) 451-4926
Fax: (301) 402-0511
Abstract

Disparity-selective neurons in striate cortex (V1) probably implement the initial processing which supports binocular vision. Recently, much progress has been made in understanding the computations which these neurons perform on retinal inputs. The binocular energy model has been highly successful in providing a simple theory of these computations. A key feature of the energy model is that it is linear until after inputs from the two eyes are combined. Recently, however, a modified version of the energy model, incorporating threshold nonlinearities prior to binocular combination, has been proposed in order to account for the weaker disparity tuning observed with anticorrelated stimuli (Cumming and Parker 1997; Read et al. 2002). In this paper, we present new data needed for a critical assessment of these two models. We compare two key predictions of the models with responses of disparity-selective neurons recorded from V1 of awake fixating monkeys. We find that the original energy model, and a family of generalizations retaining linear binocular combination, are quantitatively inconsistent with the response of V1 neurons. In contrast, the modified version incorporating threshold nonlinearities can explain both sets of observations. We conclude that the energy model can be reconciled with experimental observations by adding a threshold prior to binocular combination. This gives us the clearest picture yet of the computation being carried out by disparity-selective V1 neurons.
**Introduction**

The separation of the two eyes introduces disparities between the images received by the left and right eyes. The visual system is somehow able to fuse the images so as to produce a unified percept of the visual world, while using the stereoscopic disparities in order to extract information about how far away viewed objects are. The neural circuits specific to this ability begin in primary visual cortex (V1), the first place in the visual system where inputs from the two eyes converge on individual cells. Many V1 cells modulate their firing rate according to the stereoscopic disparity of the stimulus (Barlow et al. 1967; Nikara et al. 1968). These disparity-tuned cells are believed to perform the initial processing of retinal inputs which eventually, in higher visual areas, gives rise to stereoscopic depth perception and to binocular fusion (single vision). Thus, a detailed understanding of the computations carried out by these cells represents the first step towards a complete description of stereoscopic vision.

The current best description of the operation of these cells is provided by the energy model (Adelson and Bergen 1985; Ohzawa et al. 1990; Qian 1994; Fleet et al. 1996; Ohzawa 1998), sketched in Figure 1A and described more fully below. This elegant model has been extremely successful in explaining qualitatively the properties of disparity-tuned neurons in V1, for example the shape of the binocular receptive field obtained with disparate bar stimuli and the shape of the disparity tuning curve obtained with random-dot stereograms (Cumming and Parker 1997; Ohzawa et al. 1997; Ohzawa et al. 1996; Anzai et al. 1999a; Prince et al. 2002a).

Although the energy model was originally intended as a qualitative description, its success to date suggests that it may be possible to elaborate the model so as to provide a good quantitative description of neuronal behavior. Extending the energy model requires identifying those quantitative discrepancies which have not so far been reconciled with the original structure of the model. First, the response to anti-correlated random-dot stimuli (when contrast polarity is inverted in one eye) must be accounted for. In real cells, anti-correlation inverts the disparity tuning curve and also reduces the amplitude, whereas the energy model predicts inversion only (Cumming and Parker 1997; Ohzawa et al. 1997). The amplitude reduction can be explained if we modify the energy model by incorporating threshold
nonlinearities prior to binocular combination (Read et al. 2002). This modified version of the energy model is shown in Figure 1B.

Second, the energy model predicts that monocular stimulation in either eye should always have an excitatory effect. In this paper, by making a quantitative comparison of monocular and binocular responses, we confirm that there are many cells in which input from one eye always seems to suppress the cell’s response, as has previously been reported by others (Ohzawa and Freeman 1986a; Prince et al. 2002b). Such behavior is inconsistent with linear binocular combination, but is predicted by our modified version (Read et al. 2002).

Third, the energy model predicts that the response to binocularly uncorrelated random-dot patterns should equal the sum of the responses to monocular random-dot patterns. In fact, it is generally much closer to their mean. It has been suggested (Prince et al. 2002a) that this is due to a contrast-normalizing mechanism which tends to boost the response to monocular. We show here that the relative size of the monocular and binocular response can be explained by our modification to the energy model, without the need to invoke a normalization process.

Fourth, a key prediction of the energy model is that the shape of monocular receptive fields determine the shape of the disparity-tuned response. Although this has been verified for simple cells in the cat (Anzai et al. 1999b), the situation for complex cells is less clear because it is difficult to estimate the receptive fields of the subunits. Fortunately, the model can also be tested in the frequency domain: the Fourier power spectrum of the disparity tuning curve should match the shape of the monocular spatial frequency tuning curves. Preliminary testing of this prediction has indicated a conflict with the energy model prediction (Ohzawa et al. 1997; Prince et al. 2002a); however, for a number of reasons, the seriousness of this conflict is hard to assess. Prince et al. (2002a) found that their disparity tuning curves often had spatial frequency bandwidths substantially larger than those estimated from luminance gratings in other studies (de Valois et al. 1982). However, Prince et al. measured tuning for horizontal disparity, so these data are not directly comparable with selectivity for the spatial frequency of luminance gratings at the preferred orientation. Ohzawa et al. (1997) found that the frequency of the disparity tuning curve tended to be lower than the preferred spatial frequency revealed
with monocular luminance gratings in the dominant eye, apparently contradicting the energy model. However, their definition of disparity frequency could potentially obscure an underlying agreement with the energy model (see below); also, confidence intervals were not presented. Most importantly, neither of these studies reported measures of spatial frequency tuning in both eyes. The original energy model assumes that spatial frequency tuning is identical in the two eyes, so it is possible that the discrepancies could be due to binocular differences in spatial frequency tuning. If this were the case, it would be easy to extend the energy model to take account of this difference, by allowing different spatial frequency tuning between subunits, either within an eye or between eyes. The data should then agree with this generalized version of the energy model. Thus, a more complete comparison of the spatial frequency tuning and the power spectrum of the disparity tuning is necessary in order to test the model.

In order to resolve this important question, we recorded the monocular spatial frequency and orientation tuning in both eyes. This is compared with the selectivity for disparity applied to random-dot patterns along an axis orthogonal to the preferred orientation. Both comparisons systematically violate the predictions of the energy model, even after it has been generalized to allow for differences between subunits. The disparity tuning curves show more power at lower frequencies than is possible within these models, even allowing for the presence of several subunits which may differ in position and/or spatial frequency tuning. However, once again, the results may be explained by our modified version of the energy model incorporating a threshold non-linearity prior to binocular combination.

In summary, therefore, we have compared two families of models of disparity selectivity: (a) the energy model and a set of generalizations of it, all postulating linear binocular summation, and (b) our modified version incorporating threshold nonlinearities prior to binocular combination. For a wide range of observations, the data are quantitatively at odds with the linear model, and can be accounted for by the threshold model. We conclude that adding thresholds to the energy model, before inputs from the two eyes are combined, represents a substantial step forward in our understanding of disparity selectivity in V1.
Material and Methods

Detailed descriptions of the general procedures have appeared elsewhere (Prince et al. 2002a; Cumming and Parker 1999). Briefly, single-unit activity was recorded from primary visual cortex (V1) of two awake macaques trained to maintain fixation while viewing stimuli for fluid reward. All protocols were approved by the Institute Animal Care and Use Committee and complied with Public Health Service policy on the humane care and use of laboratory animals.

Stimuli were generated on a Silicon Graphics Octane workstation and presented on two Eizo Flexscan F980 monitors (mean luminance 41.1 cd/m², contrast 99%, framerate 72 Hz) viewed via a Wheatstone stereoscope, in which the monitors are viewed through mirrors positioned in front of the animal’s eyes. At the viewing distance used (89 cm) each pixel in the 1280×1024 display subtended 1.1 min arc. Anti-aliasing was used to render with sub-pixel accuracy (pixels are colored intermediate shades of gray to represent edges that only partially cover the pixel, (Foley et al. 1990)). Glass-coated platinum-iridium electrodes (FHC, Inc.) were placed transdurally each day. Electrode position was controlled with a custom-made microdrive which used an ultra-light stepper motor mounted directly onto the recording chamber.

The monkeys initiated a stimulus presentation by maintaining fixation on a binocularly-presented spot to within ±1°. They were required to maintain fixation at this accuracy for 2.1 s to earn a fluid reward. During each such trial, four stimuli were presented, each lasting 420 ms, separated by 100 ms.

Stimuli

Sinusoidal luminance gratings were used to determine the minimum response field, spatial frequency and orientation tuning of the cell. After an initial determination of the preferred spatial frequency and orientation, the monocular orientation tuning curve in each eye was obtained using a circular patch of grating with spatial frequency reasonably close to optimal. Quantitative orientation tuning curves usually spanned a range of 180° centered around the preferred orientation (or direction, for direction-selective cells). The spatial frequency tuning
curve was then obtained using a large rectangular grating patch at the preferred orientation. The frequencies generally spanned 0.0625 to 16 cycles per degree (cpd) in steps of one octave. A pseudo-random sequence interleaving frequencies and eye of presentation was used in both cases. During monocular trials, the non-stimulated eye viewed a uniform screen of the same mean luminance.

Dynamic random-dot stereograms were composed of black and white dots, scattered at random on a gray background. The dots were usually 5×5 pixels (0.1°×0.1°); for some cells, a different size was used if this enhanced the response rate. A new random stereogram was generated every frame (72Hz). The dot density was sufficient to cover 50% of the gray background but, because the dots were allowed to overlap one another (dot location was randomly assigned with sub-pixel precision using anti-aliasing), the total coverage was slightly less. On average, 20% of pixels were black, 20% white and 60% gray. Figure 2 shows an example stereogram, together with a circle indicating the size of typical V1 minimum response fields for comparison.

The energy model assumes that all receptive fields feeding into a cell have the same orientation. Its predictions are therefore most easily framed in terms of disparities parallel and orthogonal to this orientation, rather than horizontal and vertical disparities. Accordingly, to facilitate testing of energy model predictions, experimental disparities were applied along the axis orthogonal to each neuron’s preferred orientation. These covered the range from −1.2° to +1.2° in the initial test for disparity selectivity, with the range −0.6° to +0.6° covered in steps of 0.1°, and steps of 0.2° outside this range. A larger range of disparities was used if necessary to ensure that there was no modulation at the extremes of the tuning curve (i.e. that the full response range had been explored). In neurons with preferred spatial frequencies > 4cpd, the central region of the curve was sampled more finely, to ensure that sampling exceeded the Nyquist limit predicted from the monocular SF tuning.
**Data analysis**

Data analysis such as curve fitting is greatly simplified if we can make the assumption that variance is constant across the data-set. This assumption is invalid for neuronal firing rates, whose variance tends to increase in proportion with the mean (Dean 1981). However, the square root of firing rates has variance which is roughly constant, independent of the mean (Prince et al. 2002a; Cumming and Parker 2000). This variance-stabilizing transformation greatly simplifies the analysis of neuronal data. For this reason, we performed all our analysis on the square-root of the recorded firing rates.

To quantify the strength of disparity tuning, we used the disparity discrimination index introduced by Prince et al. (2002a):

\[
\text{DDI} = \frac{R_{\text{max}} - R_{\text{min}}}{R_{\text{max}} - R_{\text{min}} + 2 \cdot \text{RMS}_{\text{error}}},
\]

*Equation 1*

where \(R_{\text{max}}, R_{\text{min}}\) are the maximum and minimum firing rate, and \(\text{RMS}_{\text{error}}\) is the square root of the residual variance around the mean firing rate recorded across the whole tuning curve, including the response to uncorrelated stimuli (effectively, infinite disparity). Like the more familiar binocular interaction index, \((R_{\text{max}} - R_{\text{min}})/(R_{\text{max}} + R_{\text{min}})\), this is a contrast measure, except that here the difference in response between the preferred and null disparity is contrasted not with the mean response, but with the variability of the firing rate \(\text{RMS}_{\text{error}}\). This means that cells in which the range in firing rates is largely due to random fluctuations are not wrongly classified as being highly sensitive to disparity; equally cells in which the change in firing as a function of disparity is relatively small but highly reliable are correctly described as strongly disparity-tuned. The term \((R_{\text{max}} - R_{\text{min}})\) in the denominator of Equation 1 ensures that the index is bounded at 1 when the variability is small.

For most cells, monocular random-dot stimuli were also presented, in trials interleaved with binocular stimuli. Blank stimuli were also usually interleaved, where both eyes viewed a blank screen of the same mean luminance as the random-dot patterns. These were used to obtain an estimate of the spontaneous firing rate.
To allow a cell into the study, we required that binocular random dots at the optimal disparity elicit a response of at least 10 spikes/s. In order to proceed to quantitative analysis of the response’s shape, we further required (i) ANOVA indicates a significant (p<0.05) main effect of disparity; (ii) the disparity discrimination index exceeds 0.375. The second condition removes neurons with weak but significant disparity tuning, as these tend to produce noisy estimates in the quantitative analysis that follows. Including these weakly tuned neurons did not change any of the substantial results; it only increased the scatter. In order to subject monocular spatial frequency data to quantitative analysis, we required that (i) the optimal drifting grating in that eye elicits at least 10 spikes/s; (ii) ANOVA indicates a significant (p<0.05) main effect of spatial frequency. We do not require tuning to be bandpass, and our sample included a few neurons which showed a lowpass spatial frequency tuning.

**Fitting tuning curves**

We summarized our tuning curves by fitting them with analytical functions. If we fitted the function directly to the mean firing rates, we would have to reduce the weight given to residuals at higher firing rates, to take account of the higher variance there. As explained above, we avoided this complication by, instead, fitting the square-root of our chosen fit function to the mean of the square-root firing rates. Since \( \sqrt{\text{firing rate}} \) has approximately constant variance, we could then just minimize the sum of the squared residuals, without needing to weight them differently.

Spatial frequency tuning curves were fitted with Gaussians, in either log or linear frequency space, whichever minimized the residuals. These had four free parameters: frequency of the peak \( f_0 \), standard deviation \( \sigma \), baseline and amplitude above the baseline. The baseline was assumed to represent the spontaneous firing rate; thus, it was not allowed to be negative. The peak frequency was constrained to lie within the range of stimulus frequencies. The amplitude was not allowed to exceed twice the range of the response.

These fitted curves were used to extract a peak frequency, a low frequency cut-off, and a high frequency cut-off, defined as the positions where the tuning curve falls to half its maximum.
Where the spatial frequency tuning curve was fitted with a Gaussian in linear frequency, with peak at \( f_0 \) and standard deviation \( \sigma \), the high and low cuts are \( f_0 \pm \sigma \sqrt{\ln 4} \). Where the tuning curve was fitted with a Gaussian in log frequency, with standard deviation \( \sigma \) in log space, the high and low cuts are \( f_0 \exp(\pm \sigma \sqrt{\ln 4}) \).

Disparity tuning curves were fitted with half-wave rectified one-dimensional Gabor functions (the product of a sinusoid with an exponential, cf. Appendix B). The original energy model predicts a Gabor disparity tuning curve, provided that the monocular receptive fields are narrow-band Gabor functions differing only in their position and phase. However, our main motivation for using Gabor functions is that they provide a succinct description of most experimental tuning curves (Cumming and Parker 1997; Ohzawa et al. 1990; Prince et al. 2002b). In the Results section, we verify that our conclusions do not depend on the use of a fitted Gabor. 1d Gabors have six free parameters: the spatial frequency \( f \) and phase \( \phi \) of the carrier cosine, the standard deviation \( \sigma \), amplitude \( A \) and centre \( \delta_0 \) of the Gaussian envelope, and the baseline firing rate \( B \). Uncorrelated responses, if available, were included in the fitting; the expected response to uncorrelated stimuli is just \( B \). \( \delta_0 \) was constrained to lie within the range of stimulus disparities, and the amplitude \( A \) was not allowed to exceed twice the difference between the maximum and the minimum response. The spatial frequency of the fit was not allowed to exceed half the Nyquist limit (i.e., one-quarter of the maximum spatial sampling rate of the data). Although these curves generally gave good descriptions of the tuning curves, the parameters of the fitted Gabor must be interpreted with care (see Prince et al. (2002a)). When using these fits to summarize some property of the tuning curve, we therefore used appropriate measures applied to the fitted curve (illustrated by our measurement of disparity peak frequency in the next section), rather than using the parameters of the fit.

**Disparity Fourier spectrum**

The original energy model predicts that the *disparity peak frequency*, the frequency at which the disparity modulation has most power, should be the same as the *preferred spatial frequency* observed with monocular gratings. In making this comparison, the disparity must
be applied at right-angles to the cell’s preferred orientation. Most previous work using random-dot stimuli in awake animals has employed only horizontal disparities. To enable a test of the energy model prediction, all disparities in the present study were applied orthogonal to the cell’s preferred orientation (Cumming 2002). It is also important that the disparity tuning is measured with a broadband stimulus such as random dots, in order to ensure that the disparity tuning curve shape is not trivially determined by the stimulus. If disparity tuning were measured with a grating, for instance, the periodicity of the stimulus would guarantee a periodic response (Cumming and Parker 2000).

The disparity peak frequency is slightly different from the “disparity frequency” – a term used by two previous authors in related but distinct senses. Ohzawa et al. (1997) used a bar as the broadband stimulus to test this prediction in the anesthetized cat. They used the term “disparity frequency” to mean the carrier frequency of the Gabor fitted to the disparity tuning curve, which they then compared with the monocular spatial frequency tuning. Note, however, that this carrier frequency does not necessarily equal the disparity peak frequency. Thus their finding that the carrier frequency of fitted Gabors was systematically lower than the preferred spatial frequency in the dominant eye is not necessarily at odds with the energy model. For sufficiently narrow-band Gabor functions, the carrier frequency \( f \) and disparity peak frequency coincide (Appendix B), but many of the disparity tuning curves presented by Ohzawa et al. appear to be fairly broadband (e.g. their Figure 15). In this case, the disparity peak frequency and the carrier frequency diverge. Which is higher depends on the phase of the disparity tuning curve (Appendix B). The disparity tuning curves presented by Ohzawa et al. display phases across the spectrum: thus, both situations occur. Furthermore, for Gabors that are not narrowband, the fitted carrier frequency is often poorly constrained by data (Prince et al. (2002b), Figure 6). For these reasons, it is not clear that the data presented by Ohzawa et al. (1997) necessarily violate the energy model.

Prince et al. (2002a) used the term “disparity frequency” to refer to the peak frequency of the Fourier transform of the disparity tuning curve after subtraction of the mean, meaning that for these authors a disparity frequency of zero is impossible by definition. The disparity frequency of Prince et al. (2002a) was designed as a way of extracting a measure of the spatial
scale of disparity tuning which would work for both bandpass and lowpass tuning curves. Neither sense of disparity frequency provides the appropriate measure for comparison with monocular spatial frequency tuning: hence our use of the disparity peak frequency.

We compared two different ways of extracting the disparity peak frequency. The first was completely model-independent; here we used the response to uncorrelated stimuli as an estimate of the baseline firing rate, subtracted that from the disparity tuning curve and took the continuous Fourier transform of the result (by trapezoidal integration). We also estimated the disparity spectrum from the Gabor function fitted to the tuning curve. When performing the fit, the Gabor function was halfwave rectified, i.e. negative values of the fit function were replaced with zeros for the purpose of evaluating the residual (since firing rates could not be lower than zero). When obtaining the disparity spectrum, we used the unrectified Gabor, and solved numerically for the peak and half-maximal points of the Fourier spectrum.

**Bootstrap resampling**

In order to interpret scientific results, it is important to have an estimate of significance, in order to be sure that features we observe in our data are not merely due to the vagaries of finite sampling. Throughout this paper, we have used bootstrap resampling (Efron 1979) to estimate significances. Given a data-set consisting of $n$ samples of the random variable, one generates a “new data-set” by randomly selecting a member of this data-set $n$ times (with replacement). This provides a convenient, non-parametric way to estimate the distribution of some function of a random variable, avoiding the normality assumptions buried in many standard statistical tests. For resampling to be reliable, $n$ must be large. This was one motivation for presenting stimuli for relatively short periods: it provided a large number of independent samples. To increase $n$ further, we pooled the data across all disparities (or spatial frequencies, for the grating stimuli) and resampled the residuals. For this pooling to be valid, the standard deviation must be the same at each disparity, so, as before, we transformed the data by taking its square root. That is, for each disparity we calculated the mean of the square root of the firing rate, and the residual difference between this mean and the square root of the firing rate on each stimulus presentation. We then pooled all these residuals into a single population. To generate a resampled datum, we picked a residual at random from this
pool, added it to the mean square-root firing rate, and squared it to obtain the resampled firing rate. We also explored resampling the data for each stimulus condition separately, and found that this gave closely similar results. In the few cases where the results were different, the method of resampling the residuals generally gave the wider confidence interval. Since this yields more conservative estimates for significance testing, resampling of residuals was adopted throughout. This meant that the effective \( n \) was always >80 for the spatial frequency tuning curves, and always >200 for the disparity tuning. All quoted significances are at the 5% level.

**Classification as simple or complex**

Within the energy model, complex cells are viewed as being made up from the summed output of several simple cells (Equation 3 below). Our analysis holds for both simple and complex cells and our conclusions do not depend on a classification of cells as either simple or complex. For this reason we have not treated simple and complex cells differently, and hence avoided the complications of attempting to make the classification in awake animals in the face of small eye movements.

**Data-set**

We recorded monocular and binocular responses to random-dot stimuli in 210 neurons, at eccentricities between 2° and 10°. Of these, 180 produced a maximum firing rate of at least 10 spikes/s. 138/180 were disparity-selective. Adequate data on spatial frequency tuning was available for 101/138 disparity-selective cells, and in an additional 23 disparity-selective neurons we had data on spatial frequency tuning but not monocular responses to random dots.

**The energy model and our modified version**

This paper represents a critical comparison of the energy model (Adelson and Bergen 1985; Ohzawa et al. 1990; Qian 1994; Fleet et al. 1996; Ohzawa 1998) and our modified version of it, introduced in order to explain the weaker response to anticorrelated stimuli (Read et al.
In this section, we lay out the key features of both models and explain how they differ. Detailed calculations are given in the Appendices.

The building-blocks of all the models considered in this paper are binocular subunits characterized by a receptive field in each eye which performs a linear operation on the retinal image in that eye. The input from each eye, $v_L$ or $v_R$, is the result of this operation (for details, see Appendix C, Equation 12). The distinctive feature of the energy model is that the inputs from the two eyes are combined linearly: the response of a binocular subunit is a function of the sum ($v_L + v_R$) of the inputs from each eye separately. If this sum is negative, the binocular cell is silent, since it cannot signal firing rates below zero. If this sum is positive, the energy model postulates that the binocular cell outputs the square of this sum. Thus, writing $C$ for the output of the disparity-selective cell,

$$C = \left[ \text{Pos}(v_L + v_R) \right]^2,$$

Equation 2

where Pos denotes half-wave rectification. A complex cell is assumed to receive input from several of these half-squared linear binocular subunits, and its response is assumed to be the linear sum of its inputs:

$$C = \sum_{j=1}^{n} \left[ \text{Pos}(v_{Lj} + v_{Rj}) \right]^2$$

Equation 3

This is shown schematically in Figure 1A. Binocular subunits (“BS”) are shown receiving input from left and right eye receptive fields, which for illustration are shown with different phases. Several of these subunits feed into a single complex cell (“Cx”).

Our modified version (Read et al. 2002) differs from the energy model in postulating that inputs from the two eyes are half-wave rectified before being combined:

$$C = \left[ \text{Pos}(v_L) + \text{Pos}(v_R) \right]^2.$$

Equation 4
Figure 1B shows one physiologically plausible implementation of this nonlinearity. In the figure, inputs from the left and right eyes initially synapse onto monocular simple cells (“MS”), which impose an output threshold, before being combined in a binocular subunit. If the inputs are combined with an inhibitory synapse, as in the lower binocular subunit in Figure 1B, we obtain units like

\[ C = \left( \text{Pos}[\text{Pos}(v_L) - \text{Pos}(v_R)] \right)^2 \]

Equation 5

(the additional Pos means that the cell does not fire when suppression from the right eye exceeds excitation from the left). Once again, complex cells are constructed from the sums of several binocular units of the type given in Equation 4 and Equation 5 (Read et al. 2002). The distinction between the two types does not matter in the energy model: there is no need to explicitly include subtypes based on \((v_L-v_R)\) as well as \((v_L+v_R)\), because \((v_L-v_R)\) is equivalent to \((v_L+v_R)\) with a phase change of \(\pi\) in the right eye’s receptive field.

Results

Overview

Disparity selectivity and monocularly

We present evidence that some cells receive purely suppressive input from one eye. We show that this is inconsistent with the linear binocular combination of the energy model, but can be explained in our nonlinear model.

Spatial frequency tuning in the two eyes

Motivated by the assumption of the original energy model that receptive fields are identical up to phase, we investigate whether there is evidence for differences in spatial frequency tuning between eyes. We find that tuning in most cells agrees well, but a minority show significant differences.
Disparity frequency and spatial frequency tuning

If the assumptions of the original energy model hold, then the disparity frequency should equal the preferred spatial frequency in the dominant eye. We show that this prediction is systematically violated, and that vergence movements cannot account for the difference. However, this prediction applied only to the original energy model, which included strong constraints on the receptive field profiles in addition to linear binocular combination.

Generalizing the energy model

We therefore generalize the energy model to allow for receptive fields with different phases, positions, and spatial frequency tuning (both across subunits, and across eyes within a subunit). We derive a constraint which even this generalized model must fulfil. We show that the data systematically violates this constraint.

Thresholding prior to binocular combination

We finally show that our modified version of the energy model, in which a threshold precedes binocular combination, can account for the observations on disparity and spatial frequency tuning.

Disparity selectivity and monocularity

Some cells receive purely suppressive input from one eye

Cells which are sensitive to binocular disparity must receive information from both eyes. It is tempting to extrapolate from this that the cells which are most sensitive to binocular disparity must be those which respond most nearly equally to input in either eye. However, previous workers (Prince et al. 2002a; Poggio and Fischer 1977; Ohzawa and Freeman 1986b; Smith et al. 1997) have found little support for this idea. In agreement with these studies, we find no relationship between monocularity and disparity selectivity. Many cells which respond nearly equally to monocular stimulation in either eye are not disparity selective, while many cells which show little or no response to monocular stimulation in one of the eyes nevertheless show clear disparity tuning. Examples are shown in Figure 3 (see also Figure 8). The response to monocular stimulation is shown by the broken horizontal lines labeled L and R (and
marked with a leftward / rightward-point arrowhead respectively: \(<\) and \(\triangleright\). In dufo96 (Figure 3A), monocular stimulation in the left eye evokes almost no response; in dufo99 (Figure 3B), it is the right eye which is silent. Yet the black dots show the cells’ responses as a function of disparity (curve = fitted Gabor); clearly both cells are selective to disparity, and so must be receiving information from both eyes. Thus, it is important to distinguish between two common uses of the term “monocular”: the classical sense, “responsive to monocular stimulation in one eye and not the other”, must not be interpreted to mean “receiving input from only one eye” (Ohzawa and Freeman 1986a, b; Smith et al. 1997).

One natural way to explain the phenomenon of disparity selectivity in “monocular” neurons is to propose that the input from one eye always has a net inhibitory effect, and thus no spikes are produced by stimulation in that eye alone. In the absence of complications such as response normalization (which could adjust the response to monocular stimuli relative to binocular), such a scheme makes two predictions. First, binocularly uncorrelated dots should produce a weaker response than monocular dots in the dominant eye (because adding dots to the other eye produces net inhibition). This was the case for 86/138 disparity-tuned cells (significant in 44). Second, the monocular response in the non-dominant eye should not be significantly greater than the spontaneous response (it is rarely possible to observe a monocular response less than spontaneous, since the latter is so frequently indistinguishable from zero). 30/138 disparity-selective neurons showed both these phenomena. In 9/30 cells, the spontaneous response was significantly greater than zero, so that if one eye had an inhibitory influence, it would be possible to observe suppression of the spontaneous response when this eye was stimulated. In 5 of these 9, the response to random dot stimulation in the non-dominant eye was smaller than the spontaneous response. The cells shown in Figure 3 are two examples. The broken line labeled U and marked with a square \(\Box\) shows the response to uncorrelated stimuli; in both cells this is below the response to monocular stimulation in the dominant eye, i.e. adding stimulation to the non-dominant eye has reduced the response. The broken line labeled S and marked with a circle \(\bigcirc\) shows the spontaneous rate, estimated from the response to a blank screen: in both cells, the cell fires more to a blank screen than to monocular stimulation in the non-dominant eye. Such examples are clearly indicative of a
predominantly inhibitory input from one eye. We conclude that in many cells, stimulation in one of the eyes always has a suppressive effect.

**This indicates a nonlinearity prior to binocular combination**

It is important to note that this represents a substantial deviation from any model in which binocular summation is linear. By definition, a model with linear binocular summation is of the form \( C = f(v_L + v_R) \), where \( f \) is an arbitrary function, and \( v_L, v_R \) represent the inputs from left and right eyes. If \( f(v_L) \) is never positive for any value of the input \( v_L \), either positive or negative (no possible stimulus in the left eye elicits a positive response), then \( f(v_R) \) can never be positive either, and so the cell would never respond. In a linear model, if the cell responds at all, then stimulation in each eye can exert either a suppressive or an enhancing effect, depending on the stimulus. To obtain the situation where one eye always exerts a suppressive effect, we must postulate some nonlinearity prior to binocular combination, such as halfwave rectification followed by an inhibitory synaptic connection. This is exactly what is proposed by our modified version of the energy model. Looking at Equation 5, \( C = \{\text{Pos}[\text{Pos}(v_L) - \text{Pos}(v_R)]\}^2 \), it is obvious that stimulation in the right eye always has a suppressive effect. For monocular right-eye stimulation, the response \( C \) is zero, and yet with disparate binocular stimuli, this unit is disparity-selective. Figure 4 shows simulations of two subunits described by Equation 5. The solid line shows the disparity tuning curve. In A, the left and right receptive fields are identical, so – since input from one eye is inhibitory – the disparity tuning curve is of the tuned-inhibitory class. In B, the left and right receptive fields are 180° out of phase. When combined with the inhibitory synapse in Equation 5, this results in tuned-excitatory disparity tuning. This demonstrates that an inhibitory synapse at binocular combination does not necessarily result in tuned-inhibitory tuning. Thus, our thresholding model explains the existence of cells which would classically be called “monocular” and yet are disparity-selective.
A thresholding nonlinearity can explain the relative amplitude of monocular and binocular responses

We now investigate the extent to which this model can account quantitatively for the relative amplitude of monocular and binocular responses. Prince et al. (2002a) observed that the response to binocularly uncorrelated dot patterns was often close to the mean of the responses to monocular stimulation in the two eyes, whereas the energy model predicts that it should be their sum. Prince et al. suggested that this could be due to a normalization process which lowers the response to binocular stimuli. However, our modification to the energy model already allows us to build cells in which the uncorrelated response is the mean of the two monocular responses, without incorporating any normalization. The horizontal lines in Figure 4 show the response to monocular stimuli (L□, R□) and binocularly uncorrelated stimuli (U□). In both cases, the uncorrelated response is close to the mean of the monocular responses, demonstrating that our model can explain this phenomenon, for both tuned-excitatory and tuned-inhibitory cells.

These simulations portray something of an extreme case: in both these examples, inhibition from the suppressive eye is much stronger than excitation from the excitatory eye, so that the response to monocular stimulation in the dominant eye, M, is nearly twice the response U to uncorrelated stimuli. In fact, 2U is an upper bound for M: our model predicts that M can never exceed 2U. The energy model has a similar upper bound: it predicts that M can never exceed U. We have seen that the energy model’s upper bound is violated by most cells (86/138). We now investigate whether the upper bound predicted by our model is similarly violated. Figure 5 shows the distribution of M/U for the 138 disparity-selective cells in our data-set. The vertical lines mark the upper bounds predicted by the energy model (dashed) and our model (solid). The mode of the distribution is close to M/U=1, so over half the cells exceed the energy model upper bound. However, the distribution begins to fall off after M/U=2, so that the upper bound predicted by our model is violated in only 23/138 cells. We used resampling to estimate the 95% confidence interval for M/U. If this interval lies entirely above 1, we can be 95% confident that the upper bound predicted by the energy is violated; this was the case for 44/138 cells (32%), shaded gray in Figure 5. If this interval lies above 2, we can be 95% confident that the upper bound predicted by our model is violated; this was so for only 4/138
cells (3%), shaded black in Figure 5. We conclude that almost all cells respond to monocular stimulation in the dominant eye at less than twice the rate for uncorrelated stimuli, and can therefore be accommodated within our modified model. Thus, our model can explain the observed spectrum of monocular and binocular response rates, without needing to invoke other mechanisms such as contrast normalization.

Spatial frequency tuning in the two eyes

The original implementation of the energy model (Ohzawa et al. 1990) assumed that all receptive fields have the same spatial frequency and orientation tuning and bandwidth. They differ only in their amplitude, their position and phase, and even so, the position and phase disparity between left and right receptive fields of a single subunit is assumed to be the same for all subunits. These constraints on the receptive fields have been assumed by all implementations of the energy model we are aware of (e.g. (Qian 1994; Fleet et al. 1996; Ohzawa et al. 1997; Lippert and Wagner 2001; Read 2002; Tsai and Victor 2003)). We shall use the phrase original energy model to denote Equation 3 with these additional constraints on the receptive fields. (Later, we shall consider a generalized energy model in which many of these constraints are relaxed.)

The available evidence suggests that these constraints are generally observed in simple cells (Ohzawa et al. 1996; Anzai et al. 1999b). In complex cells, the situation is harder to assess. Preferred orientation is observed to be closely matched between the two eyes (Bridge and Cumming 2001), supporting the view that all receptive fields share the same orientation. However, there is some evidence from the cat suggesting that there may be a population of cells in which spatial frequency differs between the two eyes (Ohzawa et al. 1996; Hammond and Pomfrett 1991). In this section, we investigate the agreement in spatial frequency tuning for our monkey data.

For 151 cells, the spatial frequency tuning to monocular drifting sinusoidal luminance gratings at the cell’s preferred orientation was probed in both eyes. 84 of these were sufficiently responsive and selective to permit fitting in both eyes. We defined the preferred spatial
frequency to be the frequency at which the Gaussian fitted to the tuning curve had its peak. In order to ensure this is meaningful, we required the fits in each eye to explain more than 60% of the variance of the tuning curve data. Figure 6 compares the preferred spatial frequency in the two eyes for the remaining 73/84 cells. The solid line shows the identity; the dotted lines mark difference in SF tuning of 1 octave. Clearly, spatial frequency tuning is usually well matched between eyes. The correlation coefficient is 0.87 ($p < 10^{-5}$). Nonetheless, 25/73 cells showed a significant difference ($p < 0.05$, by resampling) in preferred spatial frequency between the eyes; these are colored black in Figure 6. There was no correlation between the difference in preferred spatial frequency and the difference in peak response between the two eyes. The figure of 25 includes some cells where the difference in preferred frequency was small (but turned out to be significant because the peak positions were robust under resampling). However, for 6/25, the peak firing rates in the two eyes occurred for gratings differing in frequency by over an octave. Two examples are shown in Figure 7. The arrowheads show the response of the cell to monocular grating stimuli as a function of the grating spatial frequency (L: $\leftarrow$, R: $\rightarrow$); the curves show the fit. The 95% confidence interval for the peak of the fitted function is shaded. The confidence intervals for the two eyes do not come close, indicating significant and substantial differences in spatial frequency tuning between the two eyes. Around 10% of cells showed evidence of such a difference.

The selection criteria applied in obtaining Figure 6 exclude an interesting class of cells in which the response in the non-dominant eye was very weak, but was maximal at those frequencies which produced the weakest responses in the dominant eye. Two examples are illustrated in Figure 8AB. On the face of it, these cells show a severe mismatch in spatial frequency tuning, with the non-dominant eye being tuned to frequencies an order of magnitude lower than the dominant eye. However, we believe a more plausible explanation is that the spatial structure of the receptive fields is really similar in the two eyes (as in the vast majority of cells, Figure 6), but that the non-dominant eye exerts a suppressive effect. This interpretation is supported by the experiments with random-dot patterns. Both these cells are disparity-selective, but also show virtually no response to random dots in the non-dominant eye (Figure 8CD). Thus, such cases are further evidence for purely inhibitory input from one eye.
**Disparity frequency and spatial frequency tuning**

We now turn to possibly the most important prediction of the energy model: namely, that the shape of monocular receptive fields determine the shape of the disparity-tuned response (Ohzawa et al. 1997; Anzai et al. 1999b). Since most cells do indeed show similar spatial frequency and orientation tuning in the two eyes, we shall assume in this section that the assumptions of the original energy model hold true. Then, the original energy model predicts that the disparity tuning curve is simply the cross-correlation of the receptive fields in the left and right eyes.

For simple cells, which are single binocular subunits (Equation 2), this prediction can be tested directly. For complex cells, which represent the sum of several binocular subunits (Equation 3), the disparity tuning curve is predicted to be the sum of the cross-correlations of the receptive fields in the component subunits. This makes the prediction hard to test in complex cells, as it is difficult to obtain the receptive fields of the component subunits experimentally. Fortunately, provided all subunit receptive fields have the same preferred orientation, the comparison can be made without a direct measurement of receptive field profile. We simply need to obtain (a) the cell’s response to binocular random-dot patterns as a function of disparity along an axis orthogonal to this preferred orientation, and (b) the cell’s response to monocular sinusoidal gratings oriented parallel to this preferred orientation, as a function of spatial frequency. The energy model predicts that the shape of the Fourier amplitude spectrum of the disparity tuning curve measured in (a) will be given by the monocular spatial frequency tuning curves measured in (b). In particular, their peaks should coincide: that is, the disparity peak frequency, defined as the position of the peak in the Fourier amplitude spectrum, should be the preferred spatial frequency of the cell. This key prediction of the original energy model, which depends critically on its linear properties, holds for both simple and complex cells. Previous work (Ohzawa et al. 1997; Prince et al. 2002a) has suggested that this prediction is not fulfilled, but, as discussed above, these studies leave open a number of possible ways in which the data could be reconciled with the energy model.
We carried out a detailed comparison, using bootstrap resampling to estimate the significance of any discrepancy.

Figure 9 shows the comparison for three neurons, illustrating the common patterns observed. The lefthand column shows the disparity tuning curves. On the right, the Fourier amplitude spectrum of the disparity-modulated component is compared with the spatial frequency tuning in the dominant eye. For both these quantities, two estimates are shown: one from the raw data and one from the fitted function. The raw spatial frequency tuning curves are shown with black dots in the plots on the right, while the fits are drawn with the black curve. The disparity-modulated component can be estimated from the raw data by subtracting the mean response to uncorrelated stimuli (horizontal line labeled with the letter U and the symbol in the lefthand plots) from the mean response of the cell to random-dot stereograms at different disparities (black dots in the plots on the left). The Fourier spectrum of this is shown on the right with a dotted gray line (“FT-DMC (data)” in the legend). Alternatively, the disparity-modulated component can be obtained from the fitted Gabor (solid curve in the lefthand plots). The Fourier spectrum of this is shown on the right with a solid gray line (“FT-DMC (fit)”).

In a few cases (Figure 9A) the Fourier transform of the disparity-modulated component (FT-DMC) did closely resemble the spatial frequency tuning curve (SFTC), but for the majority of cases there were substantial discrepancies, of two types. First, the peak of the FT-DMC was often at a lower frequency than the peak of the SFTC (Figure 9B). Second, the FT-DMC was often close to lowpass in form, despite a clear bandpass SFTC (Figure 9C).

We had 105 disparity-selective neurons which were sufficiently responsive to gratings in the dominant eye, selective for spatial frequency, and adequately described (> 60% variance explained) by the Gaussian fit. In order to avoid making the assumption that all disparity tuning curves were well described by Gabor, we also used a model-independent estimate of the disparity peak frequency, using the response to uncorrelated stimuli as an estimate of the baseline of the disparity tuning curve, and taking the continuous Fourier transform of the raw data. We compared this estimate of disparity peak frequency with the SFTC peak frequency of
the Gaussian fit, for the 105/112 cells in which the response to uncorrelated stimuli was available and in which the Gaussian fitted to the SFTC explained >60% of the variance. The disparity peak frequency was less than the SFTC peak frequency in 84/105 of cells ($p<10^{-9}$ under the null hypothesis that the estimated disparity peak frequency is as likely to be above the SFTC peak frequency as below it (binomial distribution)). The frequency difference was individually significant in 43/84 cells.

This model-independent method of extracting the disparity peak frequency has two disadvantages. First, in ~10% of cells, the disparity tuning curve appeared to be truncated by the lower limit of 0 spikes/s. These cells may represent an energy model unit followed by an output threshold. It is possible that the discrepancy between the disparity peak frequency obtained from the raw data, and the SFTC peak frequency, may reflect distortions introduced into the Fourier spectrum by the threshold. For these cells, a better estimate of the underlying response may be gained from the unrectified Gabor corresponding to the half-wave rectified Gabor fitted to the data (shown in Figure 9A). Second, because the Fourier spectra of raw data are usually noisy and multimodal, it is hard to extract measures of bandwidth. Again, this is solved by using the fitted Gabor.

For those 99 cells in which the Gabor fitted to the disparity tuning curve explained >60% of the variance, we therefore repeated the analysis using the estimate of disparity peak frequency derived from the fit. The results are shown in Figure 10A, which plots the disparity peak frequency against the preferred spatial frequency in the dominant eye, both derived from the fitted functions. The solid line marks the identity line; according to the energy model, all points should lie on this line. In fact, the SFTC peak frequency was greater than the disparity peak frequency in 80/99 cells ($p<10^{-9}$, binomial), and the difference was significant in 51/80 individual cells (resampling; these are the filled symbols in Figure 10A). Thus, we obtain very similar results whether we use the fitted Gabor or the raw disparity tuning curve.

We also examined the high and low cutoff frequencies of the fitted functions. Figure 10B,C plot the cutoff frequencies for the disparity tuning curve against those for the spatial frequency tuning curves. Again, the energy model predicts that all points should lie on the
identity line (marked with the solid line). In fact, the low-cuts differed significantly in 43/99 of cells, while the high-cuts differed in 67/99 (filled symbols). Once again, these significant differences nearly all reflect relatively more power at low frequencies in the FT-DMC than in the SFTC.

In many cases, there is so little attenuation of the FT-DMC at low frequencies that the low-cut frequency is not defined (plotted as a low-cut of zero). The discrepancy in the response at very low frequencies is made clearer in Figure 10D, which compares the relative power at the lowest frequency tested monocularly. In 77/99 cells, the FT-DMC contains more power at these low frequencies than the SFTC ($p < 10^{-6}$, binomial). In 45/77 cases, this difference is significant (filled symbols). In many cases, the disparity tuning curve is close to Gaussian in form (relative power=1, i.e. no attenuation at low frequencies). That this occurs in the presence of a bandpass SFTC is a dramatic deviation from the energy model. The bandpass SFTC implies that the receptive fields of the subunits contain both “on” and “off” subregions. At a disparity equal to the separation of the “on” and “off” regions, the contributions from the left and right eyes should be negatively correlated, producing a response which is smaller than the response to uncorrelated dots. These suppressive side-lobes in the disparity tuning curves are often not found. Prince et al. (2002a) also noted that many disparity tuning curves were Gaussian in form. However, those data were not clearly at odds with the energy model for two reasons. First, disparity was applied horizontally, regardless of receptive field orientation, so it remained possible that suppressive side-lobes would have emerged if disparity had been applied orthogonally to preferred orientation. Second, their data on spatial frequency tuning were not generally sufficient to exclude the possibility of a substantial lowpass component in the monocular SFTC. The present data eliminate both of these difficulties, and clearly indicate a need for more complex processing than the original energy model can provide.

**Vergence eye movements are unlikely to explain the mismatch**

One possible explanation of the mismatch between disparity tuning and spatial frequency tuning is that the monkeys may be making small vergence movements. This would have the effect of introducing jitter into the disparity of each stimulus. Effectively, we would be
summing several disparity tuning curves of the form predicted by the energy model, each with a random disparity offset. This tends to smear out the sidelobes, shifting the peak of the disparity power spectrum to lower frequencies. This process is illustrated in Figure 11. Consider an energy-model binocular subunit, whose receptive fields in both eyes are identical, with no position disparity, both having the profile shown in Figure 11A. The thin line in Figure 11B shows the disparity tuning curve which would be obtained for this subunit in the absence of vergence movements. Now suppose the monkey makes random vergence movements, so that the disparity of his actual fixation point relative to the fixation target at any moment is a Gaussian centered on zero. This means that the disparity tuning curve actually measured is the true curve convolved with this Gaussian. The result is shown with the thick line in Figure 11B. Figure 11C shows the effect on the Fourier amplitude spectrum of the disparity-modulated component. The thin line shows the power spectrum for the original tuning curve, and the thick line for the observed version contaminated by vergence. The vergence has had two effects: it has shifted the peak of the disparity power spectrum towards DC, and it has greatly reduced the amplitude. For clarity, therefore, the broken line shows the observed disparity power spectrum scaled up to the same amplitude as the uncontaminated one. The same effect would be obtained if the neuron being recorded from represented the sum of several subunits which differed in their position disparities.

Either of these possibilities might explain why, in 80/99 cells, the peak of the disparity power spectrum is at a lower frequency than the preferred spatial frequency obtained with gratings. To estimate the vergence jitter which would be necessary to achieve this, we assume that the underlying disparity tuning curve is a Gabor with disparity peak frequency equal to the preferred spatial frequency in the dominant eye, and that the Gabor function fitted to the observed disparity tuning curve represents this underlying disparity tuning curve convolved with a Gaussian distribution of vergence. In over half the cells (48/80), the amount of vergence jitter needed to bring about the requisite shift in frequency is larger than the SD of vergence reported by the scleral search coils, even though search coils clearly overestimate variability in vergence (Read and Cumming, 2003). It therefore seems unlikely that the animal’s vergence jitter is large enough to explain the mismatch in peak frequency in most cells.
Generalizing the energy model

It remains possible that combinations of subunits with different position disparities might be responsible for the lower disparity frequency. However, even if this is the case, the energy model places an upper limit on the power at any frequency. This can be appreciated by inspecting Figure 11. The multiple subunits have shifted the peak towards lower frequencies, but they have done this by removing power at high frequencies rather than by adding power at any frequency. If multiple subunits are responsible for the mismatch between disparity peak frequency and preferred spatial frequency, the overall power in the disparity tuning should be greatly reduced. We therefore generalized the energy model to examine the possibility that such scatter in positions could explain our data.

In the original formulation of the energy model (Ohzawa et al. 1990), all binocular subunits were assumed to have the same phase disparity, position disparity, spatial frequency and orientation tuning. We now allow an arbitrary number of subunits, with different phases, positions, and spatial frequency tuning (both across subunits, and across eyes within a subunit), including monocular subunits and binocular subunits which are not tuned to disparity. We shall, however, continue to assume that all receptive fields have the same orientation, and the same profile parallel to this orientation (see Appendix E). With this much weaker set of constraints on the receptive fields, it is no longer true that the disparity Fourier spectrum must have the same shape as the spatial frequency tuning. However, it can be shown (Appendix E; Equation 29) that the disparity power spectrum, $|\tilde{D}(f)|^2$, and the product of the monocular spatial frequency tuning curves, $L_{SF}(f)R_{SF}(f)$, must still satisfy the following inequality, for every spatial frequency $f$:

$$\frac{L_{SF}(f)R_{SF}(f)}{L_{ASF} R_{ASF}} \geq \frac{|\tilde{D}(f)|^2}{U^2}.$$  

Equation 6

The monocular spatial frequency tuning curves have been normalized to unit area by dividing by the area under each curve, $L_{ASF}$ and $R_{ASF}$. The disparity power spectrum has been
normalized by dividing by the squared response to uncorrelated random-dot stereograms $U^2$. This has the advantage of canceling out any differences in the overall response to sine gratings and to random-dot stimuli (e.g. because random dots have less contrast power in the cell’s spatial bandpass, or because of stimulus-dependent change in width- or end-stopping). Since such effects apply equally to disparate and uncorrelated stereograms, they would cancel out in Equation 6. For the original energy model, Equation 6 holds with an equals sign.

This inequality allows us to detect whether the lower disparity frequency can be explained by the presence of multiple subunits with different positions, as in Figure 11. Then, the shift in the peak frequency would be achieved not by boosting the disparity power at low frequencies, but only by removing power at high frequencies. This would substantially weaken the disparity modulation (cf. Figure 11B), so that the inequality Equation 6 would be satisfied. If the inequality is violated, then multiple subunits are not a sufficient explanation.

The same generalization allows for differences in spatial frequency tuning between eyes. Our analysis in the previous section (like that of Ohzawa et al. (1997)) considered only the dominant eye, and assumed that the SFTC in the non-dominant eye differed only by a scaling of response magnitude. Yet our results (Figure 6, Figure 7) show that there is a significant minority of cells in which spatial frequency tuning differs between eyes. The inequality of Equation 6, which includes terms for the spatial frequency tuning in each eye, holds for the generalized energy model even in this case.

The top row of Figure 12 (ABC) examines Equation 6 for one cell, ruf030, which satisfied the energy model prediction reasonably well. Figure 12A shows the spatial frequency tuning in the two eyes together with the fitted functions; Figure 12B shows the disparity tuning curve, with the 95% confidence interval at each disparity shaded. Figure 12C compares the fitted grating curve $L_{SF}(f)R_{SF}(f)/L_{ASF}R_{ASF}$ (black) and disparity spectrum $|\bar{D}(f)|^2 / U^2$ (gray). The disparity peak frequency is at 0.82, whereas that of the grating curve is at 0.24. This difference was significant, so the data are incompatible with the simplest form of the energy model. Since the disparity peak frequency is, atypically, higher than the SFTC peak frequency, this cannot be explained by vergence eye movements. However, since the
inequality of Equation 6 is not significantly violated at most frequencies, the data are fairly compatible with our generalized version of the energy model, incorporating many subunits with different spatial frequency and/or disparity tuning. Note that the odd-symmetric disparity tuning in this cell cannot arise simply from a phase disparity of $\pi/2$ between receptive fields whose properties are otherwise identical, as originally envisaged by Ohzawa et al. (1990), since this would require the normalized curves in Figure 12C to peak at the same frequency. One possibility is that this cell receives input from one tuned-excitatory subunit with a position disparity of 0° and one tuned-inhibitory subunit with a position disparity of 0.5°, resulting in the odd-symmetric curve. Such a scheme is allowed within this generalized version of the energy model, but is nevertheless very different from previous explanations of odd-symmetric disparity tuning (Ohzawa et al. 1990; DeAngelis et al. 1991). Some properties of such a model are discussed in Read et al. (2002).

More typical results are shown in the bottom two rows of Figure 12. The cell in Figure 12F is an example where the peak frequencies (marked by vertical lines) are close, so that an analysis only of the peak frequency would conclude that it is consistent with the energy model. But, over a range of low frequencies, the scaled disparity power spectrum is significantly higher than is possible under the energy model. The spatial frequency tuning curves indicate no response to a DC stimulus, so according to the energy model the disparity tuning curve should have no DC component. In the energy model, the disparity tuning curve of a complex cell is the sum of the disparity tuning curves for the individual subunits. This cannot produce a curve with a DC component that is absent from all the subunits. The bottom row of Figure 12 (GHI) shows another example which severely violates the inequality. The spatial frequency tuning curves peak at high frequencies: over 10 cycles per degree. In contrast, the disparity tuning curve has most power at DC, and no power at all at 10cpd. Once again, the power at DC is far beyond what could be accounted for by multiple subunits. The three example cells shown in Figure 9 also violate the inequality.

To quantify the results across the entire data set, we used the difference

$$\Delta = \frac{L_{sf}(f)R_{sf}(f)}{L_{asf}R_{asf}} - \frac{|\tilde{D}(f)|^2}{U^2}.$$
This is the difference between the black curve and the gray curve in the right-hand column of Figure 12 (CFI). The energy model, generalized to include many subunits, predicts that $\Delta$ should be positive or zero for all frequencies (Equation 6); i.e. black curve above gray curve in Figure 12CFI. Using the functions fitted to the disparity and spatial frequency tuning, we evaluated $\Delta$ at a range of frequencies. We used resampling to estimate the 95% percentile of this difference; if this percentile is negative, we can reject the inequality of Equation 6 with 95% confidence.

We had 83 disparity-tuned cells for which spatial frequency tuning curves (SFTCs) were available in both eyes, and for which all three fitted functions explained more than 60% of the variance. For frequencies lower than the peak of the product of the SFTCs, the hypothesis $\Delta \geq 0$ could confidently be rejected for over half of the cells. The significance of the rejection in individual cells was often very high. At a frequency of 0.01cpd, the hypothesis could be rejected with 95% confidence for 47/83 cells, and with better than 99.99% confidence for 13/83 cells. Figure 13 shows how the proportion of cells for which the energy model can be rejected (i.e. $\Delta$ significantly less than zero) varies depending on the frequency at which $\Delta$ is evaluated. In order to make a meaningful definition of “low frequencies” across a population with differing spatial frequency tuning, we express frequencies relative to the peak. This shows clearly that, whereas at high frequencies most cells appear consistent with the generalized energy model, significant discrepancies emerge as we move towards lower frequencies. Thus, although a more general version of the energy model can explain the differences in the peak frequencies previously reported (Ohzawa et al. 1997; Prince et al. 2002a), the absolute power at low frequencies in the FT-DMC still allows us to reject the energy model in over half the cells. Note that this analysis also demonstrates that no amount of vergence variability could explain the data in these cells. If the vergence variability was large enough to account for the necessary shift in disparity peak frequency, it would have reduced the amplitude of disparity tuning below that observed. We conclude that the disparity-selective responses of many V1 neurons are not compatible with the energy model, even after it has been generalized to include multiple subunits with different spatial frequency and disparity tuning.
The possible sources of error cannot explain the mismatch

Spontaneous firing rates
An overestimate of the area under the SFTC would yield excessively small values for the normalized product of SFTCs, which might cause erroneous rejection of the energy model. This could occur if a substantial spontaneous firing rate was added to the response resulting from receptive field structure. However, the observed spontaneous rates were almost always very low. For 211 of the 252 cells, blank stimuli, consisting of a uniform gray screen at the mean luminance of the random dots, were interleaved with the disparate random dots. The mean blank response, averaged over these 211 cells, was 2.1 spikes/s. The mean response exceeded 10 Hz in only 13/211 cells.

Orientation misalignment
Care was taken to ensure that disparity tuning and spatial frequency tuning were measured along the same axis, but of course this may not have been exactly orthogonal to the receptive field orientation. Such a misalignment will mean that Equation 6 does not hold in general, and the effects will depend upon the monocular receptive field. For Gabor receptive fields which are longer in the direction parallel to the carrier Gabor than in the direction orthogonal to it, such misalignment always moves the peak of the product of the monocular spatial frequency curves to lower frequencies, while having less effect on the disparity frequency (Appendix F). Thus, for these receptive fields, the disparity frequency should be if anything higher than the frequency of the peak of the product of monocular spatial frequency tuning curves, the opposite of the observed pattern. Thus, it seems unlikely that our results reflect any artifact of this kind.

Thresholding prior to binocular combination can explain the mismatch
It is therefore clear that the energy model, either in its original form or as generalized here, cannot account for the data. More elaborate extensions of the energy model are needed in order to reconcile it with experimental observations. We have previously (Read et al. 2002)
extended the energy model by adding thresholds prior to binocular combination (Equation 4). We introduced this modification in order to explain the reduced response of V1 neurons to binocularly anti-correlated stimuli. We have already seen that this same modification also explains the phenomenon of disparity selectivity in classically “monocular” cells. We now show that the same modification also explains the existence of low-pass disparity tuning curves in cells which have band-pass spatial frequency tuning.

Figure 14 compares disparity and spatial frequency tuning for the generalized energy model (top row, ABC) and for our modified version with two different sets of parameters (bottom two rows: DEF, GHI). The layout is the same as Figure 12: the left-hand column shows the spatial frequency tuning curves obtained with grating stimuli, the middle column shows the disparity tuning curve and the response to uncorrelated random-dot patterns, while the right-hand column compares the normalized Fourier power spectrum of the disparity-modulated component (FT-DMC, gray curve) and the product of the normalized spatial frequency tuning curves (SFTCs, black curve). We have chosen an example in which the spatial frequency tuning is different in the two eyes: the left eye is tuned to a spatial frequency of 2.5cpd, and the right eye to 3.5cpd.

In the top row, we show results for a single binocular subunit, combining left- and right-eye inputs linearly in accordance with the energy model. Since there is only a single subunit, the normalized FT-DMC is identical to the product of the normalized SFTCs (Figure 14C) i.e. Equation 6 holds with an equals sign (cf. Appendix E). The inequality is thus satisfied.

The middle row, Figure 14DEF, shows the same quantities for a binocular subunit in which the inputs from left and right eyes have been passed through a high threshold nonlinearity before being summed and squared. This has completely removed the inhibitory sidelobes which were present for the energy model subunit in Figure 14B, resulting in a low pass disparity power spectrum. The inequality of Equation 6 is violated, just as in real data.

The bottom row, Figure 14GHI, shows a similar nonlinear model, this time with an inhibitory synapse resulting in a tuned-inhibitory-type disparity tuning curve (since the left and right
receptive fields are in phase). For a single subunit, the spatial frequency tuning curve in the inhibitory eye would be zero, so the product of the SFTCs would be zero and the inequality would be severely violated. Since it was rare for a completely silent SFTC to be recorded in one eye, this would be a rather extreme example. Instead, we make our model cell the sum of two binocular subunits, with identical left and right receptive fields. In the first subunit, the left eye is excitatory and the right eye inhibitory; in the second subunit, it is the other way around. Thus, the response observed for grating stimuli in the left eye is entirely due to the first subunit, while the response to gratings in the right eye is entirely due to the second subunit. But both subunits contribute to the disparity tuning curve. Thus, in our modified version of the energy model, power present in the disparity tuning curve does not necessarily show up in the spatial frequency tuning curve. This is why the disparity power spectrum for this model cell rises well above the product of the spatial frequency tuning curves at low frequencies (Figure 14I). This behavior is incompatible with the generalized energy model, but is commonly observed experimentally (Figure 12FI). We conclude that our modified version of the energy model provides a better match to the data.

**Discussion**

The energy model of binocular complex cells has been successful in qualitatively capturing several aspects of their function. However, several quantitative predictions are not borne out by experimental data. In this paper, we document two such discrepancies. We demonstrate that in each case, the agreement with data is better if we modify the energy model by adding threshold nonlinearities prior to binocular combination.

The first problem for the energy model is that there are many cells in which input from one eye appears to be predominantly inhibitory. The energy model is linear up to binocular combination, and this means that inputs from each eye can be both positive or negative, depending on the particular stimulus. For example, if a receptive field in one eye consists of a single ON region, then a random-dot pattern in which mainly white dots happen to fall on the receptive field will elicit an excitatory response from this eye, whereas a pattern in which the receptive field is covered predominantly with black dots will elicit an inhibitory effect. Thus,
we cannot classify this eye as being “excitatory” or “inhibitory”. However, previous workers have noted cases where one eye appears to have a purely inhibitory effect (Ohzawa and Freeman 1986a; Poggio and Fischer 1977).

We investigated this by carrying out a quantitative comparison of the response rates to monocular and binocular random-dot stimuli. In agreement with previous reports, we find that many neurons show little or no response to stimulation in one eye, despite exhibiting clear disparity selectivity when tested with binocular stimuli. Indeed, in the small number of cells which show significant spontaneous firing, stimulation in the non-dominant eye often reduces the response below the spontaneous level. We also find that in many cells which respond well to random-dot patterns in the dominant eye, adding dots to the non-dominant eye reduces the response. The lack of this second, critical piece of evidence may explain why the challenge posed to the energy model by monocular disparity-selective neurons has so far been ignored. Further support for the idea of inhibitory input from one eye is provided by our study of spatial frequency tuning curves with grating stimuli. Spatial frequency tuning in most eyes is well matched between eyes, and for some of the cells where the peak frequencies appear mismatched, the data suggest that the underlying receptive field structure may be identical in the two eyes, but that one eye has an inhibitory effect. This is impossible in the energy model.

The observation of cells in which one eye has a net inhibitory effect implies a non-linear operation prior to binocular combination. We have previously proposed modifying the energy model to incorporate such a nonlinearity: halfwave rectification followed by an inhibitory synapse (Read et al. 2002). This modified version of the model can explain our observations; we present simulations of classically “monocular” cells which are nevertheless tuned to disparity. The same modification enables us to construct cells in which the response to uncorrelated stimuli is close to the sum of the responses to monocular stimuli in the two eyes, as often observed experimentally, rather than fixed at the mean of the monocular responses, as required by the energy model. We conclude that this modified version of the energy model is a better explanation of the data.
The second discrepancy investigated in this paper is the mismatch between disparity frequency and spatial frequency tuning (Ohzawa et al. 1997). In the original energy model, the disparity peak frequency – loosely, the frequency of the undulations in the disparity tuning curve – should be the same as the preferred spatial frequency observed with monocular sinusoidal luminance gratings. In fact, the disparity peak frequency is systematically found to be lower than predicted by the energy model. Crucial to this demonstration was the use of disparities applied at right-angles to the preferred orientation of the cell. Previous work in monkeys using only horizontal disparities failed to make clear this failure of the energy model (Prince et al. 2002a).

We considered the possibility that differences in spatial frequency tuning between the eyes are responsible for this discrepancy. Although the energy model in its original form assumes identical spatial frequency tuning in left and right eyes, we found a few clear examples of excitatory inputs from both eyes with different spatial frequency tuning, as has previously been reported in the cat (Hammond and Pomfrett 1991). The energy model can easily be generalized to take such interocular differences into account. If the eyes differ in spatial frequency tuning, then the disparity peak frequency should be located at the peak of the product of the tuning curves from left and right eyes. (Note that this implies that the disparity peak frequency is always in between the preferred spatial frequencies in the two eyes.) However, we showed that even this generalized version of the energy model must obey an inequality relating spatial frequency tuning to disparity tuning. This inequality was violated by most of the cells in our data set: their disparity tuning curves had more spectral power at low spatial frequencies than was possible even in the generalized energy model, given the observed response to grating stimuli.

However, the additional power at low spatial frequencies can be explained with our modified version of the energy model. Halfwave rectification prior to binocular combination removes the sidelobes in disparity tuning curves from bandpass receptive fields, shifting power to lower frequencies and explaining how a lowpass disparity tuning curve can be observed alongside bandpass spatial frequency tuning. In addition, allowing purely inhibitory input from one eye means that subunits can contribute to the observed disparity tuning without
affecting the spatial frequency tuning observed in one eye. Thus, in our modified model, the spatial frequency tuning is decoupled from the disparity tuning. This offers an explanation of why the correlation between disparity frequency and spatial frequency predicted by the energy model is not observed.

Thus, in all the areas where we have compared the energy model and our modified version, the latter has agreed more closely with the data. In addition, the modified version also explains the weaker disparity tuning observed with anti-correlated stimuli (Cumming and Parker 1997; Ohzawa et al. 1990; Livingstone and Tsao 1999). This does not constitute a proof that binocular combination is nonlinear. It is possible that all of the discrepancies noted here and elsewhere (eg. Cumming and DeAngelis (2001)) could be reconciled with linear binocular summation if sufficiently complex subsequent processing is postulated (perhaps incorporating contrast normalization and multiple binocular subunits with different output nonlinearities). Also, there may be individual neurons in which the original energy model remains a better description than our modified version. If it is true that binocular complex cells receive input from many binocular subunits, there may be a mixture of both mechanisms at work, with some binocular subunits receiving essentially linear input from both eyes and others receiving thresholded input, so explaining the continuum of observed properties.

However, we propose that the most straightforward solution is to postulate that binocular combination is not linear in most neurons. With this one modification to the energy model, the key characteristics of disparity selectivity in striate cortex can be economically reproduced. This suggests we are close to an accurate mechanistic description of how this novel property of the visual cortex is produced.

Acknowledgements:
We thank Greg DeAngelis, Lance Optican, Nicholas Port, Marc Sommer, Bob Wurtz and Huihui Zhou for helpful comments on earlier drafts. Thanks to Mark Szarowicz and Chris Hillman for their invaluable help with the animals.
Appendices

A. Fourier transforms

We denote Fourier transforms with tildes. For example, if $\rho(x,y)$ is a receptive field function, $\tilde{\rho}(\tilde{f},\tilde{\theta})$ represents its Fourier transform, where

$$\tilde{\rho}(\tilde{f},\tilde{\theta}) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \rho(x,y) \exp\left\{2\pi i \tilde{f} \left(x \cos \tilde{\theta} + y \sin \tilde{\theta}\right)\right\}.$$ 

This expresses the Fourier spectrum in polar coordinates, where the angle $\tilde{\theta}$ represents the orientation of each Fourier component relative to the $x$ axis, and $\tilde{f}$ represents the spatial frequency of each Fourier component. (We denote these with tildes to distinguish them from the orientation and spatial frequency of the receptive field, or of the stimulus, which we shall encounter later.) We shall sometimes need the Fourier spectrum in terms of the frequency components orthogonal and parallel to the $x$ axis, $f_x = \tilde{f} \cos \tilde{\theta}$ and $f_y = \tilde{f} \sin \tilde{\theta}$. We denote this $\tilde{\rho}^c(f_x, f_y)$, where the superscript C stands for Cartesian:

$$\tilde{\rho}^c(f_x, f_y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \rho(x,y) \exp\left\{2\pi i (f_x x + f_y y)\right\}.$$ 

The Fourier transform is in general a complex quantity. We shall use its absolute magnitude, $|\tilde{\rho}(\tilde{f}, \tilde{\theta})|$ and phase $\alpha = \arg \tilde{\rho}(\tilde{f}, \tilde{\theta})$. The squared magnitude, $|\tilde{\rho}(\tilde{f}, \tilde{\theta})|^2$, gives the Fourier power spectrum.

B. Disparity tuning curves

We fit experimental disparity tuning curves with 1d Gabors:

$$D(\delta) = \frac{A}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{\delta^2}{2\sigma^2}\right\} \cos(2\pi f \delta - \phi)$$

Equation 8

where $\delta$ is disparity, $D(\delta)$ is the disparity-modulated component of the tuning curve (i.e. after the subtraction of the baseline), $f$ is the spatial frequency $f$ and $\phi$ the phase of the carrier.
cosine, $\sigma$ is the standard deviation and $\delta$ the centre of the Gaussian envelope, and $A$ is the amplitude. The Fourier power spectrum of this 1d Gabor is:

$$|\tilde{D}(\tilde{f})|^2 = \frac{A^2}{2} \exp\left(-4\pi^2 \sigma^2 \left[2\cosh(2\pi^2 \sigma^2 \tilde{f}) + \cos 2\phi\right]\right).$$

**Equation 9**

For narrow-band disparity tuning curves, $\sigma^2 f^2 >> 1$, this reduces to

$$|\tilde{D}(\tilde{f})|^2 \approx \frac{A^2}{4} \exp\left(-4\pi^2 \sigma^2 \left[2\tilde{f} - |f|\right]^2\right)$$

**Equation 10**

so the disparity peak frequency coincides with the Gabor carrier frequency $f$. For Gabors which depart sufficiently from the narrow-band approximation, the disparity peak frequency may differ from the Gabor carrier frequency, since the second term in Equation 9, the Gaussian multiplied by $\cos(2\phi)$, becomes non-negligible. If $\cos(2\phi)>0$ (closer to even symmetry than odd symmetry), the disparity peak frequency is less than the carrier frequency; if $\cos(2\phi)<0$ (closer to odd symmetry), it is greater.

**C. Monocular spatial frequency tuning for a half-squared-linear binocular subunit**

The binocular energy model (Ohzawa et al. 1990) is based on binocular subunits characterized by a receptive field in each eye: $\rho_L(x,y), \rho_R(x,y)$. The response of the binocular subunit to a particular stereogram is the square of the sum of the inner product of the image in each eye with that eye’s receptive field:

$$C = [\text{Pos}(v_L + v_R)]^2,$$

**Equation 11**

where the symbol $v$ stands for the inner product (sometimes loosely called the convolution):

$$v_L = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy I_L(x,y) \rho_L(x,y).$$

**Equation 12**
where \( I_L(x,y) \) is the image presented to the left eye, expressed relative to the mean luminance of the screen (so positive values of \( I_L \) represent bright features, and negative values dark features).

We consider how such a subunit responds when presented with a monocular sinusoidal grating of spatial frequency \( f_g \) oriented at an angle \( \theta_g \) away from the optimal orientation, drifting with a temporal angular frequency \( \omega_g \). We define the retinal coordinate system such that the \( x \) axis is orthogonal to the optimal orientation of the subunit. Then, the grating stimulus is

\[
I(x,y) = I_g \cos(2\pi f_g x \cos \theta_g + 2\pi f_g y \sin \theta_g + \phi_g - \omega_g t),
\]

**Equation 13**

where \( I_g \) is the maximum luminance of the grating relative to the mean luminance of the screen.

The instantaneous response of the subunit at time \( t \) is

\[
C(t) = I_g^2 |\tilde{\rho}(f_g, \theta_g)|^2 \left[ \text{Pos}(\cos(\phi_g + \alpha(f_g, \theta_g) - \omega_g t)) \right]^2,
\]

where \(|\tilde{\rho}(f_g, \theta_g)|^2\) and \( \alpha(f_g, \theta_g) \) are, respectively, the Fourier power and phase of the receptive field function in the stimulated eye at the spatial frequency and orientation of the grating, and we have chosen to express the orientation of Fourier components relative to the optimal orientation. The unit’s mean response averaged over a stimulus temporal cycle is

\[
\langle C \rangle = \frac{I_g^2}{4} |\tilde{\rho}(f_g, \theta_g)|^2.
\]

If we set \( \theta_g = 0 \), then this expression gives the mean response of the cell as a function of the spatial frequency \( f_g \) of an optimally-oriented grating. This is therefore the model prediction for the shape of our experimental spatial frequency tuning curves. Later, it will be convenient to express the Fourier power spectrum in Cartesian coordinates:

\[
\langle C \rangle = \frac{I_g^2}{4} |\tilde{\rho}^c(f_g, 0)|^2.
\]

**Equation 14**
Equation 14 represents the predicted spatial frequency tuning curve in terms of the Fourier power spectrum of the receptive field, in Cartesian coordinates. Below, we shall combine this with the predicted disparity tuning curve in order to derive a powerful constraint on the response of any generalized energy model cell.

D. Disparity tuning curve for a half-squared-linear binocular subunit

We now consider the response of the binocular subunit to binocular random-dot stereograms. The energy model is built from half-squared-linear subunits, $C = \text{Pos}(v_L + v_R)^2$, which respond only when the sum of inputs from the two eyes, $(v_L + v_R)$, is positive. However, with random-dot patterns, any image-pair is as likely to occur as its photographic negative, and so the cell responds on average half the time. This is very convenient because it means that, in considering the mean response of the cell averaged over many thousands of random-dot patterns, we can drop the half-wave rectification and treat the subunit as if it were simply summing and squaring its inputs, provided that we also divide by 2. Thus, in deriving the disparity tuning, we can imagine that the response of the subunit is $C = (v_L + v_R)^2/2$. When such a model is stimulated with random-dot patterns with disparity $\delta$ along the $x$ axis (i.e. orthogonal to the preferred orientation), the disparity-modulated component of its response is given by

$$D(\delta) = \langle v_L v_R \rangle(\delta) = \int_{-\infty}^{\infty} dxdy.\rho_L(x,y)\int_{-\infty}^{\infty} dx' dy'.\rho_R(x',y')I(x,y)I(x' + \delta, y'),$$

Equation 15

where the angle-brackets represent averaging over the ensemble of all possible random images. Consider the analytically-tractable case where the images are white noise: that is, each pixel is independently colored either black or white with equal probability, so that the product of the luminance of pixels at different positions averages to zero. This is not the same as the random-dot patterns used in our stimuli, but simulations suggest that it gives very similar answers when averaged over a large number of patterns. For white-noise stimuli, we may approximate the term inside angle brackets in Equation 15 by a Dirac delta function:
\[
\langle I(x, y)I(x' + \delta, y') \rangle = I_{RD}^2 \Delta^2 \delta_{\text{Dirac}}(x' + x - x') \delta_{\text{Dirac}}(y' - y)
\]

**Equation 16**

where \( I_{RD} \) is the luminance of a white pixel (and \(-I_{RD} \) of a black pixel) relative to the gray level of the screen, and \( \Delta^2 \) is the area of a pixel. Then Equation 15 becomes

\[
D(\delta) = I_{RD}^2 \Delta^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dxdy \rho_L(x, y) \rho_R(x - \delta, y).
\]

By standard techniques of Fourier analysis, the Fourier spectrum of the disparity-modulated component of the cell’s response, \( \tilde{D}(f_x) \), is

\[
\tilde{D}(f_x) = I_{RD}^2 \Delta^2 \int_{-\infty}^{+\infty} df_y \tilde{\rho}_L(f_x, f_y) \tilde{\rho}_R(f_x, f_y)
\]

**Equation 17**

where \( f_x, f_y \) are the frequencies orthogonal and parallel to the preferred orientation, \( \tilde{\rho}_C(f_x, f_y) \) is the Fourier transform of the receptive field expressed in Cartesian axes in which \( x \) is orthogonal to the preferred orientation, the suffixes \( L, R \) denote left and right eyes, and \(*\) denotes complex conjugation.

The response of a complex cell is modeled as the sum of \( n \) of the binocular subunits of Equation 11:

\[
C = \sum_{j=1}^{n} \left[ Pos(v_{Lj} + v_{Rj}) \right]^2
\]

**Equation 18**

Since Fourier transforms are linear, the Fourier spectrum of the disparity tuning curve is just the sum of \( n \) terms as in Equation 17. To find the Fourier power spectrum for \( n \) subunits, we multiply this sum by its complex conjugate to obtain:

\[
\left| \tilde{D}(f_x) \right|^2 = I_{RD}^4 \Delta^4 \sum_{j,k=1}^{n} \int_{-\infty}^{+\infty} df_y \int_{-\infty}^{+\infty} df_y' \tilde{\rho}^C_{Lj}(f_x, f_y) \tilde{\rho}^C_{Rj}(f_x, f_y') \tilde{\rho}^C_{Lk}(f_x, f_y) \tilde{\rho}^C_{Rk}(f_x, f_y') \times \cos(\Delta \alpha_j(f_x, f_y))^2 - \Delta \alpha_k(f_x, f_y')
\]

**Equation 19**
where $\Delta \alpha_j(f_x,f_y)$ is the difference between the Fourier phases of the left and right receptive fields in the $j^{th}$ subunit, at the frequency component indicated. In general $\Delta \alpha_j(f_x,f_y)$ depends on frequency, but if the receptive fields are narrow-band Gaborbs, it is constant and equal to $\Delta \phi = \phi_l - \phi_r$, the phase disparity between the Gabor receptive fields.

It will be convenient to normalize the Fourier power spectrum by the response to monocular random images. The mean response to white noise in the left eye is

$$L = \frac{I_{RD}}{2} \Delta^2 \sum_{j=1}^{n} \int_{-\infty}^{\infty} dx dy \left[ \rho_{Lj}(x,y) \right]^2 = \frac{I_{RD}}{2} \Delta^2 \sum_{j=1}^{n} \int_{-\infty}^{\infty} df_x df_y \left| \tilde{\rho}_{Lj}^C(f_x,f_y) \right|^2,$$

Equation 20

and similarly for the right eye, where the last step has used Parseval’s theorem to replace the square-integral of the receptive field function with the integral of its Fourier power spectrum. Then, from Equation 19 and Equation 20, we obtain

$$\frac{|D(f_x)|^2}{4LR} = \frac{1}{4LR} \left[ \frac{I_{RD}}{2} \Delta^2 \sum_{j=1}^{n} \int_{-\infty}^{\infty} df_x df_y \left| \tilde{\rho}_{Lj}^C(f_x,f_y) \right|^2 \right] \sum_{k=1}^{n} \int_{-\infty}^{\infty} df_x df_y \left| \tilde{\rho}_{Rk}^C(f_x,f_y) \right|^2 \sum_{j=1}^{n} \int_{-\infty}^{\infty} df_x df_y \left| \tilde{\rho}_{Lj}^C(f_x,f_y) \right|^2 \sum_{k=1}^{n} \int_{-\infty}^{\infty} df_x df_y \left| \tilde{\rho}_{Rk}^C(f_x,f_y) \right|^2$$

Equation 21

As we now show, this normalized Fourier spectrum can be compared with normalized spatial frequency tuning curves to test the generalized version of the energy model.

E. **Inequality relating monocular tuning curves to Fourier power spectrum of disparity tuning curve**

We consider the model of Equation 18, $C = \sum_{j=1}^{n} \left[ \text{Pos}(v_{Lj} + v_{Rj}) \right]^2$. We restrict ourselves to the case where all receptive fields (for all subunits, in both eyes) are Cartesian-separable in the same coordinate frame, so that every receptive field $\rho(x,y)$ can be written as the product of a function of $x$ only and a function of $y$ only. The $x$-function is arbitrary, but we require the $y$-
function to be the same for all receptive fields, and to have zero Fourier phase. One example is if all the receptive fields are Gabors with the same orientation and the same extent parallel to the preferred orientation (but note that Gabor receptive fields are not required). Orientation tuning is assumed to be the same in all subunits, but spatial frequency tuning is allowed to differ both across eyes in a single binocular subunit, and across subunits. Even though we have referred so far only to binocular subunits, this framework also includes monocular subunits as a special case: we simply set the \( x \)-function in one eye to zero, so that the subunit contributes a term \([\text{Pos}(v_L)]^2\) or \([\text{Pos}(v_R)]^2\). Non-disparity-selective subunits can be represented by the sum of two monocular terms: \([\text{Pos}(v_L)]^2 + [\text{Pos}(v_R)]^2\). Thus, this formulation is extremely general.

Because we have assumed separability, the Fourier transform of each receptive field can be written, in Cartesian form, as \( \hat{\rho}_L^C(f_x, f_y) = \hat{\rho}_L^C(f_x)\hat{\rho}_L^C(f_y) \) etc, where the subscript \( L_j \) indicates that this is the left-eye receptive field of the \( j \)th subunit. Using the results of Appendix C (Equation 14), the monocular spatial frequency tuning curves obtained with optimally-oriented drifting gratings in the left and right eyes respectively are

\[
L_{\text{SF}}(f) = \frac{I_g^2}{4} \left| \hat{\rho}_L^C(0) \right|^2 \sum_{j=1}^{n} \left| \hat{\rho}_{Lj}^C(f) \right|^2, \\
R_{\text{SF}}(f) = \frac{I_g^2}{4} \left| \hat{\rho}_R^C(0) \right|^2 \sum_{j=1}^{n} \left| \hat{\rho}_{Rj}^C(f) \right|^2,
\]

Equation 22

where \( I_g \) is the luminance modulation amplitude of the grating and \( f \) its spatial frequency. We remove the dependence on grating luminance by normalizing both tuning curves, dividing by the area under the spatial frequency tuning curve:

\[
L_{\text{ASF}} = \frac{I_g^2}{4} \left| \hat{\rho}_L^C(0) \right|^2 \sum_{j=1}^{n} \int_{-\infty}^{\infty} df \left| \hat{\rho}_{Lj}^C(f) \right|^2, \\
R_{\text{ASF}} = \frac{I_g^2}{4} \left| \hat{\rho}_R^C(0) \right|^2 \sum_{j=1}^{n} \int_{-\infty}^{\infty} df \left| \hat{\rho}_{Rj}^C(f) \right|^2.
\]

Equation 23

Notice that this integration is over all frequencies, both positive and negative, i.e. for gratings drifting in both directions. Our experimental tuning curves are expressed as a function of
positive frequencies only; the estimate of $L_{ASF}$ is therefore twice the area under an experimental tuning curve.

The product of the normalized tuning curves is

$$
\frac{L_{SF}(f)R_{SF}(f)}{L_{ASF}R_{ASF}} = \frac{\sum_{j=1}^{n} \left| \tilde{\rho}_{ij}^{+}(f) \right|^2 \sum_{k=1}^{n} \left| \tilde{\rho}_{Rk}^{-}(f) \right|^2}{\sum_{j=1}^{n} \int_{-\infty}^{\infty} df \left| \tilde{\rho}_{ij}^{+}(f) \right|^2 \sum_{k=1}^{n} \int_{-\infty}^{\infty} df' \left| \tilde{\rho}_{Rk}^{-}(f') \right|^2}
$$

Equation 24

We now use the simplifying assumptions of separability in the rather unwieldy expression (Equation 21) derived in Appendix D for the Fourier power spectrum of the disparity-modulated component of the cell’s response to binocular white noise, normalized by the monocular response. The integrals over $y$ cancel out top and bottom, while the Fourier phase depends on the $x$-function only, since we assumed that the $y$-function has zero Fourier phase (e.g. is a Gaussian). We then obtain

$$
\left| \tilde{\mathcal{P}}(f) \right|^2 = \frac{\sum_{j=1}^{n} \sum_{k=1}^{n} \left| \tilde{\rho}_{ij}^{+}(f) \tilde{\rho}_{i2k}^{+}(f) \tilde{\rho}_{i2k}^{+}(f) \tilde{\rho}_{Rk}^{-}(f) \cos(\Delta \alpha_j(f) - \Delta \alpha_k(f)) \right|}{4LR \sum_{j=1}^{n} \int_{-\infty}^{\infty} df \left| \tilde{\rho}_{ij}^{+}(f) \right|^2 \sum_{k=1}^{n} \int_{-\infty}^{\infty} df' \left| \tilde{\rho}_{Rk}^{-}(f') \right|^2}
$$

Equation 25

Now consider the difference between the normalized Fourier power spectrum, Equation 25, and the product of the normalized left and right monocular tuning surfaces, Equation 24:

$$
\left\{ \frac{L_{SF}(f)R_{SF}(f)}{L_{ASF}R_{ASF}} - \frac{\left| \tilde{\mathcal{P}}(f) \right|^2}{4LR} \right\} \sum_{j=1}^{n} \int_{-\infty}^{\infty} df \left| \tilde{\rho}_{ij}^{+}(f) \right|^2 \sum_{k=1}^{n} \int_{-\infty}^{\infty} df' \left| \tilde{\rho}_{Rk}^{-}(f') \right|^2
$$

$$
= \sum_{j=1}^{n} \sum_{k=1}^{n} \left\{ \left| \tilde{\rho}_{ij}^{+}(f) \right|^2 \left| \tilde{\rho}_{i2k}^{-}(f) \right|^2 - \left| \tilde{\rho}_{ij}^{+}(f) \tilde{\rho}_{i2k}^{+}(f) \tilde{\rho}_{i2k}^{+}(f) \tilde{\rho}_{Rk}^{-}(f) \cos(\Delta \alpha_j(f) - \Delta \alpha_k(f)) \right| + 2 \left| \tilde{\rho}_{ij}^{+}(f) \tilde{\rho}_{i2k}^{+}(f) \tilde{\rho}_{i2k}^{+}(f) \tilde{\rho}_{Rk}^{-}(f) \sin^2 \left( \frac{\Delta \alpha_j(f) - \Delta \alpha_k(f)}{2} \right) \right| \right\}
$$

Equation 26

where we have used a trigonometric identity to replace the cosine in Equation 25 with the sine of a half-angle.
The right-hand side of Equation 26 is clearly unaffected by interchanging the summation indices \( j \) and \( k \). We write it out twice, interchanging the indices the second time. It then becomes apparent that the terms other than the sine terms form a perfect square, so we can write

\[
2 \left[ \frac{L_{SF}(f)R_{SF}(f)}{L_{ASS}R_{ASS}} - \frac{\bar{P}(f)^2}{4LR} \right] \sum_{j=1}^{n} \int_{-\infty}^{\infty} df \left| \tilde{\rho}_{Lj}^{\alpha}(f) \right|^2 \sum_{k=1}^{n} \int_{-\infty}^{\infty} df \left| \tilde{\rho}_{Rk}^{\alpha}(f') \right|^2 \\
= \sum_{j=1}^{n} \sum_{k=1}^{n} \left\{ \left| \tilde{\rho}_{Lj}^{\alpha}(f) \right|^2 \left| \tilde{\rho}_{Rk}^{\alpha}(f) \right|^2 \right\} + 4 \sum_{j=1}^{n} \sum_{k=1}^{n} \left| \tilde{\rho}_{Lj}^{\alpha}(f) \right|^2 \left| \tilde{\rho}_{Rk}^{\alpha}(f') \right|^2 \sin^2 \left( \frac{\Delta \alpha_j - \Delta \alpha_k}{2} \right)
\]

It is now clear that every term in the sum on the right-hand side of this equation is the square of a real quantity, and therefore non-negative. The sum itself must therefore yield a non-negative number. The sums on the left-hand side must yield a non-negative number for the same reason. We conclude that, for every spatial frequency \( f \),

\[
\frac{L_{SF}(f)R_{SF}(f)}{L_{ASS}R_{ASS}} \geq \frac{\bar{P}(f)^2}{4LR}.
\]

**Equation 27**

Equality holds when the receptive fields in each eye have the same Fourier amplitude spectrum for all the subunits, and when the Fourier phase disparity between left and right eyes is the same for all subunits. This is the case, for example, in the original form of the energy model (Ohzawa et al. 1990); even-symmetric disparity tuning curves were obtained by setting the phase disparity within every subunit to be zero, and odd-symmetric by setting it to \( \pi/2 \).

One practical problem with Equation 27 is that it does not take account of any gain control mechanism which boosts responses to monocular stimulation relative to binocular stimulation. On expanding the squared term in Equation 18, it is apparent that the model’s mean response to binocularly uncorrelated random-dot stimuli is given by the sum of its mean responses to monocular random-dot stimuli:

\[
U = L + R
\]

**Equation 28**
Prince et al. (2002a) tested this prediction of the energy model and found that, in fact, the uncorrelated stimuli were lower than predicted by the model. Instead of being equal to the sum of the left and right responses, as in Equation 28, the uncorrelated response was close to their average, \( U_{\text{obs}} = (L_{\text{obs}} + R_{\text{obs}})/2 \) (the subscript obs denoting “observed”). A similar result is found in our present data, and can be explained by our modified version of the energy model. However, it can also be reconciled with the original energy model if we postulate a form of contrast normalization which tends to boost the response to monocular stimulation. This could be a divisive normalization in which the unnormalized response of each cell is divided by the total response from a pool of neighboring cells. We can model this simply by dividing the output of the model in Equation 18 by the total contrast of the images presented to left and right eyes. This halves the uncorrelated response relative to the monocular responses, leading to the correct relationship \( U_{\text{obs}} = (L_{\text{obs}} + R_{\text{obs}})/2 \). Thus, combining the energy model with a divisive normalization mechanism can explain this feature of the data.

However, this means that we cannot use the observed response to monocular random dot patterns in Equation 27, since the quantities L and R in Equation 27 need to be the responses to left and right monocular random-dot stimulation before any contrast normalization, and we do not have access to this experimentally. We therefore recast Equation 27 using the response to binocularly-uncorrelated stimuli, \( U \). The energy model (before response normalization) states that \( U = L + R \). It then follows from \( (L-R)^2 \geq 0 \) that \( U^2 \geq 4LR \). Thus,

\[
\frac{L_{sf}(f)R_{sf}(f)}{L_{ASF}R_{ASF}} \geq \frac{\hat{d}(f)^2}{U^2}.
\]

**Equation 29**

The right-hand side involves the ratio of responses to two binocular stimuli, and is thus unaffected by a normalizing mechanism which boosts monocular responses. In fact we also examined the inequality of Equation 27, which gave similar results.

In summary, for a significantly generalized version of the energy model, we have derived a useful upper limit on the normalized power spectrum of the disparity tuning curve. The normalized disparity power at any frequency cannot exceed the normalized response to
monocular drifting gratings at that frequency. This holds for a model of the same form as that proposed by Ohzawa et al. (1990), but generalized to allow any number of subunits which may differ in their spatial frequency tuning, in the phase of their receptive fields, in the position of their receptive fields on the retina, and in the position and phase disparity between left and right receptive fields. The only property which is not allowed to vary between subunits is the orientation of the receptive fields, and their extent along this axis. The upper limit on disparity power holds even in the presence of response normalization mechanisms which scale the responses to grating stimuli relative to the responses to random-dot patterns, or which scale the responses to monocular stimuli relative to the responses to binocular stimuli.

F. Effect of stimulus misalignment

Here we examine the consequences of using a stimulus orientation that does not exactly match the orientation of the receptive field under study. The results apply to the model of Equation 18, \( C = \sum_{j=1}^{g} \left[ Pos(v_{Lj} + v_{Rj}) \right]^2 \), restricted to the case where all receptive fields are narrow-band Gabor functions with identical spatial frequency and orientation tuning (i.e., differing only in phase and position on the retina), and where the position and phase disparity between left and right receptive fields is the same for all subunits. This is the form of the energy model used by all previous workers that we are aware of (e.g. Ohzawa et al. (1990; 1997), Fleet et al. (1996), Qian and Zhu (1997)). Since, for Gabors with a spatial frequency bandwidth of less than \( \sim 2 \) octaves, the Fourier power spectrum is independent of position or phase, Equation 19 of Appendix D simplifies to

\[
\tilde{D}(f_x) = \int_{-\infty}^{+\infty} df_y \left| \tilde{\rho}^C (f_x, f_y) \right|^2
\]

Equation 30

while the left and right monocular spatial frequency tuning curves are identical and equal to

\[
L_{SF}(f_x) = R_{SF}(f_x) = \frac{I}{4} \left| \tilde{\rho}^C (f_x, 0) \right|^2.
\]
As before, the equations are for disparities along the $x$ axis, and for the spatial frequency tuning curves obtained with gratings oriented parallel to the $y$ axis. However, we now allow for the possibility that the $y$ axis is not aligned along the true preferred orientation of the cell – for instance, because of experimental error in assessing this true orientation. The power spectrum $|\hat{\rho}(f_x, f_y)|^2$ of this cell is shown in Figure 15. The spatial frequency tuning curve (Equation 27) is simply a slice through this surface, along the $f_x$ axis. The Fourier spectrum of the disparity tuning curve, on the other hand, is given by the line integral of this surface along lines parallel to the $y$ axis (Equation 30). Referring to Figure 15, and using the symmetry of the Gabor power spectrum, it is apparent that the peak of the disparity power spectrum will occur at the horizontal frequency $f_0 \cos \theta_0$, where $f_0$, $\theta_0$ are the spatial frequency and orientation of the Gabor. In contrast, it can be shown that the peak of the monocular spatial frequency tuning curve will occur at

$$f_x = \frac{\sigma_\perp^2}{\sigma_\parallel^2 \cos^2 \theta_0 + \sigma_\perp^2 \sin^2 \theta_0} f_0 \cos \theta_0,$$

where $\sigma_\parallel$, $\sigma_\perp$ describe the extent of the Gabor respectively parallel and orthogonal to its preferred orientation. It follows that, provided $\sigma_\parallel > \sigma_\perp$, misaligning the stimuli will cause the measured monocular spatial frequency tuning curves to peak at a lower frequency than the power spectrum of the disparity tuning curve. Thus, at least for the class of model considered here, such misalignment cannot be responsible for the observation that the disparity frequency is usually lower than the preferred monocular spatial frequency.

G. Glossary of symbols used in the Appendices

- $C$: firing rate of model cell, e.g. Equation 11.
- $\delta$: disparity along an axis orthogonal to receptive field orientation.
- $D(\delta)$: disparity-modulated component of response to binocular random-dot patterns, Equation 15.
- $\tilde{D}(f)$: Fourier amplitude spectrum of the disparity-modulated component of the disparity tuning curve, Equation 17.
- $f$: spatial frequency (cycles per degree).
$f_g$ \hspace{1cm} \text{spatial frequency of grating stimulus, Equation 13.} \\
$I(x,y)$ \hspace{1cm} \text{image luminance, relative to mean, as a function of retinal position, e.g. Equation 13} \\
$I_L(x,y), I_R(x,y)$ \hspace{1cm} \text{left and right retinal images.} \\
$I_g, I_{RD}$ \hspace{1cm} \text{maximum luminance of grating / random-dot pattern relative to the mean screen luminance, cf Equation 13 and Equation 16.} \\
$j$ \hspace{1cm} \text{index used to enumerate subunits feeding into a cell, Equation 18.} \\
$L, R$ \hspace{1cm} \text{mean response to monocular random-dot patterns in left, right eye, Equation 20.} \\
$L_{SF}(f), R_{SF}(f)$ \hspace{1cm} \text{monocular spatial frequency tuning curve, i.e. response as a function of frequency to drifting gratings in left, right eyes; Equation 22.} \\
$L_{ASF}, R_{ASF}$ \hspace{1cm} \text{area under monocular spatial frequency tuning curves, Equation 23.} \\
\text{Pos} \hspace{1cm} \text{halfwave rectification: Pos}(x) \equiv x \text{ if } x > 0, = 0 \text{ otherwise.} \\
\rho(x,y)$ \hspace{1cm} \text{receptive field function.} \\
$\tilde{\rho}(\tilde{f}, \tilde{\theta})$ \hspace{1cm} \text{Fourier transform of receptive field (polar coordinates; } \tilde{f}, \tilde{\theta} \text{ are the spatial frequency and orientation of each Fourier component).} \\
$\tilde{\rho}^c(\tilde{f}_x, \tilde{f}_y)$ \hspace{1cm} \text{Fourier transform of receptive field (Cartesian coordinates; } \tilde{f}_x, \tilde{f}_y \text{ are the components of spatial frequency orthogonal and parallel to the preferred orientation).} \\
$\tilde{\rho}^s(\tilde{f}_x), \tilde{\rho}^s(\tilde{f}_y)$ \hspace{1cm} \text{Fourier transform of one-dimensional receptive field profiles along axes orthogonal and parallel to the preferred orientation.} \\
$\theta_g$ \hspace{1cm} \text{orientation of grating stimulus relative to the preferred orientation of the cell, Equation 13.} \\
$U$ \hspace{1cm} \text{mean response to binocularly uncorrelated random-dot patterns, Equation 28} \\
$v_L, v_R$ \hspace{1cm} \text{inner product (convolution) of retinal image with receptive field in left eye, right eye, Equation 12} \\
$x, y$ \hspace{1cm} \text{retinal coordinates, in a frame where the } y \text{ axis is the preferred orientation of the cell.}
Figure Legends

Figure 1
Block diagrams of the energy model (A) and our modified version (B). The grayscale plots represent receptive fields, which are shown differing in phase. The arrows show the results $v_L, v_R$ of a linear operation performed on the image in each eye. The models do not specify the physiological details of how this linear operation is calculated, so they are shown simply with arrows. Subsequently, triangles represent excitatory synapses, and disks inhibitory synapses. A: In the energy model, linear inputs from both eyes converge onto a single binocular simple cell. Each binocular simple cell computes the linear sum of its left and right inputs, and outputs the half-squared sum to the complex cell. B shows a possible implementation of our modified version: linear inputs from left and right eyes pass through monocular simple cells, and are thus half-wave rectified, before converging on a binocular simple cell. After this rectification, the type of synapse (excitatory/inhibitory) at the binocular simple cell has a profound influence on the type of disparity tuning observed. In B, the upper binocular subunit is shown receiving excitatory synaptic input from both eyes; the lower subunit is shown with one excitatory and one inhibitory synapse.

Figure 2
Random-dot stereogram of the type used in our experiments. The dots are $0.1\times0.1$° square. The circle indicates the size of a typical V1 receptive field for comparison. The diameter is 1.2°, based on a subset of cells in which we obtained the one-dimensional receptive field envelope using a long luminance-modulating bar stimulus at the preferred orientation. The mean SD of the Gaussians fitted to the receptive field envelopes was 0.3°, and we took four SDs as a suitable estimate of receptive field diameter.

Figure 3
Two tuned-inhibitory cells which show evidence of an inhibitory input from one eye. Stimulation in the non-dominant eye seems always to reduce the firing rate: the response to monocular random dots in the non-dominant eye is less than that to a blank screen, while the response to binocularly uncorrelated random dots is less than that to monocular random dots in the dominant eye. Disparity discrimination index = 0.61 for duf096 (A), 0.64 for duf099 (B).
Black dots● represent the mean firing rate as a function of disparity; the black curve is the fitted Gabor. Horizontal lines represent responses to stimuli without disparity: the line labeled L and marked with a leftward arrowhead ◀ indicates the response to monocular random dots in the left eye (R▷: right eye); U
indicates the response to binocularly uncorrelated random dots; SO = spontaneous rate (response to a gray screen of the same mean luminance). All error-bars are ±SEM.

**Figure 4**

Our new model, incorporating a threshold linearity prior to binocular combination, can explain disparity selectivity in classically monocular cells, and cells in which the response to binocularly uncorrelated dots (U) is close to the mean of the response to monocular stimuli (L<,R>) rather than to its sum as predicted by the energy model. Both plots show simulations for a single binocular subunit receiving inhibitory input from the right eye (Equation 5); mean response over 100,000 different random-dot patterns. A: the monocular receptive fields are identical, so the inhibitory synapse in Equation 5 results in a tuned-inhibitory cell. B: the left eye’s receptive field is the inverse of the right eye’s receptive field, so with the inhibitory synapse this results in a tuned-excitatory cell.

**Figure 5**

Frequency histogram for the ratio of the response to random-dot stimulation presented monocularly to the dominant eye, M, to the response to binocular uncorrelated random-dot patterns presented binocularly, U. The dashed vertical line marks M/U=1, the upper bound predicted by the energy model; the solid vertical line marks M/U=2, the upper bound predicted by our modified version. The shading indicates cells for which we can be 95% confident that the ratio exceeds the upper bound: gray shows cells where the 2.5% percentile exceeded 1, black where it exceeded 2. Thus, the gray+black regions indicate the 44/138 cells which significantly violate the upper bound predicted by the energy model; the black regions indicate the 4/138 cells which significantly violate the upper bound predicted by our model.

**Figure 6**

Scatter plot of preferred spatial frequency in left eye against that in right, on log axes. The preferred frequency is defined by the peak of the Gaussian fitted to the monocular tuning curve obtained with sinusoidal luminance gratings. The solid line shows the identity; the dotted lines mark difference in SF tuning of 1 octave. Filled symbols indicate the 25 cells whose preferred frequencies in left and right eyes differed significantly (p<0.05, resampling). Circles indicate cells from monkey Duf, squares from monkey Ruf. The cells shown in Figure 7 are indicated. Two cells which had very low preferred spatial frequencies fall outside the range shown in the figure.

**Figure 7**
Two example cells (A: duf156; B: ruf127) which respond well to grating stimulation in either eye, but show a particularly extreme difference between the spatial frequencies evoking the maximum response. The triangles show the mean firing rate to monocular drifting luminance gratings, as a function of spatial frequency. Error bars are ±SEM. The curves are the Gaussian fitted to this data; the peaks are indicated with vertical lines. The shaded regions show the 95% confidence interval for this peak (estimated by refitting resampled data-sets). Left eye: leftward empty arrowheads <, dotted line; Right eye: rightward filled arrowheads ➤, solid line.

Figure 8

Two disparity-tuned cells (A, C: duf092; B, D: ruf065) which seem to show an inhibitory influence from one eye.
A, B: spatial frequency tuning (symbols as in Figure 7)
C, D: disparity tuning (symbols as in Figure 3). For ruf065.0, the disparity-tuning curve is not well fitted by a Gabor (fit explained < 60% of variance); the fit is therefore not shown. Disparity discrimination index = 0.58 for duf092 (C), 0.52 for ruf065 (D). For both cells, the maximum binocular response is less than the response to monocular stimulation in the dominant eye. Thus, adding dots to the non-dominant eye, at any disparity, always reduces the response.

Figure 9

Comparing disparity and spatial frequency tuning. Left: disparity tuning curves (symbols as in Figure 3). For ruf072 (A), the dashed curve shows the unrectified Gabor, which dips below zero. Right: Fourier transform of the disparity-modulated component of the disparity tuning curve (FT-DMC; gray) compared with the spatial frequency tuning curve in the dominant eye (SFTC; black), scaled to the same peak value. Black dots: mean firing rate as a function of spatial frequency; black curve: log-Gaussian fitted to this data. Dotted gray line (“FT-DMC (data)”: Fourier amplitude spectrum of raw disparity tuning curve minus mean response to uncorrelated stimuli. Dashed gray line (“FT-DMC (fit)”: Fourier amplitude spectrum of fitted Gabor minus fitted baseline.
A: ruf072: FT-DMC resembles the SFTC.
B: duf065: FT-DMC appears shifted towards lower frequencies than the SFTC.
C: duf067: FT-DMC is lowpass, SFTC is bandpass.
Figure 10

These four panels compare the Fourier spectrum of the disparity tuning curve (FT-DMC), after subtraction of the fitted baseline, with the grating spatial frequency tuning in the dominant eye. In each case, the identity line is marked, and the filled symbols show cells where the properties of the disparity tuning curve differed significantly from those obtained with gratings (p<5%, resampling). Circles indicate cells from monkey Duf, squares from Ruf. All quantities were estimated from fits to data (99 cells).

A: Frequency at which the Fourier spectrum peaks.

B,C. Low cut and high cut, i.e. the frequencies on either side of the peak at which the fitted spectrum falls to half its maximum value. In panels A-C, one cell falls outside the range of the axes; in each case its FT-DMC values were higher than the SFTC values, but this difference was not significant.

D: Relative power at lowest frequency tested monocularly (i.e., power at this lowest frequency, normalized by power at the peak of the frequency spectrum).

Figure 11

The effect of vergence eye movements, or several subunits with different position disparities. A: a receptive field profile: a Gabor function yielding a spatial frequency bandwidth of 1.5 octaves. B: The thin line shows the autocorrelogram of this receptive field, which is the disparity tuning curve predicted by the energy model (Equation 11). The heavy line shows the disparity tuning curve which would be obtained by combining the output of many such subunits, as in Equation 18, each with the receptive field profile shown but differing in position disparity. To obtain the curve shown here, 1000 subunits were used, with random position disparity drawn from a Gaussian with mean zero and standard deviation equal to half the spatial period of the Gabor carrier. This simulates random jitter in the monkey’s vergence. C: compares the Fourier power spectrum for a single subunit (thin line) and for the 1000 subunits (heavy line). The dotted line shows the 1000 subunit result scaled up to the same peak. For the single subunit, the disparity power spectrum peaks near the frequency of the Gabor carrier. Adding more subunits with a scatter in position disparity has removed power at high frequencies, shifting the peak to lower frequency.

Figure 12

Comparison of spatial frequency and disparity tuning for three example cells.

Left column (ADG): monocular spatial frequency tuning curves (Left eye: leftward empty arrowheads ◄, dotted line; Right eye: rightward filled arrowheads ►, solid line.).

Middle column (BEH): Black dots● show disparity tuning curve. Curve shows the fitted Gabor; shaded region shows the 95% confidence interval for fitted functions. Broken horizontal lines show the response to
uncorrelated stimuli (upper line, □) and to a blank screen (lower line, ○). In each case, symbols and error bars show mean±SEM.

Right column (CFI): The black line shows the product of the left and right spatial frequency tuning curves, normalized by the area underneath each. The gray line shows the Fourier power spectrum of the disparity tuning curve minus the fitted baseline, normalized by the square of the fitted baseline. The shaded regions around the curves show the 95% confidence intervals, estimated by fitting resampled data. The position of the peak of each curve are marked with vertical lines. The generalized energy model predicts that the gray curve should always lie below the black curve (Equation 6).

Top row (ABC): ruf030: a cell which is mostly consistent with the energy model. At low frequencies the normalized power spectrum of the disparity tuning curve is not significantly greater than the product of the monocular spatial frequency tuning curves, in accordance with the generalized energy model. (At high frequencies, the inequality is in fact violated, so even this cell may not be entirely compatible with the energy model.)

Middle row (DEF): ruf139: a cell which violates the generalized energy model. At low frequencies, the normalized disparity spectrum (gray region in F) is significantly higher than the normalized product of the spatial frequency tuning curves (dark region in F), in violation of Equation 6. Note that an analysis only of the peak frequency of these two curves would not have revealed this discrepancy.

Bottom row (GHI): duf096: another cell which violates the generalized energy model. The spatial frequency tuning curves peak at high frequencies: over 10 cycles per degree. In contrast, the disparity tuning curve has most power at DC, and no power at all at 10cpd.

Figure 13

At low frequencies, over half of cells are incompatible with the energy model. The energy model predicts that ∆ (Equation 7) is non-negative at every frequency. For each of 83 cells, we evaluated ∆ at a range of frequencies. The solid curve shows the percentage of cells for which the energy model hypothesis that ∆ is non-negative could be rejected at the 5% level (dotted line). For low frequencies, the energy model can be rejected for around half the cells. We used the Gaussian/Gabor functions fitted to the data. The spatial frequency tuning curves were assumed to be zero outside the range of frequencies tested (even though the fitted curve does not necessarily fall to zero, since a baseline was included in the Gaussian fit). In order to make a meaningful definition of “low frequencies” across a population with differing spatial frequency tuning, we evaluated ∆ at 32 frequencies, of which the lowest was always 0.01cpd, while the 16th was the peak of the product of the spatial frequency tuning curves (SFTCs) (marked “peak” in the figure). Other frequencies were scaled at equal increments in ln(frequency). The graph shows that, at frequencies below the
SFTC peak, the energy model can usually be rejected, because the disparity tuning contains more power than expected on the basis of the spatial frequency tuning (cf. Figure 9, Figure 10).

**Figure 14**

Comparison of simulated spatial frequency and disparity tuning for three model cells.
Left column (ADG): monocular spatial frequency tuning curves (Left eye: dotted line; Right eye: solid line.).
Middle column (BEH): disparity tuning curve (black curve) and response to uncorrelated stimuli (broken line, U). Note that the horizontal axis has a different scale in B, E and H.
Right column (CFI): The black line shows the product of the left and right spatial frequency tuning curves, normalized by the area underneath each. The gray line shows the Fourier power spectrum of the disparity tuning curve about the baseline, normalized by the square of the response to uncorrelated stimuli.
The receptive fields are two-dimensional Gabor functions, tuned to 2.5cpd (left eye) and 3.5cpd (right eye).
Top row (ABC): energy model binocular subunit, \( C = [\text{Pos}(v_L + v_R)]^2 \).
Middle row (DEF): same binocular subunit modified to include a high threshold prior to binocular combination: \( C = [\Theta(v_L) + \Theta(v_R)]^2 \) where \( \Theta \) represents a threshold function: \( \Theta(x) = (x-\theta) \) if \( x \) exceeds the threshold \( \theta \), and 0 otherwise. The value of the threshold \( \theta \) was set so high that only 5% of monocular random-dot patterns evoked a response.
Bottom row (GHI): sum of two binocular subunits including the same threshold nonlinearity as above, plus an inhibitory synapse: \( C = [\text{Pos}(\Theta(v_L) - \Theta(v_R))]^2 + [\text{Pos}(\Theta(v_R) - \Theta(v_L))]^2 \). A pair of complementary subunits is chosen in order to give a monocular response in both eyes (clearly, a single inhibitory subunit would violate the inequality even more dramatically, since the product of the spatial frequency tuning curves would be zero).
In each case, the disparity tuning curve represents the mean response to 100,000 different random-dot patterns.

**Figure 15**

Contour plot of Fourier power spectrum for a Gabor function oriented at 18° to the vertical, tuned to frequency 0.2cpd, spatial frequency bandwidth 1.5 octaves and orientation bandwidth 30° (both full-width at half-maximum of power spectrum).
Figures

A  energy model  
(linear binocular combination)

B  modified version  
(thresholds before binocular combination)

Figure 1

Figure 2
Figure 3

Figure 4
Figure 5

Figure 6
Figure 7

Figure 8
Figure 9
Figure 10

Figure 11
Figure 12
Figure 13

spatial frequency tuning curves (SFTCs)

left eye
right eye

disparity tuning curve (DTC)

normalized FT-DMC and product of SFTCs

---

63
Figure 14

FT-DMC peaks here

SFTC peaks here

SFTC is a slice along this axis

Figure 15
References


