

Suppressive Mechanisms in Monkey V1 Help to Solve the Stereo Correspondence Problem

Seiji Tanabe, Ralf M. Haefner, and Bruce G. Cumming

Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892

Neurons encode the depth in stereoscopic images by combining the signals from the receptive fields in the two eyes. Local variations in single images can activate neurons that do not signal the correct disparity (false matches), giving rise to the stereo correspondence problem. We used binocular white-noise stimuli to decompose the responses of monkey primary visual cortex V1 neurons into the elements of a linear–nonlinear model (via spike-triggered covariance analysis). In our population of disparity-selective neurons, we find both excitatory and suppressive elements in many of the neurons. Their binocular receptive fields were aligned in a specific push–pull manner for disparity. We demonstrate that this arrangement reduces the responses to false matches but preserves the responses to true matches. The responses of the cells to the noise stimuli were well explained by a linear summation of the elements, followed by a nonlinearity. This model also explained the shape of independently measured disparity-tuning curves, although it overestimated the response magnitude. This study constitutes the first direct physiological evidence for the contribution of suppressive mechanisms to disparity selectivity. This new mechanism contributes to solving the stereo correspondence problem.

Introduction

The different geometric positions of left and right eyes lead to small differences between the images on the two retinæ (binocular disparities). These disparities are used by the visual system to infer depth. A central challenge consists in identifying corresponding features in the two eyes, called the stereo correspondence problem (Julesz, 1971; Marr and Poggio, 1979). Visual neurons suffer from the problem that responses to noncorresponding images (false matches) can be as large as those to correct matches (Cumming and Parker, 1997, 2000).

The disparity-energy model has been widely used to explain the disparity tuning of neurons in the primary visual cortex (V1) (Ohzawa et al., 1990; DeAngelis et al., 1991; Fleet et al., 1996; Zhu and Qian, 1996). This model passes the image through linear filters in each eye and then passes the binocular sum through an output nonlinearity. It is a member of a widely used class of linear–nonlinear (LN) models (Hunter and Korenberg, 1986; Marmarelis et al., 1986; Sakai et al., 1988; Schwartz et al., 2006). The original disparity-energy model placed strong constraints on the linear filters: there were exactly two parallel elements (a quadrature pair), both of which are excitatory. Both elements used the same rule (e.g., a simple translation) to apply a disparity between left and right eye filters: a receptive field (RF)

disparity. These two elements elegantly capture many properties of disparity-selective neurons. However, such a simple model is inevitably an approximation; understanding how real neurons deviate from this approximation has helped clarify how they compute disparity.

The energy model responds best (on average) to stimuli with a disparity that matches the RF disparity. Nonetheless (as with many other detectors), in any one image, stronger activation may be produced by a disparity that does not match (Cumming and Parker, 1997, 2000). This means that, in a neural population of various RF disparities, the neuron with the incorrect RF disparity may respond the strongest to a given image. It is therefore unclear how the correct depth can be inferred from a population of such neurons (Fleet et al., 1996). Recent work from our group has suggested that responses to false matches may be attenuated by adding elements to the original model (Read and Cumming, 2007; Haefner and Cumming, 2008; Tanabe and Cumming, 2008).

To characterize these putative elements with as few assumptions as possible, we analyzed the spiking responses of neurons in the primate V1 with a spike-triggered covariance approach (de Ruyter van Steveninck and Bialek, 1988; Touryan et al., 2002; Horwitz et al., 2005; Rust et al., 2005; Schwartz et al., 2006). We find that neuronal responses are characterized by a combination of both excitatory and suppressive elements. Furthermore, filters of the elements are arranged in a way that results in a suppression of neuronal responses to false matches. We show how the combination of excitatory and suppressive elements helps to reduce the problem of false matches in the stereo correspondence problem.

Materials and Methods

Subjects. Two male rhesus macaques (*Macaca mulatta*) were used in the experiments. We implanted a head-restraining post, scleral search coils

Received Sept. 23, 2010; revised April 5, 2011; accepted April 14, 2011.

Author contributions: S.T., R.M.H., and B.G.C. designed research; S.T. and B.G.C. performed research; S.T. and B.G.C. analyzed data; S.T., R.M.H., and B.G.C. wrote the paper.

This work was supported by the Intramural Program of the National Eye Institute/National Institutes of Health. We thank H. Nienborg for comments on this manuscript and D. Parker and B. Nagy for excellent animal care.

Correspondence should be addressed to Seiji Tanabe, 1410 Pelham Parkway South, Room 822, Bronx, NY 10461. E-mail: seiji.tanabe@gmail.com.

R. M. Haefner's present address: Volen National Center for Complex Systems, Volen 208/MS 013, Brandeis University, Waltham, MA 02454.

DOI:10.1523/JNEUROSCI.5000-10.2011

Copyright © 2011 the authors 0270-6474/11/318295-11\$15.00/0

in both eyes, and a recording chamber over the operculum of V1. Surgical procedures were done under general anesthesia and sterile conditions. All protocols were approved by the Institutional Animal Care and Use Committee and complied with Public Health Service policy on the humane care and use of laboratory animals.

The subjects perched on a primate chair, with their head fixed. They viewed a separate CRT monitor (Eizo Flexscan F980) with each eye through a haploscope. Their task was to maintain the conjugate eye position within a box of $0.8 \times 0.8^\circ$ for 2.1 s, at the end of which a liquid reward was delivered. Stimuli were generated on a Silicon Graphics Octane workstation. Gamma correction was used to linearize the luminance response. The mean luminance was 42 cd/m^2 , contrast was 99%, and the frame rate was 96 Hz. To minimize the impact of any onset transient, the stimulus was displayed continuously between trials, and only frames occurring after fixation had been maintained for at least 100 ms were used.

Recording. A tungsten microelectrode (typically $1.0 \text{ M}\Omega$ at 1 kHz; Alpha Omega) was lowered through the dura on each recording session. Voltage signals were amplified (Bak Electronics), bandpass filtered (0.1–10 kHz), and stored to disk (Datawave Discovery System). Spike waveforms were recorded at 32 kHz sampling rate. Single-unit isolation was checked offline with custom-built software.

On isolation of a single unit, we characterized the ocular dominance, orientation tuning, direction selectivity, and spatial frequency (SF) tuning with a drifting sinusoidal grating. Using the preferred grating inside a long, thin patch, we measured the minimum response field (MRF) of the cell along the axis perpendicular and parallel to the preferred orientation, respectively (Read and Cumming, 2003). The disparity tuning was then tested with a dynamic random-line stereogram (RLS) ($4 \times 4^\circ$; line width, 0.04°) centered on the MRF. If the cell fired fewer than 10 spikes/s to all RLS or did not show significant disparity selectivity (ANOVA, $p < 0.05$), it was excluded from this study.

Stimulus. We generated our noise stimulus, $I^{(L)}(x)$ and $I^{(R)}(x)$, by summing 10 sinusoidal gratings that formed a harmonic series and one offset term. These images were calculated and presented at the resolution of the screen pixels but, because they only produce 21 independent values in each eye's image, the images were downsampled to 21 values for the analysis. We shall refer to these downsampled values as "pixels" below, i.e., x is a 21-dimensional vector (x_i with $i = 1, 2, \dots, 21$). The sampling interval was $1/21$ th of the period of the harmonic series. The amplitude $a_m^{(L)}$ of the m th harmonic of the left eye was a Bernoulli random variable, 0 or c , each with a probability of 0.5. The phase of the harmonic $\varphi_m^{(L)}$ was sampled uniformly from $0 \leq a_m^{(L)} < 2\pi$. The image in the left eye was thus

$$I^{(L)}(x_i) = c_0^{(L)} + \sum_{m=1}^{10} a_m^{(L)} \sin(2\pi m f_0 x_i + \varphi_m^{(L)}), \quad (1)$$

where $c_0^{(L)}$ is sampled from the same distribution as the sum of sinusoids on the right-hand side of Equation 1:

$$c_0^{(L)} = \frac{1}{\sqrt{20}} \sum_{m=1}^{10} b_m^{(L)} \sin(\theta_m^{(L)}). \quad (2)$$

The $b_m^{(L)}$ and $\theta_m^{(L)}$ values were sampled independently of Equation 1. The resulting 21 stimulus pixels were each approximately Gaussian distributed (by the central limit theorem) and uncorrelated with each other, as required by the reverse correlation technique we are using.

The fundamental frequency f_0 was chosen such that the series covered the SF pass band as measured with gratings. For neurons that responded to a broad range of low frequencies, this fundamental was chosen such that the highest frequency was outside the pass band. A new noise image was generated on every video frame. The patch size was equal to the period of the fundamental component f_0 . The contrast of the components, c , was fixed at 0.17. With this value, the probability that an image would saturate the monitor was 0.005. On these image frames, the value of c was lowered such that the image did not saturate the dynamic range of the monitor.

In the right eye, the amplitudes $a_m^{(R)}$ and the DC component $c_0^{(R)}$ were assigned independently of the left eye on each video frame. The phase of

the m th component in the right eye was the sum of the $\varphi_m^{(L)}$ in the left eye and a randomized interocular phase difference $\Delta\varphi_m$, thus

$$I^{(R)}(x_i) = c_0^{(R)} + \sum_{m=1}^{10} a_m^{(R)} \sin(2\pi m f_0 x_i + \varphi_m^{(L)} + \Delta\varphi_m). \quad (3)$$

$\Delta\varphi_m$ was randomly sampled from a discrete uniform distribution with equal probability at $\{0, \pi/3, 2\pi/3, \pi, 4\pi/3, 5\pi/3\}$ (for the purposes of another study).

In a subset of the cells, we also measured responses to interleaved anticorrelated and correlated RLS. A trial lasted 2.1 s. There were four periods of stimulus presentation within a single trial. Each period had duration of 420 ms, followed by a blank interval of 100 ms. A new RLS was generated every frame.

Identification of the LN model. The noise image was converted to an array of numbers. The axis of the image parallel to the stimulus orientation was ignored because the luminance was uniform. Because the actual stimulus was shown at screen resolution computed directly from the sinusoidal components, the luminance pattern along the perpendicular axis was downsampled to 21 locations for each frame in each eye (the number of independent values generated by our method) for the purpose of our analysis. The image values were the luminance differences from the background gray. A single binocular image can thus be represented as a point in a 42-dimensional space.

We triggered the noise stimulus backward in time from each spike. There was one spike-triggered ensemble (STE) of frames for each trigger delay, $\tau = \{20, 25, \dots, 95 \text{ ms}\}$. For each τ , we calculated the spike-triggered covariance (STC) matrix. We chose the τ that maximized the variance across the values in the STC matrix. The STE with this τ was then used to summarize the responses of each cell.

The average of the STE, or the spike-triggered average (STA), is the identified filter of a simple-cell-like element of the LN model. The output of this element is half-wave rectified, instead of full-wave rectified as in the other elements. We tested the significance of this element by shuffling the trials, i.e., randomly reassigning the spikes recorded in one trial to the stimuli presented in another. Once a trial of spikes was reassigned, it was not replaced in the possible pool of reassignments. We created 1000 sets of shuffled data. For each shuffle, we calculated the STA and the distance of the STA from the origin. If the distance of the original STA exceeded the 99.5 percentile of the distances of the shuffled ones, the STA was regarded significant.

The axis along the STA was projected out from all the images in the STE; that is to say that the vector component parallel to the STA was subtracted from each frame in the STE (Schwartz et al., 2006). The subtraction guaranteed that the linear filter of the simple-cell-like element was orthogonal to the linear filter of any other element of the LN model. We calculated the STC matrix of the new STE. The eigenvectors and eigenvalues of the STC matrix are the principal components of the STE and their variances, respectively. The principal components with significant variances are the identified filters of our LN model.

The significance of the eigenvalues was tested in a nested sequence. Initially, the null hypothesis was that all eigenvalues are not significant (Rust et al., 2005; Schwartz et al., 2006). We shuffled the trials to create 1000 sets of data. This made 1000 sets of eigenvalues, each sorted into rank order. The 0.5 percentile of the lowest rank was the lower bound, and the 99.5 percentile of the highest rank (first rank) was the upper bound of the shuffled eigenvalue. We checked whether any of the original eigenvalues exceeded the bounds. If none of them did, the null hypothesis was regarded as correct, and the sequence of tests was stopped. Otherwise, the null hypothesis was rejected, and the eigenvalue that deviated most from the bounds was tagged as being significant. If the tagged eigenvalue was above the upper bound, its eigenvector was added to the list of excitatory elements. If the eigenvalue was below the lower bound, the eigenvector was added to the list of suppressive elements. The second round of the sequence started by projecting out the identified eigenvalues from the STE. The updated null hypothesis was that the remaining eigenvalues were not significant. The lowest rank in this round moved up by one. The sequence of tests was repeated until the null hypothesis was not rejected.

The outputs of the linear filters are each passed through a squaring nonlinearity. We refer to the linear filters, followed by their output nonlinearity, as “elements” to emphasize their hypothetical nature (Rust et al., 2005); they may not correspond with real physiological subunits that provide input to a cell, but the summed response of these elements describes the subspace in which the images of the STE lie. The majority of our analysis therefore uses the linear sum of element responses to characterize the model; this model structure is analogous to the traditional disparity energy model (Ohzawa et al., 1990). Thus, the structure of our model is as follows:

$$f(s_1, s_2, \dots, s_M) = [a_0 \cdot [s_0]_+^2 + a_1 \cdot s_1^2 + a_2 \cdot s_2^2 + \dots + a_M \cdot s_M^2]_+, \quad (4)$$

where s_k denotes the output of the k th linear filter ($k = 0, 1, 2, \dots, M$). The square brackets $[\cdot]_+$ represent half-wave rectification. The 0th element is simple-cell like. The value of s_k is called the feature contrast.

To estimate the weights a_k for the elements, we applied the same methods as used for the traditional energy model (Schwartz et al., 2006). For each element, we calculate the feature contrast, s_k , on every frame. To estimate the probability that a given feature contrast is associated with a spike, we binned images into groups with similar s_k and measured the fraction of these images appeared in the STE as $n(s_k)/n_{\text{STE}}(s_k)$, where $n(s_k)$ and $n_{\text{STE}}(s_k)$ are the number of frames in this group of the original stimulus ensemble and the STE, respectively. This gives the probability that a spike will be elicited by this feature contrast over the duration of a single frame. Multiplying by the frame rate ($R_{\text{frame}} = 96$ Hz) converts this into firing rate in spikes per second. The border of the bins were the 0th, 5th, 10th, ..., 100th percentiles. The center of a bin was the median value within that bin. These response curves,

$$\frac{n_{\text{STE}}(s_k)}{n(s_k)} R_{\text{frame}} = a_k \cdot s_k^2 + b_k, \quad (5)$$

were constructed for all elements, and then the parameters a_k and b_k were fitted by minimizing the least-square error.

Analysis of the identified model. The identified elements were classified as either excitatory or suppressive, depending on whether increasing feature contrast increased or decreased the probability of eliciting a spike. Too much importance should not be attached to the shape of individual filters. The space spanned by the set of filters is what the data constrains, and quite different individual filters could be used to span the same space. To summarize this space, we analyze the signal that was carried within each pool by simulating the summed response of all elements to RLSs. This was done separately for the pooled excitatory elements only, the suppressive elements only, and all elements. We generated 1000 independent frames of RLS for each disparity value. The model disparity-tuning function was calculated by averaging the response to all the frames.

The disparity tuning of each pool was summarized with two numbers, a position disparity and a phase disparity, inferred from the symmetry of the tuning function. We calculated the centroid μ and the symmetry phase φ of a disparity-tuning function $v(z)$ of a pool (Read and Cumming, 2003), as

$$\mu = \frac{\int v(z) z dz}{\int v(z) dz}$$

$$\varphi = \arg\left(\sum_w F[v(z - \mu)]\right), \quad (6)$$

respectively. $F[\cdot]$ is the complex Fourier transform, and \sum_w is the summation across frequency components. The values of μ and φ were the estimates of position and phase disparities signaled by the pool, respectively. We used circular statistics for the analyses of symmetry phase φ (Mardia and Jupp, 1999). To estimate the preferred SF of each binocular

element, we calculated the cross-correlation between the filters of the left and right eyes and took the peak in the amplitude spectrum. We determined the SF of a pool by calculating a weighted mean of the elements in that pool, with weights based on eigenvalues.

The inclusion of the suppressive pool had only a modest effect on the shape of the disparity-tuning curve, which averages the responses to many stimuli. To explore the role of the suppressive pool in particular images, we analyzed the model responses frame by frame. The responses of one element to an image presented at different disparities are equivalent to a population response of a map of similar detectors, with different preferred disparity, to a single image. The stimuli were RLSs convolved with a low-pass filter (to ensure that the response maps were smooth). The low-pass filter was a Gaussian function with $\sigma = 1.25\Delta$, where Δ is the pixel width. The disparity of the RLS was fixed to be equal to the preferred disparity of the identified model. The true peak was estimated as the peak after averaging the population response to 1000 independently generated stimuli. Local peaks were detected as an increment followed by a decrement. A local peak was classified as a false peak if it was at least two pixels away from the true peak.

Comparison with conventional disparity tuning. The above analyses explore the subspace spanned by the images that were associated with spikes. To compare the observed responses of a neuron with the predictions of its identified model, it is useful to include a final nonlinearity that determines the relationship between the summed response of excitatory elements, r_{exc} , the summed response of the suppressive elements, r_{sup} , and the firing rate. Rust et al. (2005) found that the firing rate depended on both r_{exc} and r_{sup} . Following their example, we fitted functions that convert combinations of $(r_{\text{exc}}, r_{\text{sup}})$ into firing rate. The values r_{exc} and r_{sup} were computed for every frame and then were grouped into bins. The mean response of the neuron in each bin was compared with the model response. We explored two types of nonlinearity. The simplest was just a threshold followed by an expansive nonlinearity:

$$g(r_{\text{exc}}, r_{\text{sup}}) = \alpha + [\beta \cdot (r_{\text{exc}} - r_{\text{sup}}) + \delta]_+^\gamma. \quad (7)$$

The parameters, α , β , γ , and δ were fitted using the least-square method. The second nonlinearity, which includes normalization, was the same as used by Rust et al. (2005):

$$h(r_{\text{exc}}, r_{\text{sup}}) = \alpha + \frac{\beta \cdot r_{\text{exc}}^\gamma - \delta \cdot r_{\text{sup}}^\gamma}{\gamma \cdot r_{\text{exc}}^\gamma + \varepsilon \cdot r_{\text{sup}}^\gamma + 1}. \quad (8)$$

Both nonlinearities operate on the same bivariate input $(r_{\text{exc}}, r_{\text{sup}})$, but the model of Equation 7 has two fewer parameters than the model in Equation 8.

Results

The identified LN model

The LN model has been widely used to characterize sensory neurons, especially in cases in which the spike-triggered average is informative. Even when responses are composed from the summed output of multiple LN elements, white-noise analysis can be used to estimate those elements, using spike-triggered analysis of covariance (de Ruyter van Steveninck and Bialek, 1988; Touryan et al., 2002; Rust et al., 2005; Schwartz et al., 2006). Figure 1A illustrates the underlying model. The response of the neuron is assumed to be the summed output of a number of elements, each of which consists of a linear filter in each eye, followed by a static nonlinearity. The elements can be either excitatory with a positive coefficient or suppressive with a negative coefficient. The summed response is half-wave rectified, and the resulting value determines the probability that a spike is generated. Under the assumptions of the model, spike-triggered analysis of covariance successfully reconstructs the underlying filters.

We recorded the spike train of V1 neurons as the monkey maintained fixation. One-dimensional noise images (independent samples with 21 lines in each eye) were presented at the preferred orientation over the receptive fields of the left and right

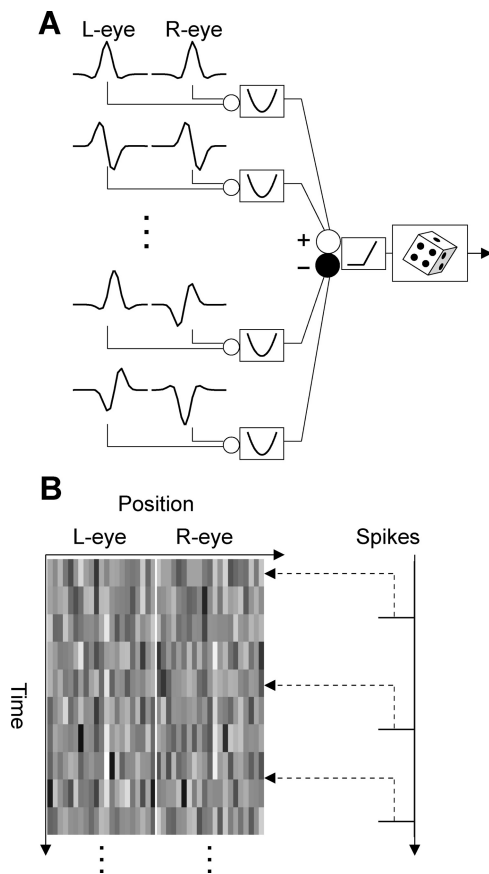


Figure 1. Identification of the elements with spike-triggered analysis. **A**, The elements of the model have a linear filter in each eye followed by an output nonlinearity. The outputs of the elements are combined linearly, with some of their coefficients being negative. After a rectified nonlinearity, the signal intensity determines the firing rate. **B**, Sequences of noise patterns were presented to both eyes. The patterns were uncorrelated between the eyes. One noise frame preceding each spike was extracted. The spike-triggered stimulus ensemble consists of all such frames preceding a spike. L, Left; R, right.

eyes. After the data were collected, we looked backward in time and triggered one noise frame with every spike (Fig. 1B). This generates a large ensemble of images, the STE. The common properties of images in the STE indicate the features in the stimulus that caused the neuron to fire (Schwartz et al., 2006). Each frame was represented as a vector in a 42-dimensional space. The STA is the average of the STE. The STC matrix represents the second-order interactions present in the STE.

Figure 2A shows the STC matrix of an example cell that showed strong binocular interaction. In this matrix, each element records the covariance between pairs of pixels in the STE. In the quadrant in which the pixels come from different eyes (bottom left, top right), such covariances reflect binocular interactions. The other quadrants represent interactions within the image of one eye. In this example, there is a narrow region of positive covariance along a diagonal strip in the binocular quadrant of the matrix. This indicates that the cell was excited by images that had the same luminance in the appropriate positions in the two eyes. The fact that the points lie on a line with slope close to -1 indicates that the same disparity was effective at all the locations. Small regions of negative covariance flanked the positive strip. The negative regions indicate that the cell was either excited by images that had the opposite luminance and/or suppressed by images that had the same luminance, in slightly disparate positions between the eyes. A striking feature of this example is that

the monocular quadrants show weaker signals than the binocular quadrant. This implies that interactions between pixels in different eyes have a stronger influence on the firing of the neuron than interactions between pixels within one eye. This property is consistent with the idea that an important function of this neuron is to signal stereoscopic depth.

We used principal components analysis to compactly summarize the STC matrix. First, for each image in the STE, the vector component along the axis of the STA was subtracted to make any principal component orthogonal to the STA. The principal components were calculated by an eigenvalue decomposition of the resulting STC. This STC matrix in Figure 2A had four eigenvalues that were significantly larger than chance and three eigenvalues that were significantly smaller than chance (nested resampling, $p < 0.01$; see Materials and Methods) (Fig. 2B). These seven LN elements capture the statistically significant modulation in the STC matrix. Finally, the STA yields the filter describing a traditional simple-cell model.

In the example shown, the filters of the excitatory elements had similar shapes in the left and right eyes (Fig. 2C). In contrast, the filters of the suppressive elements had inverted shapes between the eyes (Fig. 2D). Thus, the relationship between left and right eye filters is inverted in the suppressive element compared with the excitatory elements. This was also true for another example cell in which the excitatory elements had dissimilar RFs between the eyes (Fig. 2E–H). The inversion was a common property of suppressive elements, as we show below.

Suppressive elements in the population

We recorded 70 disparity-selective cells from two monkeys. There were 66 cells (36 from monkey duf and 30 from monkey ruf) that had at least one significant element. Only one neuron behaved like a traditional simple cell, i.e., the only significant element was the STA. Of the remainder, 88% (57 of 65) had more elements than the traditional complex model, similar to the proportion reported by Rust et al. (2005). Even this figure is probably an underestimate: for statistical reasons, more spikes (more images in the STE) are required to identify additional elements, so the number of significant elements increased as the size of the STE increased ($r_s = 0.50$, $p = 0.0019$) (Fig. 3). We found at least one excitatory element and at least one suppressive element in 29 cells. They constituted 51% (29 of 57) of all the cells that had more than two elements. Again this figure represents a lower bound of the true fraction, because identifying suppressive elements requires large STEs. For neurons in which we were able to record at least 20,000 spikes, 24 of 35 (69%) revealed at least an excitatory and a suppressive element.

Functionally push–pull-like combination of signals

Each half of the filter of an element describes the receptive field in one eye. For these LN elements, the cross-correlation function between the left and right RFs predicts the disparity tuning of the element (Fleet et al., 1996; Zhu and Qian, 1996; Prince et al., 2002) and generates the structure seen in the binocular quadrant of the STC matrix (Ohzawa et al., 1990). For the cell shown in Figure 2A–D, the cross-correlation functions of all four excitatory elements were even symmetric and maximal at -0.1° disparity (Fig. 4A). In contrast, the cross-correlation functions for the suppressive elements were minimal at -0.1° disparity (Fig. 4B). The inverted shape of the cross-correlation functions for the suppressive element demonstrates a push–pull-like organization for disparity; disparities that produce excitation also produce withdrawal of suppression. Although we describe the functional

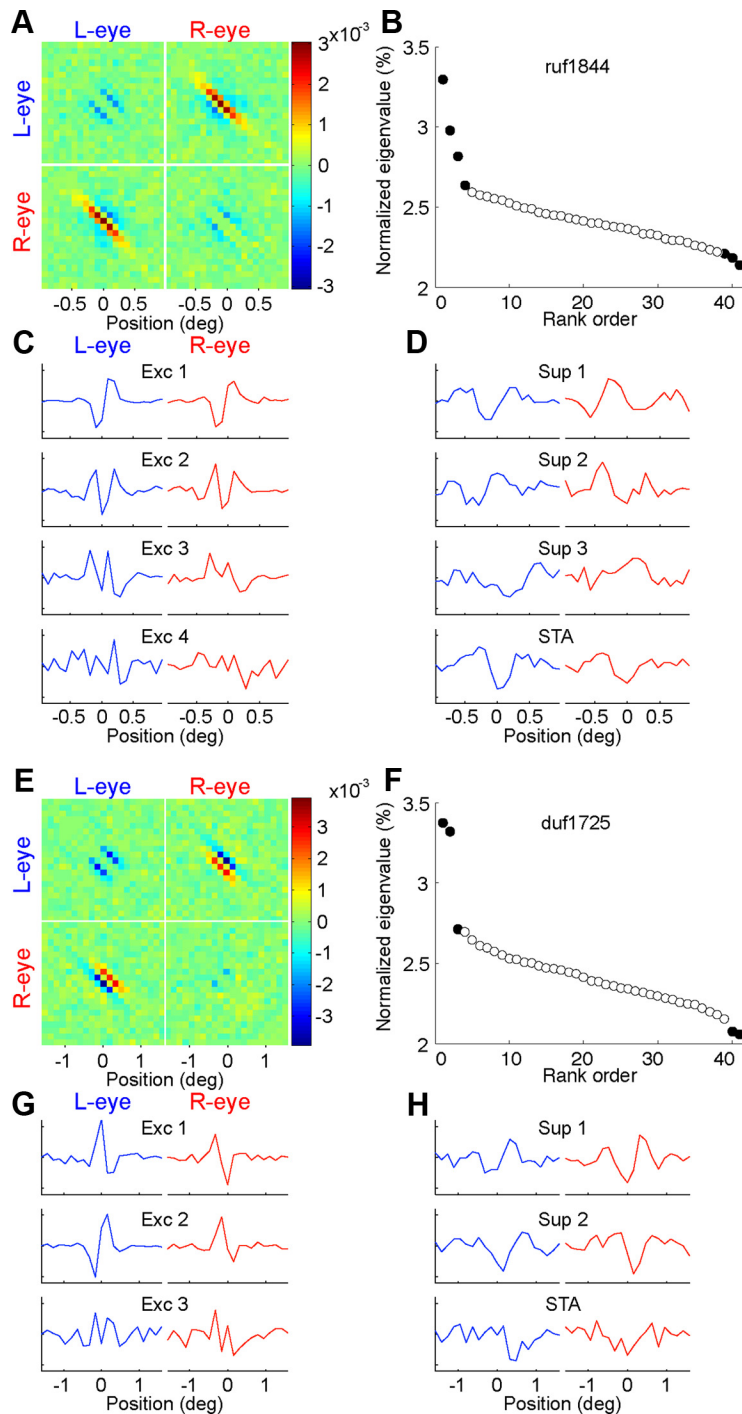


Figure 2. Identified elements of two example cells. **A**, The STC matrix has four quadrants. The top left and the bottom right quadrants show correlations between pixels shown to the left eye (L-eye) or to the right eye (R-eye). Structures in the top right and bottom left quadrants show binocular interaction. The diagonal was replaced with zeros in this figure to show the important structures in the matrix. The diagonal contained large values in the actual calculations (corresponding to the variance of each pixel). The color bar shows the scale of covariance values. **B**, After projecting out the STA from the ensemble, the eigenvalues of the STC matrix reveal the significant elements (filled symbols). The eigenvalues are normalized to a sum of one. The last eigenvalue was always zero, corresponding to the axis that had been projected out (data not shown). **C**, Four elements were identified as excitatory (Exc) elements. The left column shows the eigenvectors (receptive field elements). The blue and red segments show the filters of the left and right eyes. The ordinate is in arbitrary units. **D**, Three elements were identified as suppressive (Sup) elements. The STA is shown on the bottom row. **E–H**, Another example cell. The plotting conventions are the same as in **A–D**. This cell revealed two suppressive elements.

relationship between excitation and suppression as push–pull-like, the description does not necessarily imply a push–pull relationship in the synaptic inputs to the cell (Priebe and Ferster, 2005). The synaptic mechanism of suppression can be either

inhibition or withdrawal of excitation. Excitation is withdrawn when the inhibition takes place earlier in the pathway.

To demonstrate push–pull organization for each neuron, in a way that does not depend on analyzing the shapes of individual filters, we first summarized the contribution of all excitatory elements (including the STA, if present) by simulating the summed signal in response to an RLS. We then compared this with the disparity tuning of the pooled suppressive elements. The example in Figure 4C shows a tuning curve typical of “tuned–excitatory” neurons for the excitatory pool, whereas the suppressive pool shows a pattern typical of “tuned–inhibitory” neurons (Poggio et al., 1988). The tuned–inhibitory response arises in these LN elements when receptive fields have dissimilar shapes in the two eyes (DeAngelis et al., 1991). Here left and right RFs are related by a phase shift close to π (large phase disparity). The excitatory elements have similar shapes in the two eyes (little phase disparity). Note that the sign convention of the ordinate is the response of the suppressive pool, not its effect on the recorded cell. This tuning function is subtracted from, not added to, the tuning function of the excitatory pool and hence reinforces the disparity selectivity. The inclusion of the suppressive pool had little effect on the shape of the disparity tuning (Fig. 4D). For this demonstration, the suppressive pool was combined linearly with the excitatory pool. The linear combination allows us to separate contributions from each pool, without any interactions introduced by the final output nonlinearity.

The second example cell had different disparity selectivity, with odd–symmetric tuning for the excitatory pool (phase disparity near $\pi/2$). Nonetheless, the difference in phase disparity between excitatory and suppressive responses was close to π .

We selected the subset of neurons that had at least one excitatory and at least one suppressive element ($n = 29$) and estimated the phase disparity of each pool from the symmetry of the disparity–tuning function [symmetry phase (Read and Cumming, 2003)]. The disparity response of suppressive elements was systematically out–of–phase with that of the excitatory elements: there was a significant circular correlation ($r_\theta = 0.71$, $p = 3.6 \times 10^{-4}$) (Fig. 5A), with a mean circular difference of -0.99π . We also explored disparity coded by simple translations of the RF (position disparity) by measuring the centroid of the disparity tuning (Read and Cumming, 2003). These were mostly distributed around zero and not correlated between the two pools ($r_s = 0.07$, $p = 0.71$) (Fig.

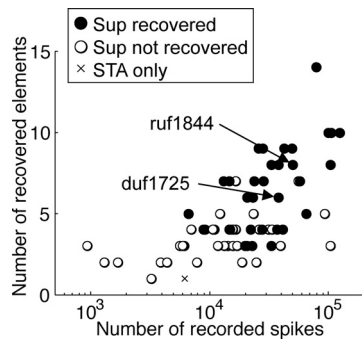


Figure 3. The dataset of this study. The number of recovered elements is plotted against the number of spikes recorded for each cell. A larger number of spikes more frequently revealed more elements than the original energy model. The additional elements frequently included suppressive (Sup) elements.

5B). The results suggest that the pools are organized in a push–pull manner. The disparity that maximally excites the excitatory pool, maximally inhibits the suppressive pool and vice versa. This mechanism is analogous to the organization of thalamocortical inputs to V1 (Priebe and Ferster, 2005) but operating in the disparity domain.

We found only one other property that was systematically different between the pools: the SF of the suppressive pool was lower than the excitatory pool (Wilcoxon’s signed rank test, $p = 2.9 \times 10^{-4}$) (Fig. 5C). The geometric mean of the frequency ratio was 0.51 (0.98 octave difference). This suppression from disparities at a coarser scale has long been known to be a useful strategy in solving the correspondence problem (Marr and Poggio, 1979). A straightforward advantage for disparity tuning is that the suppressive elements reduce the side lobes in the disparity-tuning function. The effect of suppression on the tuning function is the same as adding a same-sign function with low SF. Recent evidence has suggested that such “coarse-to-fine” mechanisms do operate early in binocular processing (Menz and Freeman, 2004a,b). Our study extends these by characterizing the spatial structure and suppressive nature of this interaction.

Reduced responses to false matches

The relationship between phase disparities of excitatory and suppressive pools suggests a possible role in solving the correspondence problem. Binocular LN elements will, on average, produce their strongest response to stimuli with a disparity that matches the RF disparity of the filter. However, any given single image (e.g., a single RLS) may produce even stronger responses at other disparities (false matches). Our group recently pointed out that neurons with phase disparities may respond more strongly to these false matches (Read and Cumming, 2007; Haefner and Cumming, 2008). For this reason, suppression from neurons with phase disparity onto neurons without may help eliminate responses to false matches (Read and Cumming, 2007). Because the suppressive elements we identified have different phase disparities from the excitatory elements, they may serve this same function.

The identified model allowed us to evaluate this possibility directly. For each neuron, we explored the matching problem by considering the response of a population of detectors that were identical to the excitatory pool for that neuron, but each was given a different position disparity. The response of this population to a single image at one disparity then illustrates the problem of false matches (Fig. 6A). For the example shown, there are three

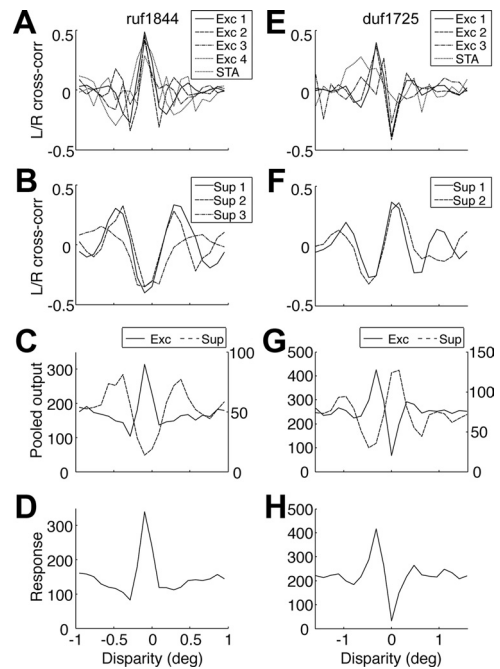


Figure 4. Prediction of the disparity tuning with the identified elements. **A**, The cross-correlation function between the left (L) and right (R) eye segments of the filter are shown for each element. The cross-correlation functions for the four excitatory (Exc) elements and the STA are superimposed. They are all in-phase. This is the same cell as in Figure 2A–D. **B**, The cross-correlation for the three suppressive (Sup) elements are shown. They are in-phase with each other. Same cell as in **A**. **C**, The identified elements were used to simulate the output of the excitatory pool (solid, left axis) and suppressive pool (dashed, right axis) in response to a random-line stereogram. The average output of 1000 frames is shown. **D**, The disparity tuning of the full model was simulated. **E–H**, The same simulations as in **A–D** but with the elements identified in the same cell as Figure 2E–H.

local maxima in the population response (disparity, -0.9 , -0.1 , and 0.5°). Indeed, the overall maximum response was to a false match (0.5°). If the visual system identified the disparity from the maximum response in such a population, the wrong depth would be perceived.

Comparing these responses with a population that linearly sums excitatory and suppressive pools, detectors of all false disparities are suppressed to some degree (Fig. 6A); only the detector of the true disparity is consistently free of suppression (-0.1°). As a result, the false peaks are moderately suppressed on average, whereas the correct peak remains unchanged. Thus, the inclusion of the suppressive pool helps disambiguate the correct peak from the false peaks.

We repeated this simulation for 1000 independently generated images, locating local peaks in each map. We then measured the response difference between the correct peak and the closest false peak. Figure 6B compares this measure for the excitatory pool with the measure when suppression is included. Almost every point lies above the diagonal, indicating that the suppressive pool systematically reduces responses to false matches more than it reduces responses to the correct disparity.

The results for each of the 29 cells that revealed suppressive elements were summarized with the mean (Fig. 6B, white cross). The selective suppression of responses to false matches was consistent across the population (Wilcoxon’s signed rank test, $p = 5.3 \times 10^{-6}$) (Fig. 6C). The best suppression was seen in the even-symmetric cells (Fig. 6C, filled symbols). These are the cells that are maximally excited by naturally occurring dispar-

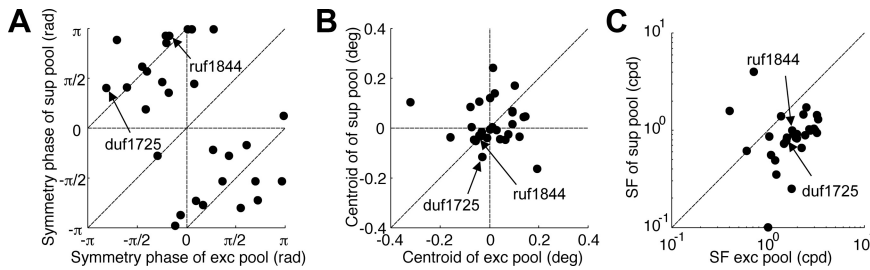


Figure 5. Disparity-tuning characteristics of excitatory and suppressive pools. Each panel shows the symmetry phase (an estimate of phase disparity) (**A**), centroid (an estimate of position disparity) (**B**), and SF (**C**), respectively. The value for the (exc) excitatory pool is plotted on the abscissa, and the value for the suppressive (sup) pool is plotted on the ordinate. Arrows indicate the example cells in Figure 2, *A–D* (ruf1844) and *E–H* (duf1725). The dashed lines traversing the top left and bottom right quadrants in **A** represent an antiphase relationship.

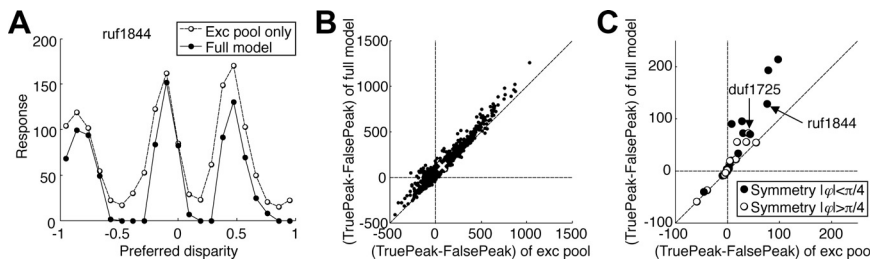


Figure 6. Reduction of false peaks by the suppressive pool. **A**, A population of disparity detectors was created using an identified model (ruf1844). One population had only the excitatory pool, whereas the other used the full model. The responses of the populations to a single static image are shown. **B**, The height of the true peak (at -0.1°) relative to the false peak was calculated for both models in response to 1000 different images. The difference in height is plotted for the full model against the excitatory pool only. The white cross indicates the mean of the response difference pairs. **C**, The simulations are summarized by the mean for each cell. The filled symbols indicate the cells with an even-symmetric excitatory pool.

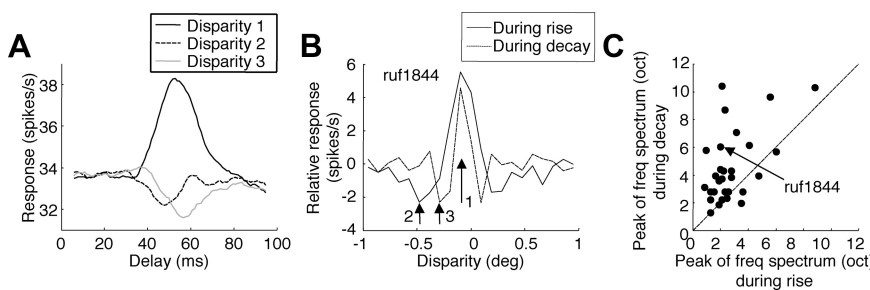


Figure 7. Dynamics of disparity tuning with stimulus-triggered spikes. **A**, Spike density functions triggered on the presence of positive correlation in the images at a particular disparity. These correlations at the preferred disparity (disparity 1) produce a transient increase in activity. At disparities 2 and 3, they produce a transient reduction in activity. Disparity 2 produced the weakest response during the rising phase of the response, whereas disparity 3 produced the weakest response during the decay. The minimum response for disparity 2 precedes the minimum for disparity 3. **B**, From spike density functions for all disparities, two cross-sections were extracted showing disparity selectivity at different times. A delay of 10 ms before the optimal represented the rising phase. The decaying phase was 10 ms after the optimal. The disparities shown in **A** are marked with the corresponding numbers. In this example, the disparity tuning sharpened over time. **C**, Population data. The disparity-tuning functions were transformed to frequency spectra (after zero padding to improve the frequency resolution). The peak frequencies were expressed as multiples of the fundamental frequency (units of octave). The peaks of the spectra shifted to higher frequencies over time.

ities, for which Read and Cumming (2007) suggested that the suppression we observe might help solve the stereo correspondence problem.

This analysis of peak responses shows a selective strengthening of responses to true matches. How effectively this reduces the matching problem depends on what decoding rule is applied to the population response. We chose our analysis of peak responses so as to avoid presenting results that depend on the particular decoding rule.

Coarse-to-fine dynamics of disparity tuning

We noted above that the coarser scale of the suppressive subunits might represent a coarse-to-fine matching mechanism. If this coarse-to-fine interaction evolves over time, it should result in a systematic sharpening of the disparity-tuning curve with time. We explored this by analyzing the dynamics of disparity tuning by forward correlation of stimulus and response. For each stimulus in the original ensemble and each disparity value, we computed the correlation between the images in the left and the right eyes. Then for a given disparity, we searched through the ensemble for positive values. We triggered the spike train on every occurrence of a positive value. The triggered spike trains were averaged and then filtered (Gaussian window with $\sigma = 1$ frame width). The result of this forward correlation analysis is a spike density function (SDF) in response to positive binocular correlation for a given disparity (Fig. 7A).

The difference between two SDFs (the response to positive with respect to the response to negative binocular correlation) shows the dynamics of disparity tuning. The tuning function was broad during the rising phase of the response. The tuning function gradually became sharper with longer delay times (Fig. 7B). We characterized the tuning width by the peak of the frequency spectrum. The peak shifted by a median of 1.5 octaves between the rising and decaying phases (Fig. 7C) (Wilcoxon’s signed rank test, $p = 1.8 \times 10^{-4}$). The disparity tuning had a temporal order of coarse-to-fine. This result is consistent with data from anesthetized cats (Menz and Freeman, 2004a,b).

Comparison with conventional disparity tuning

We tested whether the model was capable of explaining conventional disparity tuning measured with RLSs. These stimuli were composed of high contrast lines and so were different from the stimuli used for the white-noise analysis. The ability of the model to predict steady-state responses to this stimulus is therefore a strong test of its ability to explain neuronal responses. Because this analysis compares observed and

predicted firing rates, it requires that we first estimate any output nonlinearity that relates the summed response of the elements to the firing rate of the neuron. For this purpose, we calculated the response of each cell to different combinations of summed excitation and suppression in the model (following Rust et al., 2005). Figure 8A shows two cross-sections through this 3-D surface: the response as a function of excitatory pool input, for maximal suppressive input and for minimal suppressive input. The downward displacement between these curves reflects the subtractive input

from the suppressive elements. High suppressive input also reduces the gain of responses to excitatory input. We quantified this with linear regression. All 29 cells showed the reduction in slope, with a median slope 20% lower for maximal suppression compared with minimal suppression.

Rust et al. (2005) accounted for these slope changes with a normalization model (Fig. 8*B*). Our data are almost equally well described by a simpler model that applies an expansive nonlinearity after summing the two pools, requiring fewer parameters (Fig. 8, compare *A*, *B*). The normalization model was modestly better, explaining 97.5% of the variance (median) compared with 96.9% for the output nonlinearity alone (Wilcoxon's signed rank test, $p = 3.6 \times 10^{-4}$), but this is not surprising with two more parameters in the normalization model. Because both models produce very similar fits to the data, none of the conclusions we present are affected by the choice of model.

The analysis illustrated in Figure 8 also allowed us to quantify how strongly the suppressive elements influenced firing rate, using the fractional suppression described by Rust et al. (2005). This is the ratio of responses with maximal excitation and maximal suppression to the response with maximal excitation and minimal suppression. The fractional suppression had a median value of 0.68 (Fig. 8*D*). This was similar to the 0.63 reported by Rust et al. (2005).

Having reconstructed the output nonlinearity, we simulated the responses of the cell to RLSs presented at various disparities. We compared these simulations with separate measurements, made on the same cells using traditional tuning curves (Fig. 9*A,B* shows examples). The median correlation coefficient between predicted and observed response was 0.88 for the 29 cells with suppression (0.82 for all 64 cells), indicating that the model predicted the shape of the disparity-tuning curve well (Fig. 9*D*).

To compare predicted response magnitudes, we took the slope of a type II linear regression of predicted versus observed responses. The median slope was 4.7 (larger than 1.0, Wilcoxon's signed rank test, $p = 2.6 \times 10^{-6}$ for the 29 cells with suppression; median of 2.9; $p = 3.5 \times 10^{-12}$ for all 64 cells). This indicates that the predicted responses were generally larger than the observed responses. This probably reflects the effect of contrast normalization: the root mean square stimulus contrast of the white-noise stimulus was lower than that of the RLS used for the disparity tuning. Any saturation of responses at high contrast will therefore result in observed responses that are smaller than the model predictions. Thus, although the responses to the white-noise stimulus alone do not clearly indicate the operation of a normalization mechanism, our data as a whole do.

No attenuation with anticorrelated stimuli

Simulating the responses of the identified models (Fig. 6) revealed that responses to false matches were suppressed. We therefore explored whether the model might also explain responses to a different type of false match: those produced by anticorrelated stereograms. Anticorrelated stereograms have no corresponding patterns between the eyes yet evoke responses in disparity-selective neurons (Cumming and Parker, 1997). These responses are on average weaker than responses to correlated stereograms, suggesting that responses to false matches are attenuated. We

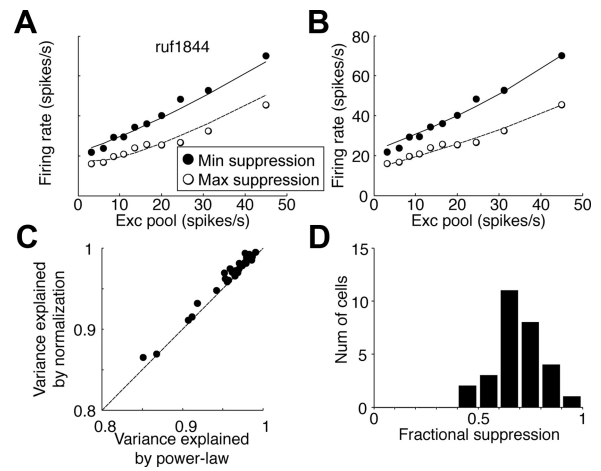


Figure 8. Combining excitatory and suppressive signals. **A**, The observed firing rate is shown as a function of the calculated responses of the excitatory (Exc) pool and the suppressive pool. The filled and open symbols are the two cross-sections of the response function. Filled symbols show responses when the response of the suppressive pool was at a minimum. Open symbols show responses for maximal suppression. The curves are the responses of the model with a power-law function in Equation 7. **B**, The symbols are copied from **A**. The curves are the fitted model with a normalization function in Equation 8. Both models captured the reduced slope with suppression. **C**, The variance explained by the power-law model and the normalization model was comparable, although divisive normalization was slightly better. **D**, Fractional suppression is the ratio of the two symbols at the right end of the response function in **A**. The distribution shows that suppression was clearly present in most cells and made a substantial contribution to the responses.

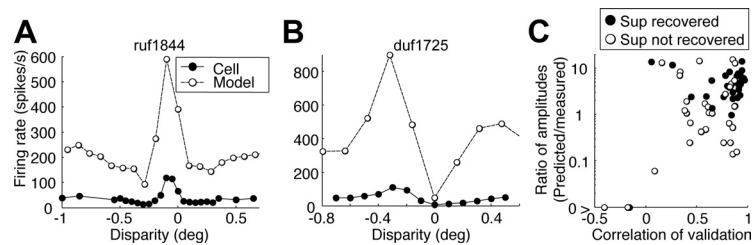


Figure 9. Comparing disparity tuning of the models with the neurons. **A**, The disparity tuning of the model is superimposed on the actual disparity tuning. Error bars indicating SEM are smaller than the data points in this scale. This is the same cell as in Figure 2*A–D*. **B**, This is the same cell as in Figure 2*E–H*. **C**, The ratio (slope of regression line) between the response of the model and the response of the cell is plotted as a function of the correlation. Three cells had a negative ratio, falling outside the plotted range. Sup, Suppression.

simulated the responses of the identified models to anticorrelated RLSs, to see whether our models explain this property. Figure 10 shows data for the two example cells and the corresponding identified model. Both identified models showed minimal attenuation of disparity tuning with anticorrelated stereograms (Fig. 10*A,B*), whereas the neurons show substantial attenuation (Fig. 10*E,F*).

To pursue this discrepancy, we estimated neuronal responses to correlated and anticorrelated disparities in our white-noise stimulus. We analyzed the stimulus-triggered response (forward correlation) as in Figure 7 but used a higher threshold of binocular correlation. The responses to images that were greater than the 75th percentile or smaller than the 25th percentile represented the responses to positive and negative correlations, respectively. These two SDFs were then separately compared with that produced by near zero correlations (25th to 75th percentiles). The SDF for each disparity was then integrated over time to arrive at a disparity-tuning function. These estimates showed no attenuation in the response to anticorrelation, very similar to the

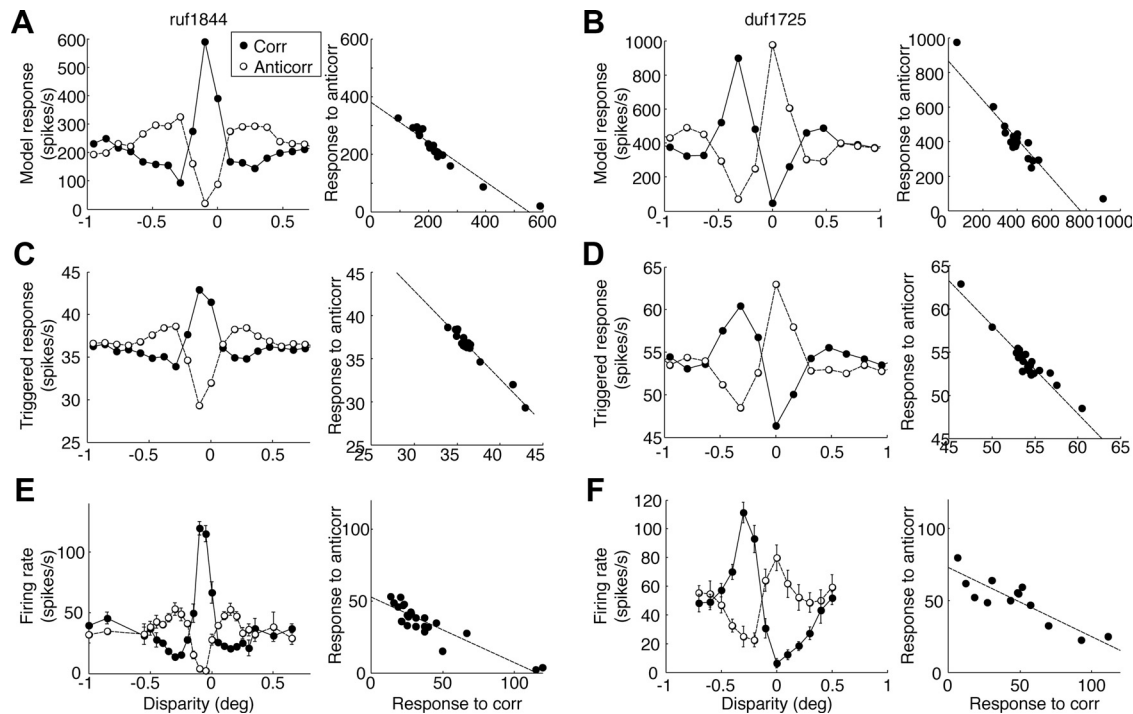


Figure 10. Neuronal and model responses to anticorrelated stimuli, for the two example cells. **A, B**, The prediction of the model of the disparity-tuning function for correlated (Corr, filled) and anticorrelated (Anticorr, open) RLSs (left). The scatter plot of the responses between same disparity (right). The slope of the regression line (dashed) is the amplitude ratio of the two tuning functions. The predicted slopes were both close to -1 (-0.69 , -1.1). **C, D**, Disparity tuning estimated from responses to noise stimuli. The regression slopes were -1.0 in both cells, indicating no attenuation of response with anticorrelation. **E, F**, The actual response of the cell to correlated (filled) and anticorrelated (open) RLSs in a traditional tuning curve. The regression slopes (-0.45 , -0.48) were lower than in **A** and **B**.

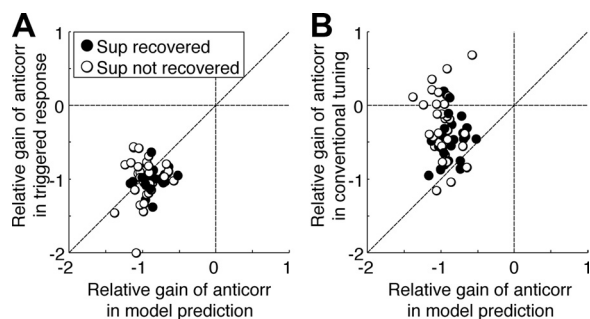


Figure 11. Responses to correlated (Corr) and anticorrelated (Anticorr) stimuli. The ratio (amplitude of response modulation to anticorrelated patterns)/(amplitude of response to correlated stimuli) defines “relative gain.” **A**, The relative gain predicted by the model is close to 1, as are the spike-triggered responses. **B**, In traditional tuning curves, the gain was substantially lower than predicted by the identified model. Thus, the model did not account for the attenuated response of the cells to anticorrelated RLS in traditional tuning curves. Sup, Suppression.

model response (Fig. 10C,D). Because the data from which the model was derived did not exhibit attenuation with negative correlations, it is unsurprising that the models themselves did not show attenuation.

The bottom row shows the responses of the cells to the RLSs in traditional tuning curves (Fig. 10E,F). Both cells showed substantial attenuation with anticorrelated stereograms (amplitude ratios were -0.45 and -0.48 , respectively). Figure 11 summarizes this analysis for all 60 disparity-selective cells. The amplitude ratios of the model prediction and the triggered response were both near -1 , with the model predicting a slightly lower ratio (median of -0.95 vs -1.0 ; Wilcoxon’s signed rank test, $p = 0.012$; $n = 58$). The ratio in the traditional tuning curves indicated a significant attenuation (median of -0.41 ; $p = 1.6 \times$

10^{-10}), similar to that reported previously. We examined the subset of neurons for which we identified at least one excitatory and one suppressive element. The amplitude ratios were slightly lower with the model prediction than the triggered response (median of -0.90 vs -1.0 ; $p = 0.005$; $n = 25$). The response in traditional tuning curves was significantly more attenuated (median of -0.45 ; $p = 2.0 \times 10^{-5}$).

Thus, the failure of the model to describe response to anticorrelated RLS in tuning curves reflects the fact that the neurons respond differently to negative correlation in the noise stimulus and in traditional tuning curves, and the model correctly describes the neuronal responses in the data to which it is fit. The difference between the neuronal responses in the two cases presumably represents the effect of some additional nonlinearity. What difference in the stimuli is responsible for this difference in response will be the subject of a future investigation.

Simulation of fixational eye movements

In generating our STE, we have assumed that the monkeys maintained perfect fixation throughout the trial. In practice, small fixational eye movements will generate random translations of the receptive field of the neuron relative to the stimulus (median of within-trial SDs, 0.053° and 0.056° for monkeys duf and ruf, respectively, for horizontal conjugate movements). We have explored the effects of this jitter in simulations, using both idealized energy models and using the identified excitatory elements. In no cases did adding random jitter to the stimulus location lead the STC analysis to produce suppressive elements like those we observed, so we do not believe that fixational eye movements can account for our findings.

We also asked whether fixational eye movements bias the excitatory and suppressive elements to a push–pull organization.

For this test, the model combined excitatory and suppressive elements that were not given a push–pull organization but were given different position disparities and both zero phase disparities. The STC analysis correctly recovered the elements, as long as the difference in the position disparities was sufficiently large. We also tested models with elements containing phase disparities that differed by $\pi/2$, rather than π . Again, the analysis correctly recovered the subunits, even in the presence of simulated eye movements. Hence, the push–pull pairing of excitatory and suppressive elements we find seems unlikely to be an artifact of eye movement.

Discussion

This study took advantage of white-noise analysis to identify a general LN model for disparity-selective cells in monkey V1. Our goal was to identify the binocular elements of the model with as few assumptions as possible. We found not only elements that represent excitatory drives, as expected from the disparity energy model, but also additional elements that reflect suppressive drives. Unexpectedly, we found that, in many neurons, the excitation and suppression were combined in a push–pull organization. Whenever a stereogram with the preferred disparity of the cell is presented, the excitatory elements drive the cell (push), whereas the suppressive elements disinhibit the cell (pull). We found that this push–pull organization helps suppress responses to false matches (spurious responses to stimuli not at the cells preferred disparity) in binocular images.

These spurious responses complicate the interpretation of responses across a population of disparity detectors. If one examines a population of energy-model like disparity detectors each tuned to a different disparity, the detector with the maximum response frequently does not have the same disparity as the stimulus. The suppressive elements identified in this study specifically attenuated these false peaks in the response, helping to disambiguate the true peak from the false ones. Thus, the push–pull organization represents a physiological step toward solving the correspondence problem in stereo computation (Marr and Poggio, 1979).

A recent computational study proposed an algorithm that estimates the correct disparity from a map of disparity-energy models (Read and Cumming, 2007). They proposed that the algorithm could be physiologically implemented by suppressing the signal of the even-symmetric (phase disparity, ~ 0) detectors with the signal of hybrid ($|\text{phase disparity}| > 0$) detectors. This is compatible with our finding that the suppressive signal is from detectors that are in antiphase (difference in phase disparity, $\sim \pi$). Although the hybrid detectors in their model spanned a range of phase relationships, our simulations here (Fig. 7) indicate that phase disparities in antiphase are sufficient to produce a similar effect. Furthermore, the computational study predicted that the effect of this suppressive signal is to reduce responses to false matches only in neurons with near zero phase disparity (Read and Cumming, 2007), a pattern we also found (Fig. 6C).

An important difference between our results and the computational algorithm is that the algorithm considered interactions between filters that were identical, apart from their binocular position and phase disparities. Our suppressive elements also showed a coarser spatial scale (lower preferred spatial frequency) than their excitatory counterparts. This allows information at coarse scales to influence matches made at finer scales. This coarse-to-fine constraint has long been recognized as useful for solving the correspondence problem (Marr and Poggio, 1979). We showed a sharpening of the disparity response over time,

consistent with this coarse-to-fine process and similar to that reported in V1 of anesthetized cats (Menz and Freeman, 2004a,b). Some human psychophysical work suggest that these coarse-to-fine interactions only operate between frequencies < 2 octaves apart (Wilson et al., 1991). This is similar to the range of frequency differences we observed, compatible with the idea that the suppressive elements we identified mediate coarse-to-fine interactions in stereo processing. Thus, it appears that the suppressive elements we identified simultaneously combine two principles for reducing responses to false matches: push–pull and coarse-to-fine organizations.

Anzai et al. (1999) used a related technique to study functional subunits of disparity-selective complex cells in cat V1. However, their analysis focused on excitatory elements, and it is not clear whether their data contained evidence for suppression. Other studies in the cat (not exploring binocular interaction) have also reported only excitatory elements (Touryan et al., 2002), but it is not yet clear whether this might be explained by methodological differences between that study and those in the monkey (Rust et al., 2005). Therefore, whether or not there is a species difference in the prevalence of suppressive elements remains an open question.

If the mechanisms we have identified correctly explained the responses of V1 neurons to all types of false matches, they should explain another phenomenon. When stimulated with random-dot stimuli of opposite polarity in the two eyes (anticorrelated stereograms), V1 neurons still show disparity selectivity but with reduced modulation amplitude (Cumming and Parker, 1997). We explored the responses of the reconstructed models to anticorrelated stereograms and found that they showed only very slight attenuation of this sort. We also found that the neuronal responses to anticorrelation in the white-noise stimulus showed little attenuation, although many of the neurons showed substantial attenuation when tested with traditional tuning curves. This suggests that mechanisms not engaged by the white-noise stimulus contribute to the stronger attenuation of tuning curves measured using anticorrelated stereograms. Nonetheless, our results identify mechanisms that effectively reduce the matching problem for more natural stimuli.

The push–pull arrangement of filters that we find is reminiscent of the push–pull mechanism found in thalamocortical projections. In the case of orientation selectivity (Hirsch et al., 1998; Troyer et al., 1998) and direction selectivity (Priebe and Ferster, 2005), the push–pull arrangement is thought to sharpen selectivity for these attributes. The binocular push–pull mechanism we describe may sharpen disparity selectivity. It also achieves something quite different: a greater robustness to false matches, or nuisance parameters in the stimulus. It will be interesting to see whether a push–pull arrangement of inputs might be also beneficial in other cases in which the brain has to infer a particular aspect of the stimulus while discarding others.

In summary, we report a physiological mechanism that helps solve the stereo correspondence problem at an early stage of binocular processing (V1). The mechanism is a push–pull organization in binocular receptive fields. A similar organization with a different functional implication has been shown for monocular processing in the thalamocortical projections (Priebe and Ferster, 2005) but has never been implicated in binocular processing. We show that simple operations based on the output of the energy model can explain how this mechanism is exploited in the cortex to solve a well-known computational problem faced by the visual system.

References

- Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for processing binocular information II. Complex cells. *J Neurophysiol* 82:909–924.
- Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389:280–283.
- Cumming BG, Parker AJ (2000) Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J Neurosci* 20:4758–4767.
- DeAngelis GC, Ohzawa I, Freeman RD (1991) Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* 352:156–159.
- de Ruyter van Steveninck R, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc R Soc Lond B Biol Sci* 234:379–414.
- Fleet DJ, Wagner H, Heeger DJ (1996) Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res* 36:1839–1857.
- Haefner RM, Cumming BG (2008) Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron* 57:147–158.
- Hirsch JA, Alonso JM, Reid RC, Martinez LM (1998) Synaptic integration in striate cortical simple cells. *J Neurosci* 18:9517–9528.
- Horwitz GD, Chichilnisky EJ, Albright TD (2005) Blue-yellow signals are enhanced by spatiotemporal luminance contrast in macaque V1. *J Neurophysiol* 93:2263–2278.
- Hunter IW, Korenberg MJ (1986) The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biol Cybern* 55:135–144.
- Julesz B (1971) Foundations of cyclopean perception. Chicago: University of Chicago.
- Mardia KV, Jupp PE (1999) Directional statistics (Wiley series in probability and statistics). Chichester, UK: Wiley.
- Marmarelis VZ, Citron MC, Vivo CP (1986) Minimum-order Wiener modeling of spike-output systems. *Biol Cybern* 54:115–123.
- Marr D, Poggio T (1979) A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci* 204:301–328.
- Menz MD, Freeman RD (2004a) Temporal dynamics of binocular disparity processing in the central visual pathway. *J Neurophysiol* 91:1782–1793.
- Menz MD, Freeman RD (2004b) Functional connectivity of disparity-tuned neurons in the visual cortex. *J Neurophysiol* 91:1794–1807.
- Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249:1037–1041.
- Poggio GF, Gonzalez F, Krause F (1988) Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci* 8:4531–4550.
- Priebe NJ, Ferster D (2005) Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron* 45:133–145.
- Prince SJ, Pointon AD, Cumming BG, Parker AJ (2002) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J Neurophysiol* 87:191–208.
- Read JC, Cumming BG (2003) Testing quantitative models of binocular disparity selectivity in primary visual cortex. *J Neurophysiol* 90:2795–2817.
- Read JC, Cumming BG (2007) Sensors for impossible stimuli may solve the stereo correspondence problem. *Nat Neurosci* 10:1322–1328.
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46:945–956.
- Sakai HM, Naka K, Korenberg MJ (1988) White-noise analysis in visual neuroscience. *Vis Neurosci* 1:287–296.
- Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. *J Vis* 6:484–507.
- Tanabe S, Cumming BG (2008) Mechanisms underlying the transformation of disparity signals from V1 to V2 in the macaque. *J Neurosci* 28:11304–11314.
- Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. *J Neurosci* 22:10811–10818.
- Troyer TW, Krukowski AE, Priebe NJ, Miller KD (1998) Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *J Neurosci* 18:5908–5927.
- Wilson HR, Blake R, Halpern DL (1991) Coarse spatial scales constrain the range of binocular fusion on fine scales. *J Opt Soc Am A* 8:229–236.
- Zhu YD, Qian N (1996) Binocular receptive field models, disparity tuning, and characteristic disparity. *Neural Comput* 8:1611–1641.