The Neural Coding of Stereoscopic Depth

Izumi Ohzawa, Gregory C. DeAngelis*, and Ralph D. Freeman#

Group in Vision Science, School of Optometry,
University of California, Berkeley, California 94720-2020

izumi@pinoko.berkeley.edu
gregd@monkeybiz.stanford.edu
freeman@pinoko.berkeley.edu

* Present address: Department of Neurobiology
Stanford University School of Medicine
Stanford, CA 94305-5401

#Reprint request


Running title: Neural Coding of Stereoscopic Depth
Key words: binocular disparity, stereopsis, binocular vision, receptive field, cat, visual cortex
Abstract
Stereopsis is a process by which the visual system gauges the relative depth of objects in three-dimensional space by measuring minute positional differences between left and right images. According to the standard notion, this information is thought to be encoded in the primary visual cortex by differences in receptive field (RF) positions for the two eyes. We have developed an alternative model by which stereoscopic information is coded and transformed through a hierarchical chain of processing in the primary visual cortex. Initially, first-order neurons of the visual cortex, simple cells, encode depth information by a scheme based on differences in internal receptive field structure between left and right eyes. Further abstraction of information is achieved by a subset of second-order neurons, complex cells, that are well suited for the detection of depth information in a manner unaffected by positional variations of objects. We review physiological evidence from studies of the cat and monkey that are relevant to the proposed scheme.

Background
One of the distinguishing characteristics of animals with frontally positioned eyes is that each of the pair has a slightly different view of visual space. The resulting difference in the images for the two eyes, which is called binocular disparity, is the necessary and sufficient condition for stereoscopic depth perception, as shown clearly from psychophysical measurements with a stereoscope (1). Although monocular cues to depth are numerous and often strong, stereoscopic depth discrimination is an order of magnitude finer than standard visual resolution (e.g., ability to resolve two small dots placed close together as two, instead of one) and it permits perceptual functions that could not otherwise be carried out.

What is the neural foundation of this process? In the original description of the binocular function of the visual cortex, Hubel and Wiesel (2) speculated that cortical cells may be involved in coding depth information but no mechanism was proposed. The first specific proposal for a neural basis of binocular depth discrimination was made by Barlow, Blakemore and Pettigrew (3). Their study suggested that binocular cells responded to different depths in space relative to the fixation point, i.e., different retinal disparities. They also suggested that this coding was specialized for horizontal disparities. These suggestions were based on the finding that there are neurons with receptive fields (RFs) that possess different positional offsets from the corresponding retinal points, and the distribution of preferred disparities was larger for the horizontal dimension than for the vertical. A considerable extension of the idea of depth processing in the visual cortex was made by Poggio and Fischer (4) who studied responses of cortical cells (V1 & V2) in behaving monkeys. They identified four types of responses (near, far, tuned excitatory, and tuned inhibitory) and suggested that relative depth processing could be accomplished with these cells. Several studies of the cat (5-7) and the monkey (8, 9) have pursued this idea.

The details of these investigations have evolved mainly with respect to response type classifications. For example, the original four categories noted above were expanded to six (9). Other studies suggest a continuum of response types (7). In all of the original work, an assumption is made that a position difference in right and left eye RFs constitutes the necessary and sufficient condition for the encoding of binocular disparity. These studies do not include a theoretical framework by which monocular and binocular processes may be integrated and which optimizes neural efficiency and function, as attempted in suitable models (e.g. 10).

We consider this issue in the current review. The idea we have developed is as follows. Stereoscopic depth information is encoded in the visual cortex by a representation that relies in part on differences in the internal structure of simple cell RFs for left and right eyes. This representation, which may be described in terms of differences in spatial phase, enables encoding of the visual cues required for stereoscopic depth detection. We refer to this scheme as the phase encoding model. However, the representation at this simple cell stage is not yet specialized for stereoscopic depth. In fact, the information
encoded by simple cells can be used as a source for many aspects of visual perception. This information is further processed by a subset of complex cells in the visual cortex into a form more suited for the detection of stereoscopic depth.

An alternative approach

The data presented here are from extracellular recordings of responses from individual cells in the cat's visual cortex. Standard physiological procedures were used to maintain animals for extended recording sessions. Spikes were isolated and analyzed in real time by computer so that results of individual runs could be assessed during the study of each cell. These methods have been described in detail previously (11).

A special technique was used to provide details of RF structure. This method, reverse correlation analysis, was developed originally for audition (12) and it has also been applied to vision recently (13). For standard peri-stimulus time histogram (PSTH) analysis, stimulus-response correlation is sought in the forward direction of time by averaging responses in a manner time-locked to the stimulus onset. With the reverse correlation analysis, the direction in which a correlation is sought is reversed. For each spike generated, a correlation with a stimulus is sought up to a few hundred milliseconds prior to the occurrence of the spike. In our case, the stimulus is a sequence of small, briefly presented and optimally oriented bars that are brighter or darker than the background. The bright and dark bars are presented one at a time in a randomized order at all locations within a two-dimensional (2D) grid that covers the RF. The process is repeated with randomized sequences of bars until a clear RF profile is obtained; this typically requires 15-20 minutes. By choosing an appropriate correlation delay for each neuron (typically 50-80 msec), a detailed RF plot is obtained in two dimensions of space (X and Y). The time delay that yields maximum correlation varies from one neuron to the next, and depends on a series of delays introduced at various stages of the visual pathway. By repeating the analysis for different temporal delays, a 3D receptive field is obtained for space and time (X, Y and T). Details of this procedure are given elsewhere (11).

To study binocular properties of cortical neurons, we use a dichoptic version of the reverse correlation method. A long, thin bar-shaped stimulus is presented to each eye simultaneously at different combinations of left and right locations in a randomized sequence of brief flashes. Analysis similar to the monocular method produces a complete profile of binocular interactions, from which we are able to obtain disparity selectivity profiles (14).

Requirements for phase encoding

In order for our proposed scheme to be plausible, cortical neurons must exist for which the internal structure of the receptive fields for left and right eyes is spatially dissimilar. However, this possibility is at odds with the accepted view that left and right fields are uniformly well matched (2, 15). If this is the case, then a preference to non-zero disparity must arise from minute offsets of receptive field positions for the two eyes (5, 16).

Fig. 1 illustrates the difference between disparity encoding according to the classical notion and our phase encoding model. The conventional view, which we refer to as a position encoding model, is depicted in Fig. 1A. Receptive fields for simple cells are represented by adjacent ON and OFF regions filled, respectively, with plus and minus signs. The example here shows an even symmetric RF profile with an OFF-ON-OFF sequence of subregions. The fields are also shown as 3D surface plots (middle). In this case, OFF and ON flanks respectively dip below and rise above the horizontal plane. Response strength is denoted by the height above and the depth below the surface. In these surface plots, the RF profile is modelled as a Gabor function (the product of a sine wave and a Gaussian) (13, 17, 18). Note that RF structure is identical for the left and right eyes. The RFs are offset with respect to each other so that the two RFs are centered at non-corresponding points. The response of this cell to sweeping bar stimuli of convergent, divergent, or zero disparities is depicted in the bottom panel of Fig. 1A. The peak response is obtained for a small convergent disparity. If the RFs were aligned at corresponding points,
zero disparity targets would produce a maximum response. Disparity information is therefore encoded in this case by lateral offset of simple cell RFs with identical internal structure.

Similar depictions are shown in Fig. 1B for the alternative scheme we propose, the phase encoding model (14, 19-23). In this case, the internal structure of left and right RFs differs substantially. The sequence of ON and OFF flanks are different for left and right eyes. For the left eye, the profile is similar to those of Fig. 1A (even symmetry). On the other hand, the RF profile for the right eye has only two major flanks (odd symmetry) that are of the same amplitude but with the opposite sign. Note, however, that the receptive fields are centered at corresponding points in the two eyes, unlike those for Fig. 1A. Since we represent the RF by a Gabor function, it is possible to specify a phase relationship between the left and right RFs. In the case of Fig. 1B, the phase difference between the left and right RFs is 90°. The predicted disparity tuning profile (bottom) is also different from that of Fig. 1A, but the maximum response is obtained for the same convergent disparity in both cases.

---

**Fig. 1**

Two alternative schemes are shown for disparity encoding by simple cells. **(A)** Position encoding: According to the conventional view, a neuron becomes selective to non-zero disparity by having left and right receptive fields (RFs) offset laterally from corresponding points. As shown at the top, the shape of the RF is the same for the two eyes with a sequence of OFF-ON-OFF flanks. However, they are not located at corresponding points indicated by vertical gray lines. The same RFs may be depicted using a 3-dimensional surface plot (middle). The height of the peaks and the depth of the troughs indicate the responsiveness of the cell to a stimulus delivered at each location. A predicted disparity tuning curve for this simple cell is shown at the bottom. The response is largest at a non-zero (convergent or crossed) disparity. **(B)** Phase encoding: Non-zero disparity is encoded by simple cells having different profiles for the left and right RFs. The two RFs differ in phase by 90°. The left eye field in this figure has an OFF-ON-OFF sequence of flanks (even symmetry), while the one for the right eye has an ON-OFF sequence (odd-symmetry). Note that these depictions are idealized for illustration purposes. Real neurons do not necessarily show exact even or odd-symmetric receptive fields nor a left-right phase difference that is a multiple of 90°. The disparity tuning curve (bottom) for this simple cell has a different shape from that of A, but the optimal disparity is the same.
Data relevant to the two models

To distinguish between the two schemes for disparity encoding illustrated in Fig. 1, we require detailed information about the internal structure of the left and right RFs of simple cells. For this purpose, we have used a reverse correlation procedure to demonstrate that for a substantial proportion of simple cells in the visual cortex, left and right eye RF structure is clearly different (20-23). Representative data that lead to this conclusion are illustrated in Fig. 2. We first run monocular tests with drifting sinusoidal gratings to determine preferred values of orientation, and spatial frequency for each eye. Then the degree of tuning for binocular disparity is determined using dichoptically presented drifting gratings with a variety of relative phase offsets for the two eyes (24). Results of this test are shown at the top of Fig. 2A and B. For each cell, PSTHs are shown on the left of the graph. These simple cells exhibit typical response modulation at the temporal frequency of the drifting grating. There is both facilitation and suppression in the responses relative to the monocular response, and this is seen more clearly in the results of the harmonic analysis to the right of each set of histograms. Responses to monocular stimulation of the left (L) and right (R) eyes are indicated by arrows for comparison. For some relative phases, the cells are silent to binocular stimulation even though the stimuli are highly effective monocularly. Maximum facilitation occurs for phases around 180˚ away from the suppressed values for each cell. Note that a phase value of zero does not represent zero disparity, since the exact retinal correspondence is unknown in the anesthetized, paralyzed preparation that we have used. Data points are fit with one cycle of a sinusoid to determine the strength of binocular interaction (24).

For these two cells, a reverse correlation procedure (11, 23) is used to obtain detailed two-dimensional RF maps (see Materials and Methods). In Fig. 2 (bottom), we present results, obtained at the optimal time delay† (60 msec), for left and right eye RFs. Spatially segregated OFF (dark excitatory) and ON (bright excitatory) subregions, characteristic of simple cells, are indicated by regions that are darker or brighter than the background, respectively. (Although the preferred orientations for the cells shown here are different, for ease of presentation, RF profiles are shown upright with elongation of subregions always aligned with vertical.) The internal structure of the RFs shown in Fig. 2A is very similar, both in the appearance of the 2D maps, as well as in the shape of 1D RF profiles shown near the bottom of each 2D map. The difference of RF profiles may be quantified by a phase difference by fitting each 1D RF profile by a Gabor function and calculating the difference of the phase value (22, 23). For the cell shown in Fig. 2A, the phase difference between left and right eyes is nearly zero. On the other hand, RFs for the cell of Fig. 2B are quite dissimilar for the left and right eyes. The order and strength of ON and OFF subregions are different for the two eyes so that there is a phase difference of 103˚, which is close to a quadrature phase difference of 90˚.

Simple cells

Similar experimental tests were performed on a total of 61 simple cells. We find a continuum of phase differences between right and left eyes, with some left and right RFs closely similar and others broadly different (21-23). A substantial proportion of cells possess different internal RF structure for the two eyes. When expressed as phase differences between left and right eye RFs, 33% of cells have phase differences of 45˚ or more. We have also found a relationship between phase differences and the preferred orientations of cortical cells. For cells whose preferred orientations are close to horizontal, phase difference tend to be small whereas cells tuned to near-vertical orientations exhibit a broad range of phase differences. Thus, there is an apparent specialization for processing horizontal retinal disparities such that phase differences between left and right eye RFs are used to encode disparity information (21-23). A

†. Simple cells with space-time inseparable RFs do not have a unique spatial phase for each eye, as the phase changes continuously over the time course of the response (11). This generally does not present a problem for disparity representation because the phase change is closely matched for the two eyes, and therefore the phase difference remains nearly constant for most cells (23).
Fig. 2

Data from two binocular simple cells are shown. (A) Responses are presented from a simple cell that possesses nearly identical receptive fields for the two eyes. At the top, responses to drifting sinusoidal grating stimuli are shown. The grating stimuli are presented to the two eyes simultaneously at a variety of relative phases in 45° steps. Peri-stimulus time histograms (PSTHs) for these conditions are shown on the left. In addition, responses are shown for stimulation of each eye alone (marked as L and R), and for the null condition (N) in which the display is uniformly gray. All stimulus conditions are presented in a randomly interleaved manner. Each histogram shows the result of 8 repetitions. These PSTHs are harmonically analyzed at the temporal frequency of the stimuli (2Hz). On the right, the amplitude of the 2Hz component (first harmonic) is plotted against the relative interocular phase of the gratings. The curve represents a cycle of a sinusoid fitted to the binocular responses (filled circles). Responses to monocular stimuli (L and R) are noted at the right margin of the plot by arrow heads. Two dimensional receptive field profiles for the left and right eyes are obtained by a reverse correlation technique, and shown at the bottom. The dark area represents an OFF (dark excitatory) flank, and the white area corresponds to an ON (bright excitatory) flank. The RF maps cover a square area 5°x 5° in size, and were obtained using a time delay of 60msec. One-dimensional RF profiles, presented at the bottom of each map, are obtained for each eye by integrating the 2D maps along the Y (vertical) dimension. The phase difference between the left and right receptive fields is 0.5°. The cell had preferred orientations of 30° and 15° (horizontal=0°) for left and right eyes, respectively, but the RFs are shown upright here. The bars had the dimension of 2° x 0.5°. The RF for the right eye is slightly shifted to the left. This simply represents inaccurate centering of the stimulus grid, but not RF incongruity (see text). (B) Results are shown from another simple cell that had different RFs for the two eyes. Responses to drifting sinusoidal gratings show a very similar pattern as that for A. The relative phase values at which the cell responds most vigorously are different for A and B. However, this is inconsequential because we had no knowledge of the absolute phase of the stimuli. RF profiles presented at the bottom show a clear structural difference for the left and right eyes. The RF maps cover a square area 5°x 5° in size, and were obtained using a delay of 65 msec. The phase difference between the left and right receptive fields was 103°, which is close to a quadrature relationship (90°). Preferred orientations were 80° and 90° for left and right eyes, respectively. The bars had a dimension of 2° x 0.5°.
specialization for encoding horizontal disparities was reported in the original study, in that the range of preferred disparities was larger for the horizontal dimension than for the vertical (3). However, in subsequent work, horizontal-vertical difference have not been found (5-7). It is difficult to pinpoint a possible source of the discrepancy. Effects of eye movements may have affected the results despite precautions taken in some of the studies. The use of a bright bar stimulus alone and subjective determinations of RF boundaries may have produced inaccurate estimates of RF positions. Also, inconsistent findings may be partly the result of data that are fundamentally mismatched. In one study (3), the distributions of optimal preferred disparities were compared between horizontal and vertical dimensions. In others (5-7), estimates of RF positions were obtained for the two eyes and the resulting incongruities were compared. It is likely that the discrepancy is due to a combination of these factors. In any case, our finding establishes that there is, indeed, such a specialization for horizontal disparity, not in terms of position incongruities of RFs, but in terms of phase differences between left and right RFs.

The necessary condition for disparity information to be encoded by means of internal structural differences (phase differences) between left and right RFs is therefore fulfilled. Our results are consistent with the phase-encoding of retinal disparity but they do not rule out the alternative mechanism in which disparities are encoded by incongruities (i.e. lateral displacements) of RFs. Although the phase encoding scheme has theoretical advantages that we have described previously (20, 23), it has been argued recently that both processes are necessary (25), and both mechanisms may be present (26, 27).

**Complex cells**

The initial coding of disparity information, which appears to occur first in striate cortex, must be followed by subsequent processing and abstraction. We have put forward the notion, based on the phase-encoding scheme, that a subset of complex cells in the visual cortex may be ideally suited to detect the local disparity information encoded by a group of simple cells. These complex cells provide a representation of disparity information that is less dependent on other stimulus parameters than the representation given by simple cells (14). Considered at a monocular level, complex cells do not appear to exhibit characteristics that are suitable for fine depth encoding. A properly oriented stimulus anywhere within the RF elicits a response. Furthermore, the stimulus may be brighter or darker than the background luminance at each position within the RF, so the sign of contrast is not relevant to the cell's response selectivity.

In order for a complex cell to be useful as a disparity detector, at least three binocular characteristics are required. First, the RFs of complex cells are generally large. If disparity tuning is determined by the size of the entire field alone, only relatively crude disparity selectivity is possible. Therefore, disparity selectivity must be considerably finer than the position selectivity of the RF (28). Second, different stimulus positions within the RF must produce constant preferred disparity signals. This property is required for the cell to be a disparity detector which is not sensitive to variations in position within the RFs. Third, contrast information in the left and right eyes should be matched at the optimal disparity. There should not be a response to opposite contrast targets in the two eyes (i.e. a dark bar in one and a bright bar in the other) at the optimal disparity, because that combination should not signal a match of corresponding image features. We have conducted extensive studies of complex cells in order to determine if the above characteristics are exhibited. We find that a substantial number, at least 30% of the complex cells we studied, exhibit characteristics which make them ideally suited as local disparity detectors (14).

Examples of results of binocular tests for complex cells are shown in Fig. 3. Responses of complex cells are characterized by an increased overall discharge to drifting sinusoidal gratings, as shown in the PSTH data of Fig. 3. The cell in Fig. 3A is dominated by the right eye but stimulation through either eye elicits a vigorous response. However, for some relative interocular phases, from 120° to 180°, response is severely suppressed. On the other hand, for phases that are 180° away, there is substantial facilitation. This pattern is seen more clearly in the plot of response amplitude, and is similar to those observed for simple cells (Fig. 2). We have shown previously that this phase-specific binocular interac-
An additional binocular test was conducted with these cells. A reverse correlation sequence, as described above, was run in a binocular mode. In this case, sequences of bright and dark bars (15° by 0.5°) of optimal orientation were flashed dichoptically at various combinations of positions throughout the RFs. It was then possible to determine which disparities elicited the strongest responses at various positions within the RF. The result of this experiment is shown as density plots beneath the graph of Fig. 3.

Data are presented from binocular experiments on two complex cells. (A) Results are shown for a disparity selective complex cell. At the top, responses to dichoptically presented sinusoidal gratings are given in the same format as that of Fig. 2. Here, the relative phase of the gratings is varied in steps of 30°. The PSTHs show unmodulated discharge typical of complex cell responses to drifting sinusoidal gratings. As for the simple cells of Fig. 2, binocular responses show clear facilitation and suppression as relative phase is varied. At the bottom, response maps obtained by a binocular version of the reverse correlation method are presented. Each domain plots responsiveness of the neuron to a particular combination of left and right stimulus positions. The horizontal and vertical axes represent left and right stimulus positions, respectively. Since a bar can be either bright or dark for each eye, there are 4 possible permutations: dark bars to both eyes (Dark-Dark), bright bars to both eyes (Bright-Bright), and two cases where one eye received a dark bar while the other was shown a bright bar. This cell shows an elongated diagonal region ideally suited for disparity detection in the case of matched contrast stimuli (Dark-Dark and Bright-Bright). Each binocular domain spans 6° of visual angle for both the left and right eyes. (B) Results are presented for a non-disparity selective complex cell. The PSTHs and the plot of responses against relative phase show vigorous responses for all interocular phase values. As expected from the grating responses, binocular response maps obtained by the reverse correlation method also do not show disparity selectivity, i.e., no diagonally elongated response region is observed. The cell simply responds well if either the left or the right stimulus, or both, are placed near the center of the receptive field of the respective eye. This results in a shaded area in the shape of a "plus" sign. A facilitatory effect is observed in the Dark-Dark map in that the crosspoint of the "plus" is darker than its arms.
3A. There are four domains, each representing one of four conditions of stimulus presentation. The conditions are: a dark bar to each eye, a bright bar to each eye, a bright bar to the left eye and a dark bar to the right eye (or the reverse). Horizontal and vertical axes represent stimulus positions for the left and right eye, respectively. As in the case of the figures that follow, the darker the pixels, the stronger the response of the cell. For the cell of Fig. 3A, a clear pattern of response can be seen. Pairs of stimuli with the same polarity (two dark or two bright bars) elicit maximum responses. Furthermore, responses occur only along a diagonal within the region of joint left and right RF space.

Note that maximum responses for this cell occur along a diagonal line of slope 1, i.e., for the same disparity value independent of left and right eye stimulus positions. In addition, maximum responses occur only when polarities of the stimuli are matched for the left and right eyes, i.e., for dark-dark or bright-bright combinations. For unmatched polarities, i.e., dark-bright or bright-dark combinations, there is a distinct lack of response along the same diagonal, although there are moderately responsive regions on either side. Finally, the width of the response region along the diagonal is considerably smaller than the overall RF. From similar observations in a population of cells, we conclude that there is a sub-population of complex cells that exhibit response characteristics, as outlined above, which make them suited as ideal disparity detectors. These cells may detect disparity information encoded by an organized subset of simple cells that share a common preferred disparity. Presumably, further encoding and abstraction would then be provided by the next stage of disparity processing that examines the output of multiple complex cells like the one presented in Fig. 3A.

An additional example of a complex cell is shown in Fig. 3B. For this complex cell, overall discharge is relatively independent of the interocular phase of dichoptically presented sinusoidal grating stimuli (top). The implication is that this cell is not involved in the processing of disparity information. If this is correct, we should not obtain response patterns similar to those seen in Fig. 3A from binocular reverse correlation tests. The data of Fig. 3B confirm this expectation. There are no obvious diagonal patterns of response for any of the four domains. Instead, the general pattern of response is cross-shaped, with a tendency for maximum responses near the center of the cross. This applies for all permutations of bright and dark bar stimulus pairs. This is the type of response predicted by a simple summation of monocular responses of complex cells. In other words, there is no obvious specialization for binocular disparity encoding or detection. Responses are strongest in the region of the center of the RF, both in monocular and binocular conditions. Cells such as this appear clearly uninvolved in disparity processing.

In conjunction with the physiological findings presented in Fig. 3, we have developed a model that accounts for the physiological data. It assumes a hierarchical processing sequence in which a minimum of four simple cells provide input to a single complex cell. The simple cells are arranged in two pairs, and the pairs are arranged in quadrature phase, i.e., the spatial phase of the pairs differs by 90° (14, 30, 31). Predictions of binocular responses based on both the phase and position models, together with responses of cells, are presented below.

The disparity sensitive complex cell shown in Fig. 3A has an even-symmetric profile with respect to change in disparity, i.e., along cross sections with a slope of -1 which are perpendicular to the diagonal response region. Predictions of a disparity energy model‡ show that this type of disparity tuning can result from a combination of 4 or more simple cell subunits with the same left and right RF profiles, as shown in Fig. 4A (14). We should note that our model of disparity-sensitive complex cells employs an organization of subunits analogous to that originally used in motion energy models (32-35). Disparity tuning of the model, obtained by integrating the 2D profile, indicates a function that is clearly even-symmetric with respect to zero disparity (Fig. 4A). However, for the zero-disparity detector illustrated in Fig.

‡. The term energy model is used to denote the fact that the output of simple cell subunits (in quadrature relationship) are squared and then summed, in accordance with the formal definition of energy. This computation provides a signal that is proportional to strength or energy of stimuli at the cell's preferred disparity. The signal is localized both in space and spatial frequency, because a neuron has a RF and a limited spatial frequency pass-band.
Fig. 4

Predicted responses are shown of disparity energy detectors and data from a complex cell. (A) Predicted response is shown of an ideal zero disparity detector with even-symmetric disparity selectivity profile in the same format as the bottom panels of Fig. 3. The phase and position models behave identically for the zero disparity detector. The diagonal line indicates a disparity of zero. Dashed vertical and horizontal lines indicate the peak of monocular excitation profiles which are Gaussian for the left and right eyes, respectively. Note that the horizontal and vertical cross sections of the 2D profile near the margins give monocular excitation profiles for the left and right eyes, respectively, because the stimulus for one eye is far outside of the other eye's receptive field near the margins. The disparity tuning curve (at the lower left of A) is obtained by integrating the 2D profile along the 45° constant disparity lines. As shown on the right, the detector combines the output of 4 simple cell subunits (numbered 1-4) that have identical left and right RFs which are centered at corresponding points in the two eyes. Positive and negative parts of the subunit RFs are shown by filled and dashed curves, respectively, to indicate the fact that simple cell subunits cannot signal inhibition directly because they usually have minimal spontaneous discharge. The vertical lines going through the subunit RFs depict the centers of RFs and their positions correspond to those of the horizontal and vertical dashed lines in the 2D profile. (B) Predicted response is shown in the same format as that of A, of a non-zero (convergent) disparity energy detector based on the position encoding model. Positional offset of all 4 subunits produces a shift of optimal disparity from zero (solid diagonal line). The maximum binocular response is obtained when the optimal monocular stimuli (horizontal dashed lines) are combined, as in the case of A. The subunit organization shown on the right indicates that all right RFs are positionally offset from corresponding points. (C) Predicted response is presented of a convergent disparity detector based on the phase encoding model. This detector, which has an odd-symmetric disparity tuning profile, combines the output of 4 simple cell subunits that possess left and right RFs which differ in phase by 90°. Horizontal dashed lines indicating the peak of monocular excitation no longer cross at the peak of binocular excitation. (D) Data are shown from a complex cell that has a disparity tuning profile similar to the prediction in C. No solid diagonal line is drawn since zero disparity is not known for the experimental data. Preferred orientations of this cell are 90° and 80° for left and right eyes, respectively. The size of the domain is 8° by 8°. Size of the bar stimulus was 20’ x 0.4’. 
there is no difference between the phase and position models because no phase difference or position incongruity are present. Therefore, the existence of cells that show an even-symmetric binocular response pattern as that of Fig. 3A is consistent with both the phase and the position model. In fact, Fig. 4B illustrates a prediction of binocular responses for a non-zero disparity detector based on the position model, and the response pattern is the same as that of Fig. 4A except for an overall downward shift. Since we do not have exact knowledge of retinal correspondence in our preparation, we do not know if the cell of Fig. 3A is a zero disparity detector constructed on the phase encoding scheme (Fig. 4A), or if it is a detector for any disparity constructed on the position encoding model (Fig. 4B). Therefore, for the phase model to be applicable for complex cells as well as simple cells, there must be complex cells that combine simple cell subunits with different RFs for the two eyes. In other words, there should be disparity detectors that combine the output of subunits with different left and right RF profiles. The predicted response of such a complex cell is given in Fig. 4C. A notable difference between this profile and those in Fig. 4A and B is that the disparity tuning curve is odd-symmetric as demonstrated clearly by the curve shown at the bottom left of Fig. 4C. In addition, Fig. 4C shows that the optimal locations of monocular stimuli, indicated as dashed lines within each square, do not intersect at the position which yields the maximum binocular response. The stimulus position for the maximum binocular response is slightly shifted to the right and below the intersection of the dashed lines. This is not the case when the subunit RF profiles are identical for the two eyes (Figs. 4A and B). In our sample of complex cells, we find neurons that show approximately odd-symmetric disparity tuning profiles. Fig. 4D presents results from such a neuron. The disparity tuning profile of this cell is approximately odd-symmetric, and the maximum binocular response is obtained when stimuli in the two eyes are positioned slightly off the respective peaks of the monocular excitation profiles (see dashed lines). The presence of these cells demonstrates that complex cells are able to take advantage of disparity information encoded by simple cell subunits that possess a large phase difference between left and right RFs.

Complex cells presented in Figs. 3A and 4D function as local disparity detectors that are insensitive to the sign of contrast and small variations of stimulus positions. These characteristics are unique to a subset of complex cells and are not present in responses of simple cells (14). The insensitivity to the sign of contrast, of course, causes a loss of information, but it is beneficial for a system that tries to determine the disparity of a local region of the visual scene. Natural scenes are almost always defined by objects and contours with bright and dark areas (and various shades of grays and colors). Therefore, if both signs of contrast are present within the RFs of a disparity detector, contributions from the two should reinforce each other, rather than cancel. The complex cells appear to be designed precisely to perform such functions. A similar insensitivity to the sign of contrast is also achieved for another role of complex cells, i.e., as motion energy sensors. In this case, the direction preference of complex cells remains constant regardless of the sign of contrast (32, 33). Because these roles of complex cells do not appear to be mutually exclusive, a complete description of complex cell RFs almost certainly requires a model that incorporates both disparity and motion processing (30, 36).

Conclusion

In summary, we propose that retinal disparity information is encoded by differences in the internal structure of RFs of left and right eyes. These differences, which we have confirmed to be present in a large proportion of simple cells in the visual cortex, may be described quantitatively in terms of phase differences. Although we have found strong evidence that supports the phase encoding model, we are not able to rule out the possibility that an additional mechanism based on RF incongruities is also involved, because we did not measure the positions of left and right RFs with respect to retinal corresponding points. These locations are difficult to determine in an anesthetized paralyzed preparation. Regardless of whether one or both mechanisms are used, the next stage in processing appears to be accomplished by a subset of complex cells with specific characteristics. We have confirmed the presence of these complex cells and believe that they enable detection of local binocular disparity with high sensitivity while main-
I. Ohzawa, G.C. DeAngelis, & R.D. Freeman --- Neural Coding of Stereoscopic Depth

...insensitivity to the variation of other stimulus parameters. Complex cells appear to gain these properties by combining the output of specific arrangements of simple cells that share a common preferred disparity. The information at the complex cell level must be further processed at the next stage (30, 31, 36), as the disparity signal that each complex cell carries is local to its receptive field and is limited to the pass-band of its spatial frequency tuning. Although the relative roles of phase and incongruency encoding remain to be determined, we believe the system outlined here may be primary in the neural processing of retinal disparity.

Acknowledgement

This work was supported by research and CORE grants EY01175 and EY03176 from the National Eye Institute and by a collaborative project of the Human Frontier Science Program.

References