Quantitative Theories of the Integrative Action of the Retina

Norma Graham          Floyd Ratliff
COLUMBIA UNIVERSITY    THE ROCKEFELLER UNIVERSITY

1. INTRODUCTION

One of the great advantages of the use of mathematics in science is that it encompasses equally well both the concrete and the abstract. The mere counting of a number of tangible objects along a single dimension is as much a proper part of the mathematical world as is the formulation of an abstract multidimensional concept that may exist only in thought. Thus, since the purely mathematical aspects of both the real and the imaginary are of one and the same kind, there is no barrier to prevent transitions from the one to the other. The major purpose of this survey of quantitative studies of neurophysiology of vision is to illustrate how easy transitions from empirical data to abstract ideas and back again, via mathematics, provide a most important tool for the investigator—a ready means for the continual interplay of evolving theory and experimental advances.

This survey focuses on the retina, which—as its name implies—is a network. The functional units in it are interconnected and interact with one another; the neural activity generated by illumination of any one photoreceptor may influence, or be influenced by, the activity generated by many others. Consequently, many visual phenomena may depend as much upon the properties of the network of interconnected retinal neurons, acting as a
whole, as they do upon the properties of the individual components, acting alone.

This vital and organic activity of the retina may be summarized briefly as follows: Light absorbed by photoreceptors stimulates them to generate neural activity in the form of graded local excitatory and inhibitory influences. These opposed influences are then integrated, over both space and time, by the neural networks in the retina. By means of this simple calculus the retina abstracts information about significant features of the spatial and temporal pattern of light and shade on the receptor mosaic. Ultimately, these abstractions are expressed in a neural code signaled by a pattern of discrete identical nerve impulses that are transmitted along the optic nerve to the brain. This set of abstractions and transformations may, quite properly, be called the ‘logic’ of the retina. The fundamental process involved—the integration (summation) of opposed excitatory (positive) and inhibitory (negative) influences—is amenable to a rigorous mathematical treatment. And it is the purpose of this chapter to review the quantitative theories of the integrative action of the retina that have been advanced. The review is confined to the retinas of the three animals on which most of the quantitative experimental and theoretical work has been done—the horseshoe crab (Limulus), the cat, and the goldfish.

For the relatively simple eye of Limulus the underlying physiological processes are fairly well understood and the theory well developed; accurate predictions can now be made of the Limulus optic nerve responses to almost any spatial and temporal pattern of illumination on the retina. For the more complex eyes of the cat and goldfish, the underlying processes are less well known and the theories less well developed, but a more complete understanding of the physiological mechanisms and more general theories are now beginning to emerge.

1.1. Historical Background

The basic idea that the retina has a certain ‘logic’ of its own based on the interplay of opposed excitatory and inhibitory influences is not new. Truly scientific work on the subject began well over 150 years ago, and the quantitative approach to the problem is now more than 100 years old. A glimpse of this early work is provided by the following brief mention of the contributions of four giants in the field: the German poet and author Johann Wolfgang von Goethe (1749–1832), the French chemist Michel Eugene Chevreul (1786–1889), the German physician and physiologist Ewald Hering (1834–1918), and the Austrian physicist-philosopher-psychologist Ernst Mach (1838–1916).¹

¹This historical review was adapted from Ratliff (1973).
Goethe’s *Outline of a Color Theory*, published in 1808, seemed to contradict at every point the well-established and widely accepted Newtonian theory of light and, being a mere poet, he was severely criticized by scientists from all sides. But much of the confusion about and criticism of Goethe’s work resulted from the failure of scientists, both then and now, to understand and to appreciate the distinction between the physical and the psychophysiological aspects of color. Many of Goethe’s conceptions about what he called “physiological colors” were well founded on psychophysical experiments on brightness contrast and color contrast, and he was perhaps the first to state explicitly the fundamental ideas of the interaction of opposed influences in the retina that are to be elaborated in this chapter. He wrote: “When darkness is presented to the eye it demands brightness, and *vice versa*; it shows its vital energy, its fitness to receive the impression of the object, precisely by spontaneously tending to an opposite state.” And, “If a colored object impinges on one part of the retina, the remaining portion has a tendency to produce the compensatory color; . . . Thus yellow demands purple; orange, blue; red, green; and *vice versa*. Thus again all intermediate gradations reciprocally evoke each other; the simpler color demanding the compound, and *vice versa*.”

No matter how far from the mark Goethe may have been in his views on the physical aspects of light, it is now evident that his concept of a simultaneous reciprocal interaction of opposed or complementary influences among neighboring parts of the retina is basic to the understanding of the psychophysiology of the visual system. For a translation of and commentary on Goethe’s work on color see Matthaei (1971).

Chevreul, whose work on fats and organic analysis opened up two major fields of research in biochemistry, also made important early contributions to the psychophysics of vision. When he was called upon, in 1824, to become director of the dye plant at the Gobelins Tapestry Works in Paris there had been numerous complaints about the quality of certain pigments prepared there. These led him to undertake a study of the psychology of color. He soon discovered that many of the complaints about the pigments themselves were unfounded; some of the supposed defects were entirely due to subjective contrast effects. The evident simultaneity and reciprocity of the contrast of contiguous colors and contiguous shades led Chevreul to attempt to represent the phenomena in quasi-mathematical terms—rather like a set of simultaneous equations—which would express all the facts succinctly. These findings were published in 1839 in his great work on *The principles of harmony and contrast of colors and their application to the arts*. So important is this work for artists and artisans that it is still a standard reference—reprinted as recently as 1967.

To Hering belongs much of the credit for recognizing that reciprocal interactions in the retina underly our normal everyday visual experience as well
as the so-called illusions and distortions such as border contrast and color contrast. As he put it, “The most important consequences of reciprocal interactions are not at all those expressed in contrast phenomena, that is, in the alleged false seeing of the ‘real’ colors of objects. On the contrary, it is precisely the so-called correct seeing of these colors that depends in its very essence on such reciprocal interactions. . . . A closer familiarity with the consequences of this reciprocal interaction is essential for understanding the nature of our vision. . . .”

Hering’s extremely important hypothesis about the role of reciprocal interactions in normal color vision, although long supported by psychophysical evidence, has only recently been vindicated by direct electrophysiological experiments. The classical Young-Helmholtz trichromatic theory must be extended, as Hering’s work indicated, to include excitatory and inhibitory interactions among various pairs of the three basic cone types with their three different sensitivities (maxima in the red, green, and blue) to the visible spectrum. For a recent translation of his major work see Hering (1964).

Mach, while experimenting with rotating discs used to produce various spatial distributions of light and shade, found light and dark bands appearing where, according to physical calculations, none were expected. Formerly, such contrast phenomena as these Mach bands—as they are now called—had generally been attributed to ‘unconscious inferences’ or ‘errors of judgment.’ But for Mach these were merely various ways of expressing the still unexplained facts. He sought an explanation in terms of the interdependence of neighboring points on the retina. As he expressed it:

Since every retinal point perceives itself, so to speak, as above or below the average of its neighbors, there results a characteristic type of perception. Whatever is near the mean of the surroundings becomes effaced, whatever is above or below is disproportionately brought into prominence. One could say that the retina schematizes and caricatures. The teleological significance of this process is clear in itself. It is an analog of abstraction and of the formation of concepts.

Mach’s quantitative analysis of the Mach bands and related contrast phenomena (for translations see Ratliff, 1965) was probably the first attempt to express the integrative action of the nervous system in precise mathematical terms. But his application of mathematical modes of thought to the study of the nervous system (in 1865!) was so far ahead of the times that his papers on the subject attracted little attention when first published. Indeed, the general concept of a simultaneous reciprocal interaction of opposing or complementary influences among neighboring parts of the retina, formulated more or less independently by Goethe, Chevreul, Hering, and Mach, was not widely accepted in their own time. And for many years thereafter psychophysicists generally regarded the retina as a mere passive and mechanical transducer rather than the vital integrative organ that it is.
1.2. Modern Unit Analyses of Retinal Activity

C. S. Sherrington’s researches into the physiology of the central nervous system laid the foundation of modern neuropathology. His *Integrative Action of the Nervous System* (1906), although based largely on reflex action, put forth in a systematic way practically all of the basic concepts of the functional organization of the nervous system that we know today. In particular, his carefully worked out concepts of central excitatory states and central inhibitory states are basic to modern interpretations. It was fundamental in Sherrington’s teaching that these excitatory and inhibitory states, both capable of gradation, were simply of opposite sign and would add up algebraically. It is interesting to note that some of these ideas of Sherrington’s about interaction in the nervous system came in part from his own little known psychophysical studies entitled “On reciprocal action in the retina as studied by means of some rotating discs” (1897) and in part indirectly from Mach by way of McDougall’s (1903) studies on perceptual contrast phenomena.

But the most important single factor that advanced the theory of the integrative action of the retina, and of the visual system in general, was the relatively recent development of techniques for directly measuring the activity of the visual neurons. Less than half a century ago Adrian and Matthews (1927a, 1927b) reported the first successful recording of the electrical activity of the whole optic nerve. Even in this whole nerve (of the conger eel), consisting of a very large number of axons of retinal ganglion cells, they could discern some evidence of excitatory and inhibitory interactions within the retina (Adrian & Matthews, 1928). Precise quantitative studies became possible soon thereafter when Hartline and Graham (1932) dissected single axons apart from the optic nerve of the marine arthropod, *Limulus* (the horsehoe crab), and recorded their unitary activity. Within a few years these unit analyses were extended to the retinas of vertebrates (Hartline, 1938a) and molluscs (Hartline, 1938b). Since Hartline’s initial studies, literally thousands of qualitative and quantitative experiments have been carried out on single neurons in a wide variety of animals and at all levels of the visual system—from the photoreceptors to the brain. Since the optic nerve forms a ‘bottleneck’ through which all visual information must pass as it is transmitted from the retina to the brain, any meaningful analysis of retinal function must concern itself with the patterns of nerve impulses conducted by it.

The most striking feature of the activity of the individual axons that make up an optic nerve is that it appears in the form of discrete electrochemical pulses, each one very much like all the others. Thus the primary data in experiments on these neurons are the time intervals between impulses. Once measured, these data may be converted into other forms, typically into *rate*. For individual neurons in the optic nerve of *Limulus*, where the interimpulse
variability is not great, a useful measure—especially in studies of the dynamics—is the instantaneous rate, the computation of which is illustrated in Figure 1. For any particular interval, the reciprocal of the interval (i.e., the impulse rate over that period) is assigned to all the time between the beginning and the end of the interval. The instantaneous rate thus forms a piecewise continuous function with the mathematical advantage that it can be manipulated in the same way as the other continuous functions (light stimuli, underlying cellular processes of excitation and inhibition) to which it is related. For preparations in which interimpulse variability is large and therefore the instantaneous rate fluctuates widely, as in the cat and goldfish, a useful substitute for instantaneous rate is the impulse histogram. The histogram is constructed by counting impulses within successive equal time intervals (Fig. 1). For steady-state responses, in any preparation, simple rate measures over a long period of any arbitrary length are commonly used.

2. EXCITATION AND INHIBITION IN
THE EYE OF LIMULUS

2.1. The Steady State

The earliest records of optic nerve activity in the compound lateral eye of Limulus showed that, in general, the more intense the illumination, the higher the instantaneous rate. There are no excitatory interactions in this eye. Activity can be elicited from an optic nerve fiber only by illuminating the receptor unit (ommatidium) from which it arises. However, during the course of the experiments, it was casually observed that turning on the room lights often diminished the activity in an optic nerve fiber that was being prepared for study (Hartline, 1949). Once this observation was considered seriously, it was easy to demonstrate that shading regions of the eye neighboring the receptor unit under study restored the receptor’s activity. Evidently activating one region of the eye inhibited the activity of neighboring regions (for review, see Hartline & Ratliff, 1972). Qualitatively, experiments showed that the amount of inhibition depended on several variables: The inhibition was found to be greater, the greater the intensity of illumination on the neighbors, the larger the area of the neighboring region activated, or the smaller the distance between the activated neighbors and the receptor unit under study (Hartline, Wagner, & Ratliff, 1956). But first attempts at expressing the relationship between the amount of inhibition and these variables in some simple quantitative way were not successful. For example, plotting decrement in a test receptor’s firing rate as a function of the intensity of illumination on the neighbors yields a complicated curvilinear function.
FIGURE 1.

Center: An oscillographic record of a train of impulses discharged in a single optic nerve fiber arising from a receptor unit (ommatidium) in the compound eye of Limulus. Time marked in 0.1 sec. Period of illumination marked by blackening of white line above time marks. Above: Plot of 'instantaneous' rate of discharge constructed by plotting the reciprocal of each time interval between impulses over that same interval. The latent period (interval between onset of illumination and discharge of first impulse) is plotted similarly. Below: Impulse rate histogram constructed by counting the number of impulses within successive equal time periods (each 0.1 sec) and plotting the rate over the corresponding interval. For short counting periods or for low rates, fractional intervals between impulses should be considered. (See Lange, Hartline, & Ramer, 1956b.)
An important and useful simplification resulted when it was considered that the inhibition might be recurrent—that is, the amount of inhibition exerted on the receptor under study by a particular neighbor might depend on the firing rate of that neighbor rather than on the illumination stimulating that neighbor. (This was suggested by an apparently imperfect summation of inhibition and by the phenomenon of disinhibition—to be discussed later.) This possibility was directly tested by recording simultaneously from two interacting receptor units (see schema in Fig. 2) as the intensity stimulating

**Figure 2.** Schema of mutual inhibitory interaction of two neighboring receptor units in the compound eye of *Limulus*. The excitation e is measured by the response r when one unit alone is illuminated. The threshold frequency that must be exceeded before a receptor unit can exert any inhibition on its neighbor is represented by \( r^0 \). It and the inhibitory coefficient \( k \) are labeled to indicate the direction of the action: \( r^0_{1,2} \) is the frequency of receptor 2 at which it begins to inhibit receptor 1; \( r^0_{2,1} \) is the reverse. In the same way, \( k_{1,2} \) is the coefficient of the inhibitory action of receptor 2 on 1; \( k_{2,1} \) the reverse. It is necessary to specify the direction of the action because the mutual influences are seldom identical, although generally they are similar. Note that the inhibition is recurrent, that is, each element exerts inhibition back on the site of impulse generation in the other; inhibition of one receptor unit depends on the response r of the other rather than on the excitation e. In the actual retina of *Limulus* the interactions are not confined to adjacent neighbors, the inhibitory influences extend over a considerable field. (See Fig. 3.) (Redrawn after Ratliff, 1968.)

The decrement in the steady-state firing rate of one of the receptors turned out to depend linearly on the steady-state firing rate of the other, once a threshold had been reached. This may be expressed in terms of a pair of simultaneous piecewise linear equations:

\[
\begin{align*}
r_1 &= e_1 - k_{1,2}(r_2 - r^0_{1,2}), \\
r_2 &= e_2 - k_{2,1}(r_1 - r^0_{2,1}),
\end{align*}
\]

where \( r_1 \) is the response (rate of discharge of impulses) of receptor unit 1, \( e_1 \) is the 'excitation' of receptor unit 1 expressed as discharge rate in absence
of inhibition from neighbors. In the first equation \( k_{1,2} \) is the coefficient of the inhibitory action of unit 2 on 1, and \( r_{1,2}^p \) the threshold of that inhibition. In the second equation, the inhibitory action of 1 on 2 is expressed similarly. Note that there are certain limitations to these equations: there can be no negative rates of discharge, there is some upper bound to the rate, and whenever the discharge rate of the neighbor \( (r_e \) in the first equation) is less than the threshold of inhibition \( (r_{1,2}^p) \) the term in parentheses is taken to be equal to zero.

To extend the quantitative description to large numbers of receptor units, we must know how the inhibitory influences combine with one another. It has been shown that the inhibitory influences from two groups of receptor units, which are widely separated so that the two groups do not interact with one another, combine by simple addition when acting simultaneously on an intermediate receptor ([Hartline & Ratliff, 1958]). Consequently, the activity of \( n \) interacting receptors may be described by a set of simultaneous piecewise linear equations, each with \( n \) inhibitory terms combined by simple addition

\[
r_p = e_p - \sum_{j=1}^{n} k_{p,j} (r_j - r_{p,j}).
\]

Notice that there is a term where \( p = j \) or, in short, where the receptor unit \( p \) is inhibiting itself (see Stevens, 1964). This self-inhibition term can be disregarded in the present discussion of the steady state, but it will turn out to be important when we come to the underlying physiology of the *Limulus* receptor unit and to the dynamics of the inhibition.

The above equations contain no explicit terms for distance. No special terms are required; the effects of distance are already implicit in the equations for they are expressed in the concomitant changes in threshold \( (r^p) \) and inhibitory coefficient \( (k) \). In general, the threshold increases with distance and the inhibitory coefficient decreases. The exact form of the distribution of the inhibitory coefficients, which has been determined by recording from some 20 to 30 fibers, can best be described as a broad Gaussian curve (a broad normal distribution) from which a concentric narrow Gaussian curve (a narrow normal distribution) has been subtracted (Fig. 3). That is, it rises to a maximum some distance away from the point of excitation and then gradually diminishes to zero (Barlow, 1967). This dependence of the mutual inhibition among receptors on distance and the recurrent nature of the inhibition are, of course, most significant factors in determining the responses to various configurations of illumination. Following are two examples.

The inhibitory effect produced by the combined influences of several groups of receptors is sometimes less than the effect produced by one group—an apparent contradiction of the law of spatial summation expressed in Equation 2 above. But this results only because the inhibitory influence exerted by a particular unit depends upon its own activity \( r \) rather than directly upon the
light stimulus incident upon it—that is, because the inhibition in *Limulus* is recurrent (feeding back upon neighbors) rather than nonrecurrent (feeding forward upon neighbors). Since the amount of activity in each of several groups illuminated together may be less than when illuminated separately (due to mutual recurrent inhibition and inhibitory thresholds), their combined simultaneous influences may not necessarily equal the sum of their separate influences.

The responses to one particular configuration of illumination illustrate this clearly and demonstrate, at the same time, the phenomenon of disinhibition (Harline & Ratliff, 1957). When, in the vicinity of a mutually inhibiting pair of receptors ($A$ and $B$), additional receptors located so that they inhibit only $B$ are illuminated, the firing rate of $A$ increases. This apparent facilitation
of A actually results from inhibition of inhibition—the additional illuminated receptors inhibit B which therefore inhibits A less than before. (A supposed facilitation that was observed in the spinal cord has since been shown to be another example of disinhibition. See Wilson, Diecke, & Talbot, 1960; Wilson & Burgess, 1961.)

So-called border contrast effects and the Mach bands, well known in human vision, also appear to depend upon the spatial distribution of lateral inhibition. One can see the Mach bands at the borders of a half-shadow cast by an object in good sunlight, such as the shadow of one's own head and shoulders on a sidewalk. The objective distribution of illumination is high and uniform in the full sunlight, more or less uniformly graded in the half shadow, and uniformly low in the full shadow. Objectively there are no maxima and minima, yet there appears to be a narrow dark band at the dark edge of the half shadow and a narrow bright band at the bright edge. Analogous effects have been observed in the lateral eye of Limulus (Ratliff & Hartline, 1959), and are implicit in Equation 2. Because the inhibition diminishes with distance, the difference between the responses of neighboring receptor units is greatest at or near the boundaries of differently illuminated regions (Fig. 4). A unit within a dimly illuminated region, but near the boundary, is inhibited not only by its dimly illuminated neighbors, but also by some brightly illuminated ones. The total inhibition on this receptor unit, therefore, will be greater than on the other dimly illuminated units farther from the boundary. Consequently, its response will be less than theirs. Conversely, a unit within the brightly illuminated area, but also near the boundary, will have a higher frequency of discharge than other brightly illuminated units located far from the boundary, since they are receiving strong inhibition from brightly illuminated neighbors all around, while the one near the boundary receives some strong inhibition from brightly illuminated neighbors and some weak inhibition from dimly illuminated neighbors. Consequently, its response is the greatest of all. Thus maxima and minima appear in the neural response, even though there are none in the spatial distribution of illumination.

2.2. Physiological Mechanisms of Excitation and Inhibition

Up to this point the activity of the Limulus eye has been described with few references to the physiological processes intervening between the light stimulus and the discharge of impulses in the optic nerve. Neither the original formulation nor the description of the abstract model embodied in Equation 2 required knowledge of these processes. Now, however, the physiology is discussed in some detail, for not only is it interesting in itself but the investigation
of it led to the development of a more exact and more general model of the Limulus eye encompassing both the steady state and the dynamics.

When Purple and Dodge (Purple, 1964; Purple & Dodge, 1965) set out to study the cellular mechanisms of excitation and inhibition in the Limulus eye, they could form working hypotheses based on both the large body of knowledge accumulated about the cellular mechanisms of similar processes in other kinds of neurons and the information already known about the Limulus receptor unit (ommatidium).

It was well known that the membrane of a nerve cell plays a crucial part in all its activity. (See Katz, 1966, for a thorough presentation of the material)
that is briefly summarized in the following paragraphs.) When the nerve is
'resting' there is a constant voltage difference between the outside and the
inside of the cell— the 'resting potential'— which occurs because the mem-
brane is selectively permeable to different ions. In general, excitatory influ-
ences affect the cell by increasing the permeability of the membrane to certain
ions and thereby drive the potential difference across the cell membrane
toward zero (that is, excitatory influences depolarize the cell). A large enough
depolarization of the membrane will trigger an impulse and, if the depolariza-
tion is maintained, a train of impulses. Inhibitory influences affect the cell
similarly by increasing the permeability of the membrane to some ions, but
in this case the ions involved cause the potential difference across the cell
membrane to be driven farther from zero (that is, inhibitory influences hyper-
polarize the cell). Thus, inhibitory influences hinder the triggering of a nerve
impulse.

The action of excitatory and inhibitory influences can be modeled as the
'equivalent electric circuit' shown in Figure 5. $E_r$ is the resting potential, $E_e$
is the potential towards which excitatory influences push the cell, and $E_i$
is the potential towards which inhibitory influences push the cell. $E$ is the result-
ing potential across the membrane (the so-called membrane potential). $V$,
the difference between this resulting potential $E$ and the resting potential $E_r$,
is called the 'generator potential.' Several resistances represent the membrane's
selective permeability to various ions. $R$, a variable resistance, reflects the
amount of excitatory influence. The greater the excitatory influence, the more
permeable the membrane is to those ions associated with excitation, and thus
the lower the resistance $R$, and the higher the excitatory conductance
$g_e = 1/R$. Similarly, the greater the inhibitory influence, the more permeable
the membrane is to those ions associated with inhibition, and thus the lower
the corresponding resistance $R$, and the higher the inhibitory conductance
$g_i = 1/R$. $R$ is fixed and equals the resistance of the membrane in its resting
state. The capacitance of the membrane will be ignored because of its rela-
tively short time constant (see Purp, 1964, p. 91). The equations underneath
the diagram in Figure 5 can be derived from Kirchhoff's rules for electric
circuits.

The general model of Figure 5 can be recast in a form specific to the Limulus
receptor unit by making the following assumptions. Light absorbed by the
photoreceptor causes excitatory conductance changes leading to depolariza-
tion of the receptor nerve cell. It has been suggested that the total excitatory
conductance and potential change is the result of the superposition of 'quan-
tum bumps,' small changes each due perhaps to the absorption of one photon
of light and the subsequent release of transmitter substance (these quantum
bumps are considered further in the section on dynamics). But exactly how
light energy on the front of the ommatidium leads to the changes in the mem-
brane permeability and consequent changes in membrane potential of the
**Figure 5.**
An equivalent circuit for the integration of excitatory and inhibitory influences. \( E_e \) is the electromotive force of the resting potential, and \( R_e \) is the resting resistance value. \( E_i \) is the electromotive force of excitation and \( R_i \) is the variable resistance associated with excitation. \( E_r \) is the electromotive force of inhibition and \( R_r \) is the variable resistance associated with inhibition. \( V \) and \( R \) are the resulting potential and resistance measured across the cell membrane. The departures of the membrane potential and resistance from the resting values are given by \( V \) and \( \Delta R \). The expression for \( V \) may be derived from Kirchhoff's rules for electric circuits.

eccentric cell is still unknown. Nerve impulses coming either from neighboring receptor units (producing 'lateral inhibition') or from the receptor unit in question (producing 'self-inhibition') cause inhibitory conductance changes leading to hyperpolarization. These inhibitory conductance changes might well result from the usual kind of synaptic process as it is reasonable to suppose, on the basis of anatomical evidence, that axon collaterals both from neighboring receptor units and doubling back from the axon of the receptor unit in question form synapses with that receptor unit's axon (Gur, Purple, & Whitehead, 1972). Each inhibitory impulse, arriving at such a synapse, would cause a given quantity of neural transmitter to be released; this transmitter, after diffusing across the synapse, will cause a given conductance change in the postsynaptic membrane. Thus, in the model, every impulse from a particular receptor unit is assumed to cause an identical change (both in time-course
and magnitude) in the inhibitory conductance \( g_i = 1/R_i \). Direct measurements of the membrane potential and conductance of the *Limulus* receptor nerve cell do show changes in the appropriate directions both when the receptor unit is stimulated by light and after nerve impulses from its neighbors or itself (MacNichol, 1956; Fuortes, 1959; Purple, 1964; Purple & Dodge, 1965). However, quantitative consideration of the measurements reveals certain defects in this simple equivalent circuit model of the *Limulus* receptor unit. By taking further into account the known anatomy of the *Limulus* eye, Purple and Dodge were able to modify the model and overcome its defects.

Each ommatidium (Fig. 6A) contains one bipolar neuron—the eccentric cell—that initiates impulse potentials that proceed down its axon into the central nervous system. The dendrite of the eccentric cell is surrounded by 10 to 15 retinular cells arrayed like the segments of an orange. Since the retinular cells contain the photopigment, excitatory influences probably act on the eccentric cell at the junction of its dendrite with the retinular cells. Inhibitory influences, on the other hand, appear to affect the eccentric cell at sites on the other side of the cell body in the lateral plexus, a region of many interconnections between fine branches of the axons of all the ommatidia. This separation of excitatory and inhibitory sites has been handled by using Rall’s extension of the classical ”cable” theory of electronic spread in uniform cylindrical axons to the more complicated geometry of branching dendrites, and cell bodies (Rall, 1962). Calculations (based on properties of the eccentric cell estimated from anatomical and electrophysiological data) using Rall’s theory suggested that the dendrite that contains the excitatory site could be treated as if it were part of the cell body but that the lateral plexus containing the inhibitory site could not. Accordingly, the model of Figure 5 was expanded into the model of Figure 6B by assuming (a) there is a length of ”uniform axon” \( X \) between the cell body and the site of inhibitory conductance change, (b) there is a very long length of uniform axon \( Y \) on the other side of the inhibitory site, and (c) the point of impulse generation is at or beyond the inhibitory site. The lengths of uniform axons, shown as cylinders in the diagram, are assumed to have the properties of a passive cylindrical cable.2

A theorem, proved by Bruce Knight (see Purple, 1964, p. 91), made the above assumptions very easy to use. He showed that properties of an arbitrary length of passive axon can be represented as a three-resistor network. Further, a semi-infinite length of axon can be represented as its resistance to ground. These simplifications allow the model in Figure 6B to be reduced

---

2 This model is still a simplification of the actual situation in which there are many-branched axon collaterals each with different electrical properties instead of passive lengths of uniform axons and inhibition acts at many inhibitory synapses distributed among the collaterals instead of acting at one point on the axon.
to the model in Figure 6C. The resistance $R_e$ and $R_i$ incorporate the information about the properties, including length, of the stretches of axon $X$ and $Y$. This model is in turn simple enough that Kirchhoff's rules for circuits can easily be used to derive formulas that describe the network. From computer solutions of these formulas calculated for various sets of model parameters (various lengths of axon $X$ for example), predictions can be made about the membrane resistance and potential measured at the cell body in response to
electrical and light stimulation. As it turns out, with appropriate choice of parameters this model predicts the observed conductance and potential measurements well (Purple, 1964; Purple & Dodge, 1965).

By adding one more property to the model—that the frequency of impulses is directly proportional to the depolarization (the generator potential) at the site of impulse initiation—the model may be compared with the more abstract Equation 2. Figure 7 summarizes the predictions of this final version of the equivalent circuit model. It shows the decrement in the firing rate of a test ommatidium resulting from six different levels of inhibitory conductance plotted as a function of excitation on the test ommatidium. Since, according to the model, firing rate is a linear function of the membrane potential and change in membrane potential is a hyperbolic function of inhibitory conductance (as is easier to show for the simple model of Fig. 5 but also true in this case), the predicted decrement in firing rate due to inhibition is a hyperbolic function of the inhibitory conductance. Further, since a hyperbolic function is quite linear for part of its range, the predicted decrement in firing rate will be approximately proportional to inhibitory conductance for small values but will begin to saturate gradually as inhibitory conductance reaches higher values. In other words, in Figure 7 the curves for different levels of

![Figure 7](image)

**Figure 7.**
Prediction by the model, shown in Figure 6C, of the decrement in a test ommatidium's firing rate due to inhibition. The abscissa represents the amount of excitation on the test ommatidium expressed in terms of the excitatory conductance ($g_e$). Decrement resulting from six levels of inhibitory conductance ($g_i$) are shown. (Adapted from Purple & Dodge, 1965.)
inhibitory conductance are approximately equally spaced, although at higher levels the curves are slightly closer together than at lower levels. This approximate proportionality between decrement in firing rate and inhibitory conductance means that the equivalent circuit model agrees almost exactly with Equation 2 in two important properties.

1. 'Linear recurrent inhibition'—the property that the decrement in a test ommatidium's firing rate is a linear function of a neighbor's firing rate above inhibitory threshold (in the case of lateral inhibition) or of its own firing rate (in the case of self-inhibition). That is, \( e_p - r_p = k_{pi}(r_i - r_{pi}) \) for the case of two ommatidia \( p \) and \( i \) (see Eq. 2). The equivalent circuit model predicts this relationship because the total inhibitory conductance is the sum of the small conductance changes caused by the individual impulses. Since all the impulses coming from a particular neighbor (at frequencies above the threshold \( r_{pi} \)) cause equal conductance increments, the summed inhibitory conductance due to that neighbor will be proportional to the suprathreshold firing rate of that neighbor. (The constant of proportionality may vary from neighbor to neighbor as is reflected in the different values of \( k_{pi} \).) Since decrement in firing rate is approximately proportional to total inhibitory conductance, the property of (approximate) linear recurrent inhibition is predicted. (The cellular mechanism for this is as yet unknown. However, direct intracellular recordings of the postsynaptic lateral inhibitory potential show that at least two closely spaced antidromic impulses are necessary to produce any postsynaptic potential at all (Knight, Toyoda, & Dodge, 1970). Thus in the model it is simply assumed that below a certain frequency, impulses produce no inhibitory conductance change.)

2. 'Spatial summation'—the property that the total decrement in a test ommatidium's firing rate due to inhibition by several neighboring ommatidia is the sum of the decrements due to each individually. Since the total inhibitory conductance due to the inhibition from several neighbors is exactly the sum of the conductances due to each alone, and since the decrement in the test ommatidium's firing rate is approximately proportional to inhibitory conductance, the property of (approximate) spatial summation is predicted.

The above discussion of agreement between the abstract Equation 2 and the equivalent circuit model implicitly assumed a constant excitatory level in the test ommatidium. What happens as excitation is changed? The equation has the following property in this situation.

3. 'Inhibition independent of excitation'—the decrement in firing rate of the test ommatidium resulting from any given pattern of firing by its neighbors is independent of the amount of excitation in the test ommatidium—that is, for a given \( \sum_j k_{pj}(r_j - r_{pj}) \), \( e_p - r_p \) is the same for all values of \( e_p \). Figure 7 shows that the equivalent circuit model predicts approximate independence: the decrement in firing rate does not change very much as the excitation on the test ommatidium is changed.
In summary, the equivalent circuit model (Purple, 1964; Purple & Dodge, 1965) that explains both lateral and self-inhibition in terms of known cellular mechanisms agrees well with Equation 2 in three fundamental properties: the inhibition is recurrent and linear above threshold, there is spatial summation of inhibition, and the decrement in firing rate caused by a given amount of inhibition is independent of excitation on the test ommatidium. Hence, one can reasonably say that the cellular mechanisms incorporated into the Purple and Dodge model 'explain' the equations that were based entirely on input-output relations.

Those details on which the cellular (equivalent circuit) model and Equation 2 disagree with one another or with direct experimental observations are interesting and can lead to further refinements of both theory and experiment. Barlow and Quarles (personal communication, 1974) found that the magnitude of the response to a 'Mach band' pattern measured in the Limulus eye differed somewhat from the magnitude of the response predicted by the abstract model of Equation 2. At that time there was already some experimental evidence (Lange, 1965) that the inhibitory decrement in the firing rate of a test ommatidium depends on the level of excitation of the test ommatidium (as it does somewhat in the equivalent circuit model). Further experiments (Barlow & Lange, 1974) show this to be a substantial effect, especially at higher levels of excitation much above those generally used in routine experiments. Taking this dependence into account, Barlow and Quarles then found excellent agreement between experimental data and theoretical prediction (see Fig. 4).

2.3. Dynamics

To extend the steady-state model (either the abstract version in Equation 2 or the physiological equivalent circuit) to include the dynamics requires the inclusion of functions representing the time-courses of each of the three component processes: excitation, lateral inhibition, and self-inhibition. Within the last few years, two complementary methods have been successfully used to determine the time-courses of these processes. In one method, the time-courses were deduced by Fourier analysis from measurements of the response of the ommatidium, both membrane potential and firing rate, to stimulation sinusoidally varying in time. In a linear system or within small stimulus ranges in any system, the response of the system to a sinusoidal stimulus is itself a sinusoid of the same frequency differing from the stimulus only in phase and amplitude.¹

¹A linear model is, roughly, an adding and subtracting model—a model in which the response to the sum of two inputs is the sum of the responses to each input alone. More formally, a linear model is a model in which the rule that assigns responses to inputs is a
Thus, to specify the response of the system to all sinusoids one need only specify the 'transfer function,' a function that gives the amplitude (difference between peak and trough) and phase of the sinusoidal response as a function of the frequency of the sinusoidal stimulus. Further, the response of the system to any stimulus can be calculated from the response to sinusoidal stimuli. And, as a last advantage, the transfer function of a system containing several parts can easily be specified in terms of the transfer functions of each part separately. (See Cornsweet, 1970, for an introduction to the use of transfer functions in the study of vision. A number of textbooks describe 'Fourier analysis' or 'linear analysis.'

The second method for studying the time-courses was to measure directly the temporal variations in the membrane potential and conductance occurring during excitation and after inhibitory impulses. These experiments agree with the deductions from the measured transfer functions.

Let us now consider in turn the temporal characteristics revealed by these two methods for the processes underlying excitation, self-inhibition, and lateral inhibition, and some of the consequences for the dynamics of the retinal network.

Excitation. The excitatory component of the generator potential appears to be the sum of numerous 'quantum bumps,' discrete miniature potentials in the depolarizing direction produced by light excitation of the receptors (Yeandle, 1957; Fuortes & Yeandle, 1964; Adolph, 1964; Dodge, Knight, & Toyoda, 1968). The dynamics of excitation, therefore, are determined by the time-course of these quantum bumps. As shown in Figure 8, each quantum bump has a moderately abrupt onset and a slower exponential decay. But the amplitude and duration of these quantum bumps are not always the same. The amplitude and duration of the quantum bumps are markedly dependent on—that is, become adapted to—the mean level of steady illumination (Fig. 9). As the light becomes brighter, the amplitude of each individual quantum bump becomes smaller and the duration becomes shorter. Thus, although the rate of occurrence of these quantum bumps increases in proportion to light intensity, the concomitant decreases in their amplitude and duration are such that in the steady state there is an approximately logarithmic relation between light intensity and the sum of the quantum bumps—the generator potential (solid line of Fig. 9). Further, at low enough frequencies of sinusoidal stimulation, where there is sufficient time during each cycle for some adaptation of the quantum bumps to occur, the generator potential response

linear transformation. For example, the model expressed by Equation 2 gives the responses $r_j$ as a linear transformation of the excitations $e_j$ as long as the $r_j$ are above the thresholds $r_{th}$. Any nonlinear model can be approximated by a linear model when the possible inputs are constrained to a small range.
at each moment approaches that given by the steady-state function (solid line). The change in generator potential resulting from a rapid change in intensity, however, is too fast for significant adaptation of the quantum bumps to take place and thus depends only on rate of occurrence of the quantum bumps. Therefore, at high frequencies of sinusoidal stimulation, the generator potential (the sum of the quantum bumps) at each moment is
FIGURE 9.
Magnitude of generator potential as a function of light intensity. The solid line shows the approximately logarithmic stimulus-response function obtained for steady-state stimulation. The two dashed lines show the linear stimulus-response function obtained for transient changes in intensity level around a high mean intensity level (right-hand dashed line) and around a low mean intensity level (left-hand dashed line). The intensity levels at which the dashed lines cross the solid curve correspond to the mean intensity levels about which the transient variations occur. *Inset:* Generator potential records for long duration light stimuli of different intensities. Nerve impulses were blocked by tetrodotoxin. The generator potential appears to be the sum of small discrete changes in potential, called quantum bumps. At the lowest intensity, (-5 log), the bumps are infrequent and of maximum amplitude. Thus the generator potential, the sum of the individual bumps, is very irregular, and one can easily distinguish individual bumps in the generator potential record. As intensity increases, the frequency of the bumps increases in direct proportion but their amplitude and duration decrease somewhat. Thus, as intensity increases, the records of the generator potential become much smoother. (Graph redrawn from Dodge, 1969; records from Dodge, Knight, & Toyoda, 1968.)

directly proportional to light intensity (dashed curves). The constant of proportionality depends on the mean level of the sinusoidal stimulus, being smaller for high mean intensity stimuli where the individual quantum bumps are smaller (right-hand dashed curve) and larger for low mean intensity stimuli where the individual quantum bumps are larger (left-hand dashed curve). Consequently, the transfer functions of the excitatory process will have different shapes for high and low mean intensity stimuli (Fig. 10). At high mean levels of illumination, the (peak-trough) amplitude of the response to low frequencies will be smaller than the amplitude of the response to higher frequencies because the slope of the steady-state stimulus-response function (solid line in Fig. 9) is smaller than the slope of the high-frequency stimulus-response function (dashed line in Fig. 9). In other words, the ampli-
FIGURE 10.
Temporal transfer function for the excitatory process. The peak-trough amplitude and phase of the generator potential in response to light that was sinusoidally varying at various frequencies for two mean luminance levels. (Impulse generation was blocked by tetrodotoxin so the lateral and self-inhibitory processes could not occur.) (From Dodge, 1969.)
tude characteristic of the transfer function shows a 'low-frequency decline.'
At low mean levels of illumination, however, there is no low-frequency de-
cline because the two stimulus-response curves do not differ.

At all mean intensities, the duration of the quantum bumps becomes an
important factor as the frequency of the sinusoidal stimulus gets so high that
its period is not much longer than the duration of individual quantum
bumps. For these high frequencies, the relatively long duration of the in-
dividual quantum bumps causes 'smearing' or 'integrating' of the responses
initiated at different times in the stimulus cycle. Thus the peak-trough am-
plitude in the response to a high-frequency stimulus is diminished or, in other
words, there is a 'high-frequency' cutoff in the transfer function.

In addition to determining the temporal transfer functions for the excitatory
process, the properties of the quantum bumps determine another dynamic
feature of the ommatidium's response—the small, but significant, variability
in interimpulse intervals. That the variability in the interimpulse interval de-
creased with increased light intensity much the way that variability in the
generator potential did suggested that the variability among intervals was
coupled to the variability of the generator potential (Ratliff, Hartline, &
Lange, 1968). Using 'spectral analysis,' mathematical methods related to the
Fourier analysis methods discussed earlier, this hypothesis was confirmed.
By considering two sources of variability—that variability inherent in the
generator potential of the test ommatidium itself (resulting from the quantum
bumps, see Fig. 9) and that variability transmitted to the test ommatidium
via inhibition from the generator potentials of neighboring ommatidia—one
can completely account for the variability in the train of impulses produced
by the test ommatidium (Shapley, 1971a, 1971b).

Self-inhibition. The low-frequency decline observed in the transfer func-
tion for light intensity to impulse rate (see Fig. 11) depends not only on the
adaptation of quantum bumps but also on self-inhibition—the inhibition
that each impulse discharged exerts back on its own generator potential. The
self-inhibitory potential after each impulse is hyperpolarizing and has an
abrupt onset and an exponential decay with a relatively long time constant
of about 0.5 sec (Fig. 8). Self-inhibition therefore depresses the steady re-
sponse to steady-state stimulation and also reduces the peak-trough amplitude
of the response to low frequencies of modulation. However, due to its long
time constant, it cannot follow high frequencies of modulation and thus does
not affect the peak-trough amplitude of the response to high-frequency stimuli
(Knight, Toyoda, & Dodge, 1970; Biederman-Thorsen & Thorson, 1971).
Unlike adaptation of quantum bumps, self-inhibition produces a low-fre-
frequency decline at all mean luminances. The time constant of self-inhibition—
indeed, the concept of self-inhibition itself—was first deduced by mathemati-
cal analyses of changes in firing rate (see Fig. 1) resulting from step increments
and decrements in the stimulus (Stevens, 1964). Direct electrophysiological observations (Purple & Dodge, 1965) later confirmed this deduction. Although simple in principle, the direct linkage of the self-inhibitory potential to the occurrence of an impulse makes the theoretical analysis of the self-inhibitory feedback complex. For details, see Knight, Toyoda, and Dodge (1970).

**Lateral inhibition.** Because of the characteristics of the quantum bumps and the self-inhibition, a low-frequency decline is an inherent property of the transfer function for light intensity to impulse rate of any individual ommatidium, and the decline is more pronounced at higher mean luminances. And when neighboring ommatidia are also illuminated, the lateral inhibitory potentials that they produce contribute still more to the low-frequency decline because they, like self-inhibition, are of relatively long duration compared to the quantum bumps (Knight, Toyoda, & Dodge, 1970; Biederman-Thorson & Thorson, 1971). Thus they can depress the responses to low frequencies but not the responses to high frequencies. The lateral inhibitory potential, however, has a more complex time-course than does the self-inhibition (Fig. 8). There is first a brief excitatory depolarization followed by a long lasting hyperpolarization. The form of the lateral inhibitory potential was, like the time-course of the self-inhibitory potential, first deduced by mathematical methods—it was determined by taking the Fourier transform of the lateral inhibitory transfer function (modulation of impulse rate in neighbors to modulation of generator potential in the test ommatidium). In subsequent experiments the time-course of the lateral inhibitory potential following each impulse discharged by neighbors was directly observed by electrophysiological means and the mathematical deductions confirmed. Although the amplitude of the inhibitory potential varies greatly with distance between test and inhibiting ommatidia, the time-course evidently does not change with distance; lateral inhibitory transfer functions measured at different distances show large changes in the magnitude of the amplitude characteristic but no significant changes in phase characteristics (Ratliff, Knight, Dodge, & Hartline, 1974). The above time-course for lateral inhibition applies only to inhibition above threshold. The dynamics of the threshold for inhibition are only beginning to be explored (Lange, Hartline, & Ratliff, 1966a; Ratliff, Hartline, & Lange, 1966; Graham, Ratliff, & Hartline, 1973).

**Composite spatiotemporal properties.** Once the dynamic properties of the three component processes—excitation, self-inhibition, and lateral inhibition—have been expressed in terms of their time-course (Fig. 8) or in terms of their transfer functions, either form can be used to extend Equation 2 to include the dynamic properties. The transfer function form, used because of
its greater simplicity and its closer relationship to the majority of the experimental data, gives the following dynamic equations:

\[ r_p(f) = g(f)I(f) - k_{pp}T_o(f) - T_o(f) \sum_{p \neq p} k_{pp}r_p(f), \]  

(3)

where \( r_p(f) \) is the peak-trough amplitude in the firing rate of ommatidium \( p \) in response to a sinusoidal light stimulus of frequency \( f \), \( I(f) \) is the peak-trough amplitude of the light stimulating \( p \), \( g(f) \) is the transfer function of the excitatory process, \( T_o(f) \) is the transfer function of the self-inhibitory process, \( T_i(f) \) is the transfer function of the lateral inhibitory process, and the \( k_{pp} \) and \( k_{pp} \) are the lateral and self-inhibitory coefficients already introduced. \( T_o(f) \) and \( T_i(f) \) are normalized so that in the steady state these equations will reduce to Equation 2. Because the time-course of lateral inhibition is independent of distance, the same transfer function \( T_i(f) \) can apply to all pairs of ommatidia. Notice that these equations apply only for suprathreshold stimuli, and introduce a minor error by ignoring the direct linkage between the lateral and self-inhibitory potentials and the individual impulses that produce them.

Equation 3 can be used to predict fully the behavior of the Limulus retina, including responses to simultaneous variations of stimuli in both space and time. Figure 11 shows theoretical predictions based on these equations and an experimental confirmation of them (Knight, Toyoda, & Dodge, 1970). The predicted light-to-impulse-rate transfer function when a spot of light is confined to the test ommatidium is shown by the solid lines and the corresponding measured transfer function by the filled circles. The predicted transfer function and data points when the spot of light is enlarged to illuminate not only the test receptor but also about 20 of its neighbors are shown by the dashed lines and open circles. Notice that in both situations there is a high-frequency cutoff resulting from limitations of the generator mechanism and a low-frequency decline due to adaptation of the quantum bumps and self-inhibition as already discussed. When the neighbors are illuminated, the low-frequency decline is further accentuated by lateral inhibition from them.

In both cases the high-frequency cutoff and the low-frequency decline 'tune' the system (network) to best transmit the intermediate frequencies. Notice, however, that the peak-to-trough amplitude of the response at the best tuned frequency is greater with lateral inhibition (large spot) than it is without (small spot) (Ratliff, Knight, Toyoda, & Hartline, 1967). This 'amplification' is a direct consequence of the long delay (about 150 msec) to the peak of the inhibitory potential (Ratliff, Knight, & Graham, 1969). Because of this delay, for stimulation at the best tuned frequency (period about 300 msec), the inhibition will be maximal at the trough of the response and minimal at the peak, thus tending to produce the greatest possible peak-to-trough amplitude.
FIGURE 11.
Theoretical and observed transfer functions for light-to-impulse rate. Filled circles show peak-to-peak amplitude and phase of impulse rate for a single ommatidium illuminated alone by a small spot of light. Open circles show amplitude and phase for that same ommatidium when the spot of light centered on it was enlarged to include 18 neighboring ommatidia. Amplitude is given in sec^-1 peak-to-peak. Mean rate was 25/sec for solid circles and 20/sec for open circles (Adapted from Knight, Toyoda, & Dodge, 1970).
There is no question that these effects are due to the delayed lateral inhibition. It is easy to demonstrate this by introducing further artificial delays. This can be done simply by delaying the stimulus to the neighbors that are exerting the inhibition on the test receptor (Ratliff, Knight, & Milkman, 1970). As the delay is increased the peak of the transfer function becomes considerably amplified and shifts to lower and lower frequencies. As predicted by the theory, second-order maxima and minima also appear as the delay is lengthened.

It is easy to transfer to the spatial domain the basic concepts of Fourier analysis that were discussed above for the temporal domain. Instead of sinusoidal variations in time, one simply considers sinusoidal variations in space—and, instead of the temporal distributions of excitation and inhibition, the spatial distribution. The resulting spatial transfer function shows a pronounced maximum at intermediate spatial frequencies. The broad extent of the inhibitory field (Fig. 3) produces the low-spatial-frequency decline and the limitations imposed by the optical apparatus and dimensions of the receptor units account for the high-spatial-frequency cutoff. Theoretical predictions even show an amplification at intermediate spatial frequencies (because of the distance to the point of maximum lateral inhibition as shown in Fig. 3) analogous to the amplification at intermediate temporal frequencies. This analogy between the spatial and temporal frequency domains has not yet been demonstrated by experiment.

A most important property utilized by Fourier methods of analysis is that any function may be expressed as the sum of sinusoids of various frequencies, amplitudes, and phases. Thus in a linear system or over a small range in any system the response to any stimulus may be calculated as the sum of the responses to the component sinusoids. An example of such a calculated response along with the response measured directly is illustrated in Figure 12. A group of receptors was illuminated with temporal square-wave modulation (a train of abrupt steps above and below the mean luminance) while a neighboring unit was illuminated steadily. From the previously measured responses of the group of receptors to sinusoidal stimulation (excitatory transfer function) and the concomitant inhibitory effect of these sinusoidal responses on the steadily illuminated neighbor (lateral inhibitory transfer function), the predicted responses to the square-wave modulation were calculated using the theory of Equation 3. The actual responses to square-wave modulation, both the directly measured excitatory responses in the group and the indirect inhibitory effects that were produced in the neighbor, agree very well with the calculations (Ratliff et al., 1974).
2.4. Summary

The responses of the *Limulus* retina to various arbitrary spatial and temporal patterns of illumination have been predicted accurately using a set of \( n \) simultaneous equations—one equation for each of the \( n \) receptor units and in each of which the three basic processes of excitation, self-inhibition, and lateral inhibition are represented as separate spatiotemporal transfer functions.

Excitation of a receptor unit in *Limulus* can only be produced by illumination falling directly on it. Each absorbed quantum of light seems to produce a brief miniature excitatory potential—a quantum bump—and these quantum bumps add together to form the excitatory component of the generator potential. The nonlinearities found in the response of the *Limulus* receptor unit as light intensity is raised (decreased sensitivity and changes in temporal response characteristics) are primarily due to the properties of this excitatory process: The amplitude and duration of the individual quantum bumps become smaller as the number of recent photon absorptions becomes larger. The variability in the response of the receptor unit is also due to the excitatory process: The variability inherent in the quantal nature of light is apparently reflected in the occurrence of quantum bumps and is thus transmitted to the train of impulses discharged by the ommatidia.

Self-inhibition also involves only a single receptor unit. Each impulse discharged down a receptor unit's axon produces, in that same receptor unit, a relatively long duration (about 0.5 sec) inhibitory potential that subtracts from the excitatory potential caused by light. This self-inhibition is an important factor in many of the temporal features of a receptor unit's response, including the transient peak in the response to an increment of light.

The lateral influence of neighboring receptor units in the *Limulus* retina on one another is inhibitory. Each impulse discharged down one unit's axon

---

**Figure 12.** Theoretical predictions and experimental observations of concurrent excitatory and inhibitory responses in neighboring ommatidia. *Left-hand graph:* The excitatory transfer function (A) shows the peak-trough amplitude (left ordinate) and phase shift (right ordinate) of the rate of discharge of one member of a group of ommatidia stimulated by sinusoidally modulated illumination. The inhibitory transfer function (B) is the peak-trough amplitude and phase shift of the rate of discharge of a neighboring ommatidium, which though illuminated steadily was inhibited by the group of ommatidia responding to sinusoidal illumination. *Right-hand graph:* The excitatory response (A) is the discharge rate of one member of the group of ommatidia stimulated by temporal square wave illumination. The inhibitory response (B) is the discharge rate of the one neighboring ommatidium, which though illuminated steadily was inhibited by the ommatidia responding to the square wave illumination. The theoretical curves are predicted from the responses (left-hand graph) to the sinusoidal stimulus. The experimental observations are displaced 5 impulses/sec downward from the theoretical predictions. (From Ratliff et al., 1974.)
will produce, in another unit, an inhibitory potential of a duration (0.15 sec) that is shorter than the duration of the self-inhibitory potential. (The lateral inhibitory potential is always preceded by a small brief lateral excitatory potential.) The magnitude of the lateral inhibition depends on the distance between the two receptor units, being largest in general when the two units are three or four units apart although never nearly as large as the self-inhibitory potential. Due to the differences in their spatial and temporal characteristics, the interaction of lateral inhibition with self-inhibition and excitation can produce many complicated spatial and temporal effects, of which the best known are probably the 'Mach bands' or 'border contrast' effects and the 'tuning' to relatively narrow bands of temporal frequencies.

3. INTEGRATIVE ACTION OF RETINAL
GANGLION CELLS IN CAT AND GOLDFISH

3.1. General Model

Some characteristics of retinal ganglion cells. The Sherringtonian concept of the receptive field is basic to an understanding of the integrative action of the vertebrate retina. It was first applied by Hartline (1938a) to the study of vertebrate retinal ganglion cells (those cells whose axons form the optic nerve that extends from the eye to the central nervous system). In Hartline's words (1941): "The retinal region occupied by visual sense cells whose connections converge upon a given retinal ganglion cell shall be termed the receptive field of that ganglion cell. . . . Not only do excitatory influences converge upon each ganglion cell from different parts of its receptive field, but . . . inhibitory influences converge as well."

Hartline observed three basic types of ganglion cells in the frog retina: 'on,' 'off,' and 'on-off'—so named because they responded most actively at the onset of light, the offset of light, or at both onset and offset. He noted the existence of intermediate types, however, and it has since become evident that there are many others that are highly specialized in other respects (sensitivity to motion, orientation, color, etc.). Since vertebrate retinas are far more complex than the retina of Limulus and the recording and analysis of unit activity in them more recent, quantitative theories of their integrative action are less complete. However, in the cat and goldfish the retinal ganglion cells have been studied extensively in recent years, and a quantitative model of their discharge of impulses in response to illumination is beginning to emerge. And despite the ages of evolution separating these vertebrates from Limulus and despite the great differences between the anatomy of vertebrate retinas and that of Limulus (compare Figs. 2 and 13), the behavior of the cat and
The spatial separation of opposed excitatory and inhibitory processes that was found in the *Limulus* retina is found also in the cat and goldfish retinas.\footnote{The primary emphasis of the following is on the retinal ganglion cells themselves, rather than on the relationship of these neurons to the functioning of the whole visual system. Thus facts that may be very important in understanding significant phenomena of visual perception but that are not important in discussing the working of these individual neurons.}
Such a spatial separation in a vertebrate retina was first discovered in the cat by Kuffler (1953) and later in the goldfish by Wagner, MacNichol, and Wolbarsht (1963). From a given retinal ganglion cell, light shining in one area of the retina elicits an 'on-type' response characterized by an increase in firing rate at the onset of a light and a decrease in firing at the offset of light, and light shining in another area elicits an 'off-type' response characterized by a decrease in firing rate at the onset of light and an increase in firing at the offset of light. These two opposing types of response are illustrated in Figure 14 both as photographs of the actual nerve potentials displayed on an oscilloscope face and as poststimulus-time-histograms (counts of the number of nerve impulses occurring in each short time period). For this goldfish cell, responses of the off-type are elicited from a central circular area; responses of the on-type are elicited from a surrounding annular area. Therefore the cell is said to have an off-center on-surround receptive field for light of 710 nms. About an equal number of cells of the opposite organization (on-center, off-surround) are also found. In the cat, response type is (almost) independent of the wavelength of light, but in the goldfish a cell may have an on-center off-surround organization for some wavelengths and an off-center on-surround organization for others. (Spekreijse, Wagner, & Wolbarsht, 1972; present in detail the various wavelength dependencies found in goldfish retinal ganglion cells.)

In any of these various types of cells, two spots of light that individually elicit opposite types of response tend to cancel each other's effect when simultaneously presented. Thus, for example, for the cell of Figure 14 both the 'off' response to a spot of light in the center of the cell's receptive field and the 'on' response to an annulus of light in the surround of the receptive field will be smaller when both spot and annulus are presented simultaneously than when each is presented alone. As is expected from such an antagonistic spatial organization (see Fig. 4 and discussion following Fig. 11), effects analogous to Mach bands are observed in the responses of retinal ganglion cells to edge patterns (Baumgartner, 1961; Enroth-Cugell & Robson, 1966), and the spatial transfer function of these cells typically shows a peak sensitivity at medium spatial frequencies (Enroth-Cugell & Robson, 1966).

Temporal characteristics of the responses of cat and goldfish retinal ganglion cells are also similar to the characteristics of Limulus responses. The
transients in the responses to steps of illumination (Fig. 14) resemble those transients in the \textit{Limulus} responses caused by 'self-inhibition' and adaptation of quantum bumps (Figs. 1 and 9). And the response of the surround mechanism of a retinal ganglion cell may well be delayed with respect to the response of the center mechanism just as the lateral inhibition in the \textit{Limulus} is delayed with respect to the excitation.

\textit{Linear model.} Rodieck and Stone (1965a, 1965b; Rodieck, 1965) proposed and tested a linear model of the cat retinal ganglion cells that incorporated the features described above and strongly resembled the model of the \textit{Limulus} eye. Figure 15 is a diagram of their model. Compare the responses
Graphical representation of on-center and off-surround mechanisms postulated by Rodieck (1965) and Rodieck and Stone (1965a, 1965b). The top graphs show the spatial distribution of responsiveness for each mechanism. The middle graphs show the temporal weighting functions (the time-course of the response of each mechanism to a brief pulse of light). The bottom graphs show the time-course of the response of each mechanism to steps of light. See text for definition of symbols.
computed with their model for the on-center and off-surround mechanisms of a cat retinal ganglion cell with those computed using Equation 3 for the excitatory responses and inhibitory effects in Limulus (Fig. 12). For an on-center off-surround retinal ganglion cell, whose receptive field is centered at point $x$, the response at time $t$, called $r(t, x)$ is postulated to be the sum of a positive input from a 'center mechanism' and a negative input from a 'surround mechanism.' For an off-center cell, the signs are reversed. Each mechanism's contribution to the ganglion cell response is the sum (or integral when writing the functions in continuous form) of the luminances at each point on the retina weighted according to that mechanism's spatial distribution and time-course. Formally

$$ r(t, x) = \iint I(y, \tau)k_c(x - y)l_c(t - \tau) \, d\tau \, dy + \iint I(y, \tau)k_s(x - y)l_s(t - \tau) \, d\tau \, dy, \tag{4} $$

where $r(t, x)$ is the response of the ganglion cell, $I(y, \tau)$ is the luminance at point $y$ and time $\tau$, $k_c(x)$ and $k_s(x)$ are the spatial weighting functions of the center and surround mechanisms, respectively, and $l_c(t)$ and $l_s(t)$ are the temporal weighting functions (the response to an instantaneous pulse of light) of the center and surround mechanisms, respectively. The spatial and temporal weighting functions that Rodieck and Stone chose are shown in Figure 15. For the center spatial distribution $k_c(x)$ they used a narrow Gaussian function, and for the surround spatial distribution $k_s(x)$, a wide Gaussian. (Notice that both mechanisms are sensitive to light shining in the central portion of the field but that the center mechanism is much more sensitive.) The center temporal weighting function $l_c(t)$ was an instantaneous pulse of excitation (a Dirac delta function) followed by an exponentially decaying inhibitory pulse (formally analogous to 'self-inhibition'). The surround temporal weighting function $l_s(t)$ was a delay of length $\tau_s$ followed by the negative of the center temporal weighting function. To provide an alternate picture of the temporal assumptions of the model, the temporal responses of each mechanism to steps of light are also shown in Figure 15. Although Equation 4 is unlike the Limulus model in having an 'excitatory center' that spreads over some distance and in having nonrecurrent inhibition rather than recurrent (the 'inhibitory surround's response depends on the luminance at each point, not on the response), the resemblance to the Limulus model is clear—summation of two opposing mechanisms, spatial summation within a mechanism, temporal transients of an exponentially decaying form, and different time-courses for the two mechanisms.

Rodieck and Stone tested this linear model by computing its responses to various moving patterns and comparing the model's behavior to the behavior of cat retinal ganglion cells (Fig. 16). Qualitatively, the model's predictions
were good. However, it is easy to show that no linear model will completely describe the cat or goldfish retinal ganglion cells.

Nonlinearities. The response of a retinal ganglion cell to a given increment (a sudden increase or 'step up' in luminance) becomes smaller as the steady background on which the increment is superimposed is made more intense. Equivalently, the intensity of an increment of light necessary to evoke a response of given magnitude increases as the background intensity increases. Figure 17, adapted from Shapley et al. (1972), illustrates this latter change with background level for the response of cat retinal ganglion cells. Background intensity is plotted on the horizontal axis and the intensity of the increment that produced a criterion response magnitude is plotted on the vertical axis. (Response magnitude was measured as maximum firing rate in response to the increment minus steady firing rate to background. The criterion response magnitude was 30 impulses/sec.) The vertical axis is also labeled in units of response magnitude divided by stimulus magnitude, a ratio which is a common definition of the 'gain' of a system. (Although other definitions are sometimes used, here 'gain' will always be used this way.) As
can be seen in Figure 17, the incremental intensity necessary to evoke a constant response (the gain) is approximately constant at low-intensity backgrounds. And then incremental intensity gradually increases (gain decreases) until at higher backgrounds, incremental intensity increases (gain decreases) in proportion to background intensity (producing a slope of one on the log-log plot). Similar functions have been found by other investigators (Barlow & Levick, 1969a, 1969b; Cland & Enroth-Cugell, 1968; Maffei, Fiorentini, & Cervetto, 1971; Yoon, 1972). In a linear system, the response to a given increment of light is always the same, or in other words, gain is the same at all background intensities. Consequently, a major 'gain-changing' nonlinearity must be introduced into the retinal ganglion cell model to explain results like those of Figure 17. Further, this is not the only type of result showing nonlinearity in cat and goldfish retinal ganglion cell responses. Evidence to be considered later suggests the existence of several different nonlinear processes, a reasonable conclusion considering the anatomical complexity of the vertebrate retina.

**General model.** To provide a framework within which to discuss these complex processes, a diagram of a general model for cat and goldfish retinal ganglion cells is presented in Figure 18. For the goldfish retinal cell, the model to be complete should be elaborated to include the opposed color-coded processes. But, whenever, as is the case in most experiments to be discussed here, light selected to affect only one of the color-coded processes is used, the model of Figure 18 will be adequate. In the general model, there are three stages: early local processing at each retinal point; combination (separately for the center and surround mechanisms) of the early local signals originating at different places on the retina; and lastly combination of the signal from the center mechanism with the opposed signal from the surround mechanism. The model of Figure 18 certainly does not include all possible models of retinal ganglion cells. For example, one might imagine that at the first combination points, signals coming from part of the surround might be combined with those coming from part of the center, and then at the final combination points these partial center-surround combinations might themselves combine. However, the model of Figure 18 is sufficiently general to serve as a framework within which to discuss the known results from the cat and goldfish retinal ganglion cells.

What is known about the processing at each stage will be examined, starting at the final combination point and working backwards. In particular, the nonlinear or 'gain-changing' effects of varying light intensity will be discussed, but other types of nonlinear processing will also be considered. Throughout most of this discussion temporal factors will be ignored and the responses of the cells talked about as if they were in a 'steady-state.' At the end of the discussion, however, the temporal properties of cat and goldfish retinal gan-
glion cells will be examined briefly. Finally, the recent division of the cat's retinal ganglion cells into two classes, X (sustained) and Y (transient), will be discussed (Enroth-Cugell & Robson, 1966). The Y cells exhibit more nonlinear behavior than do the X cells and it would be most satisfying to discuss the processing at each stage of the general model for these two classes separately. This cannot be done systematically, however, because most studies have not distinguished the two classes.

3.2. Combinations of Signals from Center and Surround

At the final combination point, the signal from the center appears simply to add to the signal from the surround. This has been well demonstrated for the cat retinal ganglion cell by Enroth-Cugell and Pinto (1970; 1972). They used an annulus flashing on and off in the periphery to elicit a response from the surround mechanism, a spot flashing on and off in the center to elicit a response from the center mechanism, and then both stimuli flashing together (in phase or antiphase). The stimuli and responses, as well as a computed sum are shown in Figure 19. The response of the cell to the two stimuli flashing together was indistinguishable from the sum of the responses to the individual stimuli as added by computer. This demonstration is particularly convincing because of certain precautions taken in the experimental procedure: (a) An artificial pupil was used and the distribution of light on the retina actually measured. (The optics of the cat eye are not good (Wässle, 1971; Bonds, Enroth-Cugell, & Pinto, 1972) and consequently the distribution of light on the retina may be very unlike that that the experimenter is trying to produce); (b) Various criteria were used to decide whether a given response was a pure-center response or a pure-surround response or mixed. Obtaining pure responses was important in this experiment because if both spot and annulus had stimulated the same mechanism (if either had elicited an impure response), early nonlinearities described in the next section would have con-
Center mechanism

Surround mechanism

... $I_1(t)$  
$f_{c1}(I_1)$  
$f_{c2}(I_2)$  
$f_{c3}(I_3)$  

... $I_2(t)$  
$f_{s1}(I_1)$  
$f_{s2}(I_2)$  
$f_{s3}(I_3)$  

... $I_3(t)$  
$g_c(f_{c1}, f_{c2}, ...)$  
$g_s(f_{s1}, f_{s2}, ...)$

$h(g_c, g_s)$

Light distribution on retina

Early local processing

Combination points for each mechanism separately

Final combination point

**Figure 18.** Diagram of general model for cat and goldfish retinal ganglion cells.
FIGURE 19.
Linear combination of responses from the center and surround mechanisms of an on-center cat retinal ganglion cell. (A) Pure center response to a spot flashing on and off in the receptive field center (indicated by filled stimulus profile on right) which was superimposed on a steadily illuminated spot (open stimulus profile). (B) Pure surround response to an annulus flashing on and off in the receptive field surround in the presence of the steadily illuminated center spot. (C) The cell's response to both the central spot and surround annulus flashing in phase (in presence of steadily illuminated center spot). (D) Computed sum of responses A and B in phase. (E) Juxtaposition of (C) and (D). (F) The cell's response to both the central spot and surround annulus flashing out of phase (in presence of steadily illuminated center spot). (G) Computed sum of responses (A) and (B) out of phase. (H) Juxtaposition of (F) and (G). Stimuli were flashing on and off at 0.4 cps. (Adapted from Enroth-Cugell & Pinto, 1972.)
sealed the linearity at the final combination point. There is evidence that a class of cells exists from which one can never get pure responses, particularly not pure surround responses (Enroth-Cugell & Pinto, 1972). If that is the case, the conclusions from this experiment cannot extend to those cells.

Another experiment (Maffei, Cervetto, & Fiorentini, 1970) indicating linearity at the final combination point for the cat retinal ganglion cell showed an interesting result of such combination. A spot in the receptive field center and a spot in the receptive field surround were flickered sinusoidally, either individually or together. The response to both was equal to the sum of the responses to each alone (Fig. 20). The differences between the transfer functions for the center and surround spots, however, lead to a complicated combination transfer function showing 'amplification' (similar to that in Limulus, see Fig. 11) and two peaks (also similar to Limulus when lateral inhibition is artificially delayed, see Ratliff, Knight, & Milikman, 1970).

Notice that linear combination of the signals from center and surround mechanisms does not guarantee that the magnitude of the combined response will be a monotonic function of light intensity, even if the magnitudes of both the center and surround mechanisms' individual responses are, for nonlinearities within the separate mechanisms may make the surround responses relatively more important in one part of the light intensity range than in another. Indeed, for on-center cells, the magnitude of the combined response (both initial transient and maintained discharge) to light of different intensities first increases as light intensity is raised but then decreases with further increases in light intensity. This decrease at higher light intensities is attributed to an increase in the relative contribution from the surround, that is, the positive signal from the center dominates at low intensities and the negative signal from the surround dominates at high intensities (Barlow & Levick, 1969b; Winters & Walters, 1970; Maffei, Fiorentini, & Cervetto, 1971). Similar results have been found for off-center cells, although there is some indication that off-center cells may differ from on-center cells in this regard.

After the basic addition of the signals from center and surround there is, at least in some situations, a 'rectification nonlinearity,' a zeroing of all responses below some fixed level. Formally,

\[
h(t) = \begin{cases} 
  g_c(t) + g_s(t), & \text{if } g_c(t) + g_s(t) \text{ greater than some fixed level,} \\
  0, & \text{otherwise,}
\end{cases} \tag{5}
\]

where \( h(t) \) is the response of the retinal ganglion cell and \( g_c(t) \) and \( g_s(t) \) are the responses of the center and surround mechanisms individually. A rectification nonlinearity was suggested for the goldfish retinal ganglion cell by Spekreijse (1969) to explain the truncated appearance of responses to flickering light. The rectification nonlinearity's instantaneous nature (its dependence only on the input at the present moment, not on past inputs) has
Figure 20:
Transfer functions for the center mechanism, the surround mechanism, and both mechanisms together of a cat on-center retinal ganglion cell. The transfer functions were obtained from the responses to sinusoidally varying illumination of a spot in the receptive field center (filled circles), of a spot in the receptive field surround (open circles), and of both spots together in phase (filled triangles). The transfer function for both spots together computed from the linear sum of the responses to the central spot alone and peripheral spot alone (considering phase as well as amplitude) is shown by the open triangles. (Adapted from Maffei, Cervetto, & Fiorentini, 1970.)

received some confirmation since the responses can be 'relinearized' by introduction of auxiliary noise signals (Spekreijse, 1969) and since the responses to white-noise stimuli and to sinusoidal flicker yield the same transfer functions as is expected if the nonlinearity is instantaneous (Schellart & Spekreijse, 1972). Spekreijse and van den Berg (1971) showed that the rectification must occur at the final combination point of center and surround after the signals from center and surround have been added. For the cat eye, Grüsser and his
co-workers (Büttner, Büttner, & Grüsser, 1971, for example) report such a
rectification of sine-wave responses, although several other investigators have
not (Cleland & Enroth-Cugell, 1966; Hughes & Maffei, 1966, for example).
This discrepancy may be due both to the depths of modulation used in the
stimuli and to differences in the physiological states of the preparations.

As Enroth-Cugell and Pinto (1972) discuss in some detail, it is hard to rule
out the possibility that some type of nonlinear combination is imitating linear
addition at this or any other point in the model. However, if that is the case,
the imitation is good.

Simple and interaction nonlinearities. The rectification nonlinearity \( h(t) \) is
an example of what will be called a simple nonlinearity—it is equivalent to a
linear combination of signals coming from different sources followed by some
nonlinear function of that sum. Nonlinear transformations that are not simple
will be called interaction nonlinearities. Examples of interaction nonlinearities
are: (a) division of the signal from the center mechanism by the signal from
the surround mechanism at the final combination point, (b) 'shunting' of the
center signal by the surround signal at the final combination point (Furman,
1965; Sperling, 1970), and (c) pairwise interaction of the early local signals
at the center or surround combination point such as

\[
g(a_1, f_2, f_3, \ldots) = \sum a_i f_i + \sum b_{ij} f_i f_j
\]

Since interaction nonlinearities have not been considered extensively as
models of retinal ganglion cell data, they will not be dealt with in detail here
(but see Sperling, 1970). However, it should be remembered that, although
interaction nonlinearities have not been considered extensively, neither have
they been entirely ruled out.

3.3. Separate Combination Points for Center
and Surround

Since the final combination of center and surround signals appears to be
linear except for a rectification, the gain-changing effect of changing light
intensity that is apparent in the cell's response (for example, Fig. 17) must
take place prior to the final combination point. In fact, there appear to be
such nonlinear effects at both the early local stage and at the separate combi-
nation points for the center and the surround mechanisms. Cleland and
Enroth-Cugell (1968) provided a particularly clear and quantitative demon-
stration that there is a change in gain resulting from the transformation at
the center combination point.

Four stimulus arrangements were used in their experiments: (a) To measure
the distribution of sensitivity across the receptive field center—a small test spot of \textit{variable} intensity at \textit{various} locations on the retina, superimposed on a very large \textit{constant} background stimulus; (b) To compare the threshold for several small spots with that for one spot \textit{variable} number and intensity of small test spots superimposed on a \textit{constant} large background; (c) To measure the trade-off between area and intensity for the test stimulus—a circular test stimulus of \textit{variable} area and intensity superimposed on a concentric circular background stimulus of \textit{constant} area and intensity; (d) To measure the trade-off between area and intensity for the adapting or 'gain-changing' background stimulus—a circular test stimulus of \textit{constant} area and intensity superimposed on a concentric circular background stimulus of \textit{variable} area and intensity. In each arrangement, the variable quantities were adjusted until the response to the intermittent test stimulus was of just threshold magnitude. (Thus, even if there were nonlinearities at the final combination point, they would not matter since the response \( h(i) \) and thus the input \( g(i) \) to the final combination point were being held constant.)

The results of the experiments demonstrated that the response of the center mechanism of a retinal ganglion cell to a test stimulus superimposed on a background stimulus depended only on two properties of the stimulus situation—the 'center effective flux' of the test stimulus and the 'center effective flux' of the background stimulus. Center effective flux is defined as the sum (or, in general, the integral) of the intensities at each point in the stimulus weighted by the sensitivities at each point of the center mechanism. For a spatially homogeneous stimulus and a center mechanism having approximately uniform sensitivity over the whole center, center effective flux is defined as \( I \times A_s \), where \( I \) is the intensity of the stimulus and \( A_s \) is the area of overlap between the stimulus and the center of the receptive field. In other words, Cleland and Enroth-Cugell (1968) showed that as long as the center effective fluxes of the test and background stimuli were held constant, the spatial distribution of stimulation could be varied without changing the response.

Of course, all possible spatial distributions of stimulation have not been tested so the above statement may not be true for all stimuli. Suppose, however, that it is. What would this dependence solely on center effective flux imply for the general model? First, it suggests that in this experimental situation the early local transformations for the center mechanism are linear and identical for all points within the center of the field (and of course are zero for points outside the center). Notice that combining intensities of light falling on separate retinal points to form the center effective flux \( I \times A_s \) is equivalent to having all light falling within the center illuminate one retinal point. If the early local transformations were not linear or differed substantially in temporal properties, the effects of light illuminating different retinal points would have been distorted relative to one another by early nonlinearities or different
time-courses and thus the light could not have acted as if it were all illuminating one point. (Although evidently linear in these experiments, the early local processing is sometimes found to be nonlinear as is discussed in detail below.)

Second, the importance of center effective flux implies that the gain-changing nonlinearity cannot be postponed until the third stage of the model (except in the trivial way of postponing some nonlinearity acting on the center signal alone until the final combination point). The experiment in which the adapting background was varied (stimulus arrangement d) shows that the responses to a test stimulus superimposed on an adapting background depend only on that part of the background illuminating the center mechanism and not at all on light illuminating the surround mechanism.

Finally, the dependence of the response on center effective flux implies that the nonlinearity acting at the center combination point must be a 'simple' nonlinearity as defined earlier. It must be equivalent to a linear addition of the early local signals coming from different points in the retina followed by some nonlinear function of that sum. For, since the early local transformations are linear and identical in this situation, the sum of the inputs arriving at the center combination \( \left( \sum s \right) \) is proportional to the center effective flux.

Thus, the experimental results can be rephrased as follows: Regardless of the spatial distribution of the stimulus, as long as the sum of inputs to the center combination point is constant, the response of the center mechanism is constant. If the nonlinearity at the center combination point were not simple, i.e., were not equivalent to a nonlinear function acting after the sum of inputs had been formed, a constant sum of inputs would not insure a constant center mechanism response.

Cleland and Enroth-Cugell (1970) made further measurements and showed that, for the range of light intensities in which they were working, the magnitude of the response of the center mechanism \( g_e \) to a test stimulus of effective center flux \( \Delta F \) superimposed on a background stimulus of effective center flux \( F \) can be expressed as

\[
g_e = \frac{k \cdot \Delta F}{(\Delta F + F)^{1.7}} \tag{6}
\]

Although these experiments could not be quantitatively repeated on the surround mechanism due to difficulties in isolating pure surround responses, Enroth-Cugell and Pinto (1972) found some evidence that the surround mechanism acts like the center mechanism: Its response to a test stimulus and the adaptation of such a response that is caused by a steady background both seem to depend on the 'surround effective flux' (intensity weighted by the spatial distribution of the surround's responsiveness) and not on other spatial aspects of the stimulus. However, in a large class of cells ('surround-conceal-
ing' cells) it was impossible to isolate a pure-surround response to any stimulus, and Enroth-Cugell and Pinto did not study these cells. Further, Cleland, Levick, and Sanderson (1973) present results suggesting that for Y cells there is no gain-changing nonlinearity at the surround combination point: An adapting background on one part of a Y cell's receptive field surround did not affect the response to a test stimulus on another part. For X cells, the results were mixed: Some cells were like the Y cells in this regard but others did show evidence of a gain-changing nonlinearity at the surround combination point (an adapting background on one part of the surround did affect the response to a test stimulus on another part). Since the relationship between the surround-concealing versus surround-revealing distinction and the X versus Y distinction is unclear, and since Cleland, Levick, and Sanderson did not use 'pure surround' responses, it is difficult at this time to compare the results of these two studies directly.

Further evidence of a gain-changing process at the center combination point of the cat retinal ganglion cells comes from experiments in which separate areas of the center of the receptive field are illuminated either individually or together (Büttner & Grüsser, 1968; Stone & Fabian, 1968). If the transformation at the center combination point were a linear summation of early local signals coming from different retinal points (remembering that the transformation at the final combination point is linear except for rectification), the response to the two small areas illuminated together would be equal to the sum of the responses to each alone. However, the response when they are illuminated together is a good deal smaller than the sum of the responses to each alone. Similar evidence comes from studies in which both the intensity and area of a centered stimulus were varied independently and the responses measured (Winters & Walters, 1970; Creutzfeldt, Sakmann, Scheich, & Korn, 1970).

To express this 'compressive' nonlinear summation, formulas of the following type have been suggested (Büttner & Grüsser, 1968; Grüsser, Schaible, & Vierkant-Glath, 1970; Creutzfeldt et al., 1970; Fischer & Freund, 1970; Grüsser, 1971).

\[
g_c = \frac{\sum j_{c,i}}{a + b \sum j_{c,i}}
\]

where \(g_c\) and the \(j_{c,i}\) are measures of the magnitudes of the center mechanism's response and the early local signals, and \(a\) and \(b\) are constants to be estimated. Such formulas apply to stimuli turned on simultaneously and do not consider the effect of a steady background. Nevertheless the similarity of the two formulations is clear: As the input to the center mechanism (\(\Delta F\) in Eq. 6 and \(\sum j_{c,i}\) in Eq. 7) increases, the response of the center mechanism does not increase in direct proportion but increases at a progressively slower
rate as the quantities in the denominators of the expressions become larger.

The evidence so far presented demonstrating a gain-changing nonlinearity at the center (and perhaps surround) combination points has all come from cat retinal ganglion cells. Corroborating evidence indicating nonlinearity at the center combination point has been found for the goldfish retinal ganglion cell: Certain temporal properties of responses to stimuli in the center of the receptive field, which are closely related to gain-change as defined earlier, depend on stimulus intensity summed over area rather than just on intensity (Schellart & Spekreijse, 1972); also, the responses to two small spots in the receptive field center indicate that there must be a nonlinearity at or after the point at which the signals from two spots combine (Levine, 1972; Abramov & Levine, 1974).

3.4. Early Processing

The evidence for some nonlinear early local processing in addition to the nonlinearity at the center combination point is quite extensive in the case of the goldfish retinal ganglion cell. Easter (1968a) demonstrated that the response to two small spots at a given intensity (where the spots illuminate retinal locations of equal sensitivity) is larger than the response to one spot at twice the given intensity. This result cannot be explained by a simple nonlinearity at the center combination point after linear early local processing, for, in that case, the sum of the early local signals arriving at the center combination point would be the same in both the one-spot and two-spot situations, and thus the responses would be the same. This result can be explained by a nonlinear early local transformation that compresses responses so that doubling the stimulus intensity at one retinal location does not double the early local signal from that location. Thus the sum of the early local signals from two spots of a given intensity will be greater than the early local signal from one spot of twice that intensity. Easter suggested that the nonlinear early local transformation was a square-root function of light intensity and this suggestion has been supported, at least above threshold, by Levine (1972). (As Easter discussed in some detail, these results could also be explained by an interaction nonlinearity at the center combination point, although some possible interaction nonlinearities could be ruled out by another experiment.)

This same two-spot experiment done in the receptive field centers of cat retinal ganglion cells has yielded results similar to those described above for the goldfish (Büttner & Grüsser, 1968; Grüsser, 1971; Stone & Fabian, 1968). And another kind of experiment using single stimuli of varying area also provided evidence that, for cat retinal ganglion cells, distributing center effec-
tive flux over a wide area was more effective than concentrating it in a small area (Creutzfeldt et al., 1970). These results can be explained, as for the goldfish, by a compressive early local nonlinearity (which can be followed by a simple nonlinearity at the center combination point). Thus there is a discrepancy between these studies and those discussed earlier (Cleland & Enroth-Cugell, 1968) in which as long as center effective flux was constant, whatever the spatial arrangement, the response was constant. The cause of the discrepancy is not clear. However, it may be due to differences in the range of luminances and responses in the various studies: The early local transformation may be linear for low values although nonlinear for high. Indeed, as is particularly clear in the studies by Creutzfeldt et al. (1970) on the cat and Levine (1972) on the goldfish, the evidence for early local nonlinearity is greatly reduced as the response decreases toward threshold.

In connection with the early local nonlinear (square-root) transformation shown in the center mechanism of the goldfish retinal ganglion cell, an odd coincidence appears to exist: The compressive nonlinearity is exactly balanced out by the spatial distribution of sensitivity thereby producing apparently linear spatial summation. Levine (1972) and Abramov and Levine (1974) have shown not only that the early local transformations are nonlinear, but also that they are not identical across the center of the receptive field. Unlike the cat center mechanism (Cleland & Enroth-Cugell, 1968), the goldfish center mechanism is more responsive to light shining near the midpoint of the receptive field center than to light shining further away from the midpoint (Levine, 1972; Abramov & Levine, 1972; Easter, 1968a, as reinterpreted by Levine, 1972). This spatial distribution of responsiveness exactly balances out the early local square-root transformation for the case of circular stimuli centered in the receptive field, so that for these stimuli, \( I \times A_0 \) (intensity times area of overlap between the stimulus and the receptive field center) is exactly proportional to the summed inputs to the central combination point \( \sum f_i \). (In the case of an arbitrary stimulus, the summed inputs \( \sum f_i \) must be calculated using the fact that the \( f_i \) are proportional to the square root of intensity and that the constant of proportionality varies for different retinal locations \( i \).) And so, by this odd coincidence, whenever \( I \times A_0 \) is held constant for centered circular stimuli, the response of the center mechanism should be constant (Ricco's law). That it is constant has been found by Easter (1968a) and Levine (1972).

Early local nonlinearity for a surround mechanism has been demonstrated for at least one kind of cell: the \( Y \) type cat retinal ganglion cell. For five \( Y \) cells studied, Cleland, Levick, and Sanderson (1973) found that a steady

---

*Receptive field centers of almost all goldfish retinal ganglion cells are about the same size, 1 mm in diameter or about 12° of visual angle.*
adapting light shining on one part of the receptive field surround decreased the magnitude of the response to a test stimulus illuminating that same part much more than it decreased the response to a test stimulus illuminating some other part of the surround. A similar result was found for some, but not all X cells.

In addition to the nonlinear gain-changing effects, introduced by raising the light intensity, that seem to occur at both the early local and separate combination point stages, several other kinds of evidence suggest further complications in the mechanism of the retinal ganglion cell. Since these complications may well be due to processing prior to the separate combination points of center and surround, two of them are discussed briefly here: local adaptation pools, and early center-surround interaction.

_Local adaptation pools._ A major modification of the general model of Figure 18 is probably required to describe the goldfish retinal ganglion cells. In these cells (Easter, 1968b), steady background light on one retinal spot decreases the cell's response to an increment of light at that spot. It also decreases the response at neighboring spots, but more for nearby ones than for distant ones (when all spots are within the center and equated for sensitivity). This result cannot be attributed to an early local nonlinearity because the background light affects some neighboring spots and not just the spot it directly shines on. Nor can this result be attributed to the simple nonlinearity at the center combination point because the effect of the background light does not extend uniformly to all points within the center of the receptive field. This result could be explained by 'local adaptation pools' such as those shown in Figure 21. In this modification, the early local signal that originates at one point on the retina is affected (before arriving at the center combination point) by a signal $A_1$ from a 'local adaptation pool.' The signal $A_1$ is determined by the light in an area around the point producing the early local signal—an area usually smaller than the center of the receptive field. (Note that it is possible that the square-root transformation assigned to the early local path is actually a manifestation of the local adaptation pool.) Further information relevant to the characteristics of the local adaptation pools in the goldfish can be found in Easter (1968b). 4

Are there 'local adaptation pools' in the cat? The results of Cleland and Enroth-Cugell's (1970) experiment varying the area and intensity of a background stimulus suggest not. However, adaptation experiments using separate small spots at various distances have not yet been done in the cat and

---

4 Similar local adaptation pools have been found in the frog by Burkhardt and Berntson (1972). These authors have also suggested a function for the local adaptation pools—the pools may account for the extraordinary sensitivity of certain cells (quite common in the frog) to moving as compared to stationary stimuli.
the temporal parameters used in the cat experiments generally have been quite different from those used in the goldfish.

*Center-surround interactions.* Maffei and his coworkers have suggested that light stimulating the center mechanism of the cat retinal ganglion cell may affect the gain or adaptation level of the surround mechanism, and vice versa (Maffei, 1968; Maffei & Cervetto, 1968; Maffei, Cervetto, & Fiorentini, 1970; Maffei, Fiorentini, & Cervetto, 1971). In the case of influence of the surround mechanism on the gain of the center mechanism, this suggestion is not supported by the results of Cleland and Enroth-Cugel (1988). And some of Maffei and his co-workers' results might be explainable without such interaction if the responses that they attributed to only one mechanism were actually mixed responses coming from both mechanisms. However, it seems difficult to explain all their findings in that fashion, particularly those sug-
gesting that the responsiveness of the surround mechanism is affected by activity in the center mechanism (Maffei, Fiorentini, & Cervetto, 1971).

3.5. Temporal Properties

Since each stage in the retinal ganglion cell mechanism has its own temporal characteristics, the overall temporal properties of the cell may be quite complicated. Figure 20 has already illustrated the complexities in the responses to time-varying stimuli that can arise just from linear addition of center and surround responses. The earlier stages appear to add even greater complexity.

One major experimental result is that the temporal characteristics of the retinal ganglion cell responses change as adaptation level is changed. The latency of the response to a flash or an increment decreases as background intensity is raised (Cleland & Enroth-Cugell, 1970; Levine & Zacks, 1970; Bickel, 1965). At low luminance levels, the response to an increment of luminance in the receptive field center rises to a maximum shortly after the increment and stays there, but at high luminance levels, the response rises to a peak and then decays exponentially to a maintained steady state (Shapley et al., 1972; Stone & Fabian, 1968; Yoon, 1972). See Figure 17. As might be expected from these responses to increments, the transfer function relating peak-trough amplitude in the cell's response to temporal frequency of sinusoidal flicker (for stimuli either in the center or in the surround of the field) changes with mean luminance level, exhibiting a low-frequency decline at high mean luminances (the peak-trough amplitude in the response to low frequencies is smaller than that in the response to medium frequencies) but no low-frequency decline at low mean luminances (Schellart & Spekreijse, 1972; Enroth-Cugell & Shapley, 1973a). Actually, rather than using stimuli flickering sinusoidally, Schellart and Spekreijse (1972) introduced a new method of determining the temporal transfer function indirectly by measuring the cell's response to a visual white-noise stimulus. The transfer functions at various mean luminances for cat and goldfish retinal ganglion cells look remarkably like the comparable Limulus electrophysiological and human psychophysical transfer functions.

Shapley et al. (1972), Enroth-Cugell and Shapley (1973b), and Schellart and Spekreijse (1972) all present evidence that these changes in the temporal characteristics of the responses depend on luminance summed appropriately over area (on the quantity theoretically equivalent to the summed inputs to the center combination point) rather than on any other property of the stimulus. This suggests that much of the time-course of the response is determined by the gain-changing nonlinearity at the central combination point rather than by the early local stages (although the early local stage will also influence the time-course). Enroth-Cugell and Shapley (1973a) have accord-
ingly proposed a model for the center combination point of the cat retinal ganglion cell, which explains simultaneously the gain-change (the different magnitudes of response for an incremental stimulus superimposed on different background levels) and the different temporal characteristics of the response for different background levels. They suggested the following equations for the response of the center mechanism $g_c(t)$:

$$g_c(t) = \frac{A(t)}{\exp(B(t))};$$

$$A(t) = \int F_c(t - \tau) \left( \frac{\tau}{\tau_p} \right)^n e^{-\tau/\tau_p} d\tau;$$

$$B(t) = \int k_g(t - \tau) \frac{e^{-\tau/\tau_h}}{\tau_h} d\tau.$$

$A(t)$ is the summed input to the center combination point calculated from $F_c(t)$, the center effective flux at time $t$, where $n$ and $\tau_p$ are constants. $B(t)$ is a feedback signal which is equal to the center mechanism's response averaged over past time with an exponential weighting function specified by the constants $k$ and $\tau_h$. A limitation of this model is that it does not predict the result for the cat retinal ganglion cell that is embodied in Equation 6, which shows that the response to an increment is exactly proportional to the incremental flux ($\Delta F$) as long as incremental plus background flux ($\Delta F + F$) is held constant. Nevertheless, this model is an important advance because it describes both the temporal characteristics and the nonlinear characteristics of the ganglion cell responses simultaneously, and does successfully predict the changes in the magnitude and time-course of the response to an incremental stimulus as background luminance is changed and also predicts some of the changes in the transfer functions at different mean luminances.

The extent of this linkage between gain-changing nonlinearity and temporal properties may be a characteristic distinguishing $X$ cells from $Y$ cells (see next section): according to Enroth-Cugell and Shapley (1973a), the linkage may be more dramatic for $Y$ cells (which are the type of cells from which they present results) than for $X$ cells.

A question often raised is whether the temporal properties of the center and the surround mechanism are the same. For stimulus spots of similar size and luminance, the center mechanism of the cat retinal ganglion cell can follow flicker out to higher temporal frequencies than can the surround mechanism (Maffei, Cervetto, & Fiorentini, 1970). If the luminance and areas of the stimuli are chosen appropriately to compensate for the different sensitivities of the two mechanisms, however, this difference between the two mechanisms is not found, at least not in the goldfish. Rather the temporal transfer functions for center and surround mechanisms can always be made.
identical except for a phase difference indicating some delay of the surround response with respect to the center response (Schellart & Spekreijse, 1972). Thus the temporal properties of the center and surround mechanisms are the same, except for this delay, as long as the mechanisms are operating at the same adaptation level.

Similarly, the temporal transfer functions for the two opposing color-coded processes found in the center mechanism of the goldfish retinal ganglion cell can be made equal except for a phase difference indicating a delay of the ‘green’ process (the process with maximum sensitivity to the middle wavelengths) with respect to the ‘red’ process (the process with maximum sensitivity to the long wavelengths) (Schellart & Spekreijse, 1972).\footnote{Bickling (1965) proposed a very interesting model of goldfish retinal ganglion cells. It is quite different from the models we have been discussing in that it makes complex assumptions about the actual mechanism for generating impulses from an underlying continuously varying slow potential (rather than just assuming that impulse frequency is proportional to the magnitude of some underlying continuously varying response function $\theta$). It predicts quite well certain temporal characteristics, a number of changes with stimulus intensity level, and some aspects of the variability among responses to the same stimulus.}

3.6. $X$ (Sustained) and $Y$ (Transient) Cells

Recently two types of cells have been discovered in the cat retina (Enroth-Cugell & Robson, 1966; Cleland, Dubin, & Levick, 1971; Fukada, 1971; Fukada & Saito, 1971) and perhaps also in the primate retina (Gouras, 1968, 1969). It has been proposed that the existence of these two types of cells may reveal a basic division of visual function: $X$ cells (also called sustained because they respond throughout a maintained stimulus) being specialized for spatial information, and $Y$ cells (also called transient because they tend to respond only at the beginning of a maintained stimulus) specialized for temporal information (Cleland, Dubin, & Levick, 1971; Fukada & Saito, 1971).

An important difference between these cell types is revealed in their responses to the onset and offset of a stationary sinusoidal grating (when the grating is not present, the retina is illuminated uniformly at the average luminance of the grating). Such a sinusoidal grating is pictured in Figure 22, along with $X$ and $Y$ cell responses to four placements of the grating. Consider what happens when the grating is so placed with respect to the cell’s receptive field that as much area receives an increase in illumination as receives a decrease in illumination when the grating is turned on and off (90° and 270° phase positions in Fig. 22, situations in which the center effective flux and surround effective flux remain constant). $X$ cells did not respond at either the onset or the offset of the grating. $Y$ cells, however, showed a tran-
sient increase in firing at both onset and offset (Enroth-Cugell & Robson, 1966). An analogous result using two small spots flickering in antiphase has been demonstrated (Büttner, Büttner, & Grüsser, 1971). A further experiment yielding similar results was reported by Cleland, Levick, and Sanderson (1973) who used radially symmetric patterns of alternating black and white sectors (like windmills) centered on cells' receptive fields. No matter how such a pattern is rotated around its center, the effective flux stimulating the center of the receptive field and that stimulating the surround remains constant (assuming a circular receptive field). Consistent with the results of Figure 22, X cells' discharge rates were never affected by rotation of such a pattern but Y cells always responded to small jerky movements of the pattern with brief but definite bursts of impulses.

This behavior of the X cells in response to gratings and windmill patterns can be predicted by a model with linear early local transformations (all having the same time-course) followed either by a linear transformation or by a simple nonlinear transformation at the center and surround combination points (and with relative responsiveness of both the center and surround mechanisms assumed spatially symmetric as always appears to be the case). For, as a result of the linear early local stage, all increases in inputs to the separate combination points would be balanced out by exactly compensating decreases at both the onset and the offset of the stationary grating (at 90° and 270° phase positions) in the experiment of Figure 22 or whenever the radially symmetric windmill pattern was rotated. Hence this model with linear early local processing predicts that the responses of both the center and surround mechanisms and consequently the response of the cell should remain constant during the onset and offset of the appropriately placed stationary grating and during any rotation of the radially symmetric pattern as does the X cell response. The compressive early local nonlinearity suggested by two-spot summation experiments might not have shown up in these experiments because the luminance was being varied within a relatively small range.

The response of Y cells to the gratings and windmill patterns can be predicted by a model with early local transformations that are nonlinear even for small ranges of luminance variation (followed either by linear or by simple nonlinear transformations at the separate combination point). The Y cells' response might also be predicted if the early local transformations were linear, but their time-courses were not identical and not symmetrically distributed around the center of the field, for in that case the time-course of the response to the decreases in illumination would be different from the time-course of the response to the increases in illumination. (As mentioned in connection with the X cells, the compressive early local nonlinearity suggested by the two-spot summation experiments might not show up in this situation because luminance is being varied within a relatively small range.)

It appears unlikely, however, that either of these explanations involving
early local transformations is a complete description of the processes differentiating $Y$ cells from $X$ cells. For example, when the spatial frequency of a moving grating is varied but the temporal frequency is held constant (the number of bars per second passing each point is held constant), the average discharge rate of $X$ cells is the same for all spatial frequencies but the average discharge rate of $Y$ cells depends heavily on the spatial frequency (Enroth-Cugell & Robson, 1966; Cleland, Dubin, & Levick, 1971). The difference between $X$ cells' and $Y$ cells' behavior as spatial frequency is changed suggests that some variations in the processes of spatial summation (the processes at the combination points in terms of the general model) are also involved in differentiating the two types.

Whatever the fundamental differences between $X$ cells and $Y$ cells, the differences may well involve the surround and not the center mechanism, since $Y$ cells respond to moving radially symmetric patterns are not changed very much if the center five degrees of the pattern is covered (Cleland, Levick, & Sanderson, 1973). And, as mentioned before, the effect of a steady adapting background on the receptive field surround is always localized for $Y$ cells but can be either localized or not for $X$ cells. Further, $Y$ cells but not $X$ cells exhibit the periphery effect: stimuli moving in a part of the visual field far outside the usual limits of the cell's receptive field cause an increase in the firing rate of the cell (Cleland, Dubin, & Levick, 1971).

There are many further differences between the two types of cells. $X$ cells tend to have smaller receptive fields than $Y$ cells and their receptive fields tend to be closer to the center of the eye (Enroth-Cugell & Robson, 1966; Cleland, Dubin, & Levick, 1971). $X$ cells tend to have higher maintained discharge rates than $Y$ cells (Enroth-Cugell & Robson, 1966; Cleland, Levick, & Sanderson, 1973). For $X$ cells, the thresholds for different-sized spots of light and for annuli with different inner diameters (centered on the receptive field) show clear evidence of opposing center and surround mechanisms, but for $Y$ cells, the thresholds do not (Cleland, Levick, & Sanderson, 1973). $Y$ cells can respond to objects moving much faster than those $X$ cells can respond to (Cleland, Dubin, & Levick, 1971). As mentioned earlier, the dynamics of the $Y$ cells may be more dependent on the level of light adaptation than those of

**Figure 22.**

On the left is a photograph of a sinusoidal grating pattern and a sketch of its luminance profile with its mean intensity indicated by a horizontal line. On the right are shown responses of an off-center $X$ (sustained) and an off-center $Y$ (transient) cat retinal ganglion cell to the onset and offset of a stationary sinusoidal grating pattern. Downward deflection of the lowest tracings indicates offset of the grating and upward deflection indicates onset. The grating was turned on and off at 0.43 c/s. (When it was off, the field was uniformly illuminated at the mean intensity of the grating.) The position of the grating relative to the midpoint of the receptive field center was varied as sketched at the right of the figure. (Adapted from Enroth-Cugell & Robson, 1966.)
X cells (Enroth-Cugell & Shapley, 1973a). The conduction velocity of impulses traveling up the axons of the retinal ganglion cells is faster for Y cells than for X cells (Cleland, Dubin, & Levick, 1971; Fukada, 1971). Lastly, this division of cells into X and Y types is also found in the lateral geniculate nucleus; the X type lateral geniculate neurons receive all their inputs from X type retinal ganglion cells whereas the Y type lateral geniculate neurons receive all their inputs from Y type retinal ganglion cells (Cleland, Dubin, & Levick, 1971).

3.7. Summary

Many features of the responses of cat and goldfish retinal ganglion cells can be quantitatively described using a specific version of the general model shown in Figure 18. The signals from the center and surround mechanisms are linearly combined (and then perhaps rectified) at the final point. The signals from different points on the retina are combined by a 'simple' nonlinear transformation at the combination point for the center mechanism and perhaps also by a similar transformation at the combination point for the surround mechanism; these nonlinear transformations reduce the gain or sensitivity of the cell as light intensity is raised. The early local transformation is a compressive nonlinear function that also reduces the sensitivity of the cell as light intensity is raised.

This version of the model of Figure 18 must certainly be elaborated, however, in order to explain a number of other characteristics of the responses of these retinal ganglion cells. The differences between the X cells and Y cells must be explained. Temporal characteristics, only partially specified in the above, must be fully described, including the variability of the responses (which has been quantitatively studied although not described here, see Barlow & Levick, 1969a, 1969b). Wavelength selectivity (particularly in the goldfish; see Spekreijse, Wagner, & Wolbarsht, 1972, for example) must also be incorporated. Further elaborations might include local adaptation pools and an influence of the center mechanism on the surround mechanism's responsiveness.

4. CONCLUDING REMARKS

The optic nerve is a bottleneck through which must pass all visual information that reaches the central nervous system. Retinal ganglion cells (eccentric cells in Limulus) are thus located at a critical point in the visual system. On the one hand their behavior represents the outcome of the integration of the activity of many individual intraretinal components, and on the other hand
this very same behavior represents the individual inputs into an even more complex integrative system—the brain.

In elucidating the intraretinal mechanism, knowledge of the retinal ganglion cell’s behavior indicates the kinds of mechanisms to be looked for within the retina. Also, any presumed understanding of how these intraretinal processes form a chain of events leading from light to ganglion cell response can be tested by the extent of its agreement with the known behavior of the retinal ganglion cell. In Limulus the discovery of the underlying physiological processes represented in the equivalent circuit model was greatly facilitated by knowing the characteristics of optic nerve axons’ responses, which suggested, for example, the existence of self-inhibition and indicated the time-course it must have. In cat and goldfish the correspondence between the actual intraretinal processes and the various processes represented abstractly in the general model of retinal ganglion cells has not yet been determined. There are, by a conservative count, four kinds of neurons in the vertebrate retina that are intermediary between light and retinal ganglion cells (receptors, horizontal, bipolar, amacrine). In the last few years, with the development of appropriate microelectrodes, micromanipulators, and staining techniques, it has become possible to record from these cells and to discover many of their individual properties. Although no overall picture of the whole network feeding into a particular ganglion cell has yet emerged, work along these lines is progressing rapidly.

Knowing what happens at the retina may also simplify the task of understanding what happens higher in the visual system. First, it enables the input to these higher cells to be specified in a form that is closer to their actual input than is the original light stimulus. Second, basic processes that have been isolated and studied at the retinal level may be identifiable as subunits in the mechanism for higher order cells that are more complicated in their behavior and therefore more difficult to analyze. For example, cells responding maximally to moving lines of particular orientation (where the lines can be anywhere in a wide part of the visual field) are found in the cortex of the cat. Although not yet quantitatively understood, it appears that these cells as well as other higher order cells integrate opposed excitatory and inhibitory influences in a manner similar to that of the simpler retinal ganglion cells. Moreover, it is likely that the mechanisms of these higher order cells in the cat are not unlike those of the complicated retinal ganglion cells found in many other vertebrate species. In fact, the retinal ganglion cells of cat and goldfish are simple compared with the retinal ganglion cells in many lower vertebrate species (frog, rabbit, pigeon), although apparently quite similar to those in most primates.

Visual perception depends not on the activity of any one cell alone, but on the integrated activity of the whole visual system. Under proper stimulus conditions, however, the action of one type of cell may be emphasized relative
to that of the others. Various perceptual phenomena are known that seem indeed to reflect the characteristics of particular types of visual neurons. This correspondence has encouraged many investigators to suggest explanations for perceptual phenomena based on visual neurophysiology. The more quantitative the description of both the perceptual phenomena and the visual neurophysiology, the more likely the success in determining the actual relation of one to the other.

ACKNOWLEDGMENTS

The preparation of this chapter was supported in part by NSF Grant GB-6540 (P2B1569) and NIH Grants EY-188 and GM-1789. We are grateful to Donald C. Hood, R. Allison Ryan, and Robert M. Shapley for their careful reading and helpful criticism of early drafts of this chapter.

REFERENCES

Abramov, I., & Levine, M. W. Spatial summation in the receptive fields of goldfish ganglion cells, 1974, in preparation.


Adrian, E. D., & Matthews, R. The action of light on the eye. Part II. The processes involved in retinal excitation. J. Physiol., 1927, 64, 279–301. (b)


Hartline, H. K. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. Amer. J. Physiol., 1938, 121, 400-415. (a)


Hartline, H. K. The neural mechanisms of vision. The Harvey Lectures, 1941-42, Series 37, 39-68.


Hartline, H. K., & Ratliff, F. Spatial summation of inhibitory influences in the eye.


The Integrative Action of the Retina


Shapley, R. M. Fluctuations of the impulse rate in Limulus eccentric cells. J. Gen. Physiol., 1971, 57, 539–556. (a)

Shapley, R. M. Effects of lateral inhibition on fluctuations of the impulse rate. J. Gen. Physiol., 1971, 57, 557–575. (b)


