A Unified Neural Network Model of Spatio-Temporal Processing in X and Y Retinal Ganglion Cells. I: Analytical Results

*Biological Cybernetics, 67*, 11–22, 1992

Paolo Gaudiano

Boston University, Department of Cognitive & Neural Systems
111 Cummington Street, Boston, MA 02215 USA

Abstract

This work presents unified analyses of spatial and temporal visual information processing in a feed-forward network of neurons that obey membrane, or *shunting* equations. The feed-forward shunting network possesses properties that make it well suited for processing of static, spatial information. However, it is shown here that those same properties of the shunting network that lead to good spatial processing imply poor temporal processing characteristics. This article presents an extension of the feed-forward shunting network model that solves this problem by means of preprocessing layers. The anatomical interpretation of the resulting model is structurally analogous to recently discovered data on a retinal circuit connecting cones to retinal ganglion cells through pairs of push-pull bipolar cells. Mathematical analysis of the lumped model leads to the hypothesis that X and Y retinal ganglion cells may consist of a single mechanism acting in different parameter ranges. This hypothesis is confirmed in the companion article, wherein the model—in conjunction with a nonlinear temporal adaptation mechanism—is used to reproduce experimental data of both X and Y cells by simple changes in morphological and physiological parameters.

1 Introduction

The retina of all vertebrates is organized into three cellular (or *nuclear*) layers and two synaptic (or *plexiform*) layers (Dowling, 1987). Information flows from the photoreceptors, through a layer of bipolar cells, and finally through the retinal ganglion cells (RGCs), whose axons project via the optic nerve to subcortical and cortical areas. In addition to the feed-forward information processing carried out by these three cell types, there exist two classes of cells, horizontal cells and amacrine cells, that carry signals laterally through the retina for additional processing.

Although there is general agreement that retinal cells can be subdivided into these broad classes, morphological and anatomical studies have shown the existence of a great number of cell types within each class, including dozens of RGC and amacrine cell types (e.g., Kolb et al., 1981). In addition to morphological and anatomical classification schemes, physiological studies have shown the existence of broad functional classes. Kuffler (1953) and Barlow (1953) first reported the existence of two RGC classes: on-center (ON) and off-center (OFF) cells. Both types have a concentrically-organized receptive field (RF) consisting of a central region and a surrounding annulus of opposite polarity. Increasing luminance in the RF center causes increased activation for ON cells, and decreased activation for OFF cells. Conversely, stimulation of the RF surround decreases activation in the ON cells, and increases activation in OFF cells. The distinct response characteristics of ON and OFF RGCs appear to originate in the bipolar cell layer (Werblin and Dowling, 1969).

An additional, independent functional classification scheme was proposed by Enroth-Cugell and Robson (1966), who reported the existence of two RGC classes on the basis of spatiotemporal response characteristics: X cells, which respond to inputs in a sustained fashion, and appear to linearly summate luminance signals throughout their RF; and Y cells, which respond to inputs in a transient fashion, and exhibit a more complicated, nonlinear spatial summation of luminance signals throughout their RF. This classification has been supported and extended through a number of physiological and anatomical studies, and a relationship has been established between the morphological classes of alpha and beta ganglion cells, respectively, and the functional classes of Y and X cells (Boycott and Wässle, 1974; Cleland and Levick, 1974; Fukuda et al., 1984; Hochstein and Shapley, 1976a,b; Saito, 1983).

The correspondence between functional and anatomical classes has supported the notion that visual processing is highly modular, with different circuits designed to carry out different tasks independently. This notion is reflected in the typical segregation between network models for static spatial processing and single-cell or low-dimensional models.
for dynamic temporal processing. The work presented here bridges this gap by simultaneously analyzing spatial and
temporal processing in a network of neurons that obey shunting equations. The analyses lead to formulation of a new
model designed to carry out accurate spatial and temporal processing simultaneously.

The new model is structurally analogous to the feed-forward circuit connecting photoreceptors to RGCs through
opponent pairs of bipolar cells, and suggests that X and Y cells may consist of the same mechanism acting in different
parameter ranges. This hypothesis is more fully developed in a companion article (Gaudiano, 1992), in which the model
is used to simulate several fundamental properties of both X and Y RGCs. The model also clarifies a paradox found
in the experimental literature by showing that X cells may exhibit linear behavior in spite of nonlinear photoreceptor
preprocessing.

2 The Feed-Forward Shunting Network

The feed-forward shunting network consists of neurons whose activation is described by first-order, time-dependent
differential equations wherein excitatory and inhibitory processes act upon distinct cellular mechanisms before competing
for dynamic activation of the neuron. Historically, the term shunt as applied to neural networks refers to the presence of
multiplicative saturation terms in the equation describing membrane potential (Furman, 1965; Grossberg, 1970, 1973;

The models of Furman (1965) and Sperling (1970) use shunting only for inhibitory interactions, whereas Grossberg’s
models (1970, 1973), on which the present work is based, assume that both excitatory and inhibitory processes act
multiplicatively on membrane polarization. The simplest differential equation describing an on-center, off-surround
shunting network (Fig. 1a) can be written as follows:

\[ \frac{dv_i}{dt} = -A(v_i - v_e) + (B - v_i)I_i - (D + v_i)\sum_{k \neq i} I_k. \]

Here \( v_i \) represents the activation or potential of the \( i \)th cell in the network; \( A \) is the rate of passive decay toward
the resting potential (which is assumed here to be zero); \( B \) and \( D \) represent the excitatory and inhibitory saturation
points, respectively; \( I_i \) represents the excitatory contribution from the \( i \)th component of the input pattern to the \( i \)th cell
(on-center); and \( \sum_{k \neq i} I_k \) represents the inhibitory contribution from all other components of the input pattern to the \( i \)th
cell (off-surround). All parameters are assumed to be nonnegative. In the absence of inputs, activation \( v_i \) decays to zero
at a rate \(-Av_i\). Otherwise, (1) guarantees that activation is always bounded between the values \( B \) and \(-D\), regardless
of input intensity. Through a simple change of variables, (1) can be interpreted as a passive membrane equation (e.g.,
Grossberg, 1988, p. 35), in which case the inputs modulate membrane conductance of two ionic species which have
opposite effects on membrane polarization.

The behavior of network (1) has been analyzed through various mathematical techniques. Details on the properties
of shunting networks and their ability to model visual processes can be found elsewhere (Grossberg, 1970, 1983, 1988;
Koch et al., 1983; Pinter, 1985; Sperling, 1970). Many properties of this type of network are derived by analysis of (1)
at steady-state, since the system converges. The equilibrium potential of the \( i \)th cell in the network is found by setting
\( dv_i/dt = 0 \):

\[ v_i = \frac{BI_i - D\sum_{k \neq i} I_k}{A + \sum_{k} I_k}. \]

In the area of static visual processing, shunting network properties include faithful transmission of relative luminance
levels in a complex scene, Weber-law sensitivity, and the ability to suppress uniform backgrounds. Equation (2) can be
used for example to show that the shunting network is able to represent the relative luminance at different positions in
an image (local contrast processing) regardless of fluctuations in the overall input intensity. Denoting the total input by
\( I = \sum_k I_k \), and the relative input intensity at position \( i \) by \( \theta_i = I_i/I \), equation (2) can be rewritten as

\[ v_i = \frac{(B + D)I}{A + I} \left( \theta_i - \frac{D}{B + D} \right). \]

As long as the passive decay is small \((A \ll 1)\), equation (3) factorizes information about relative input size \( \theta_i \) from
overall intensity \( I \), because activation \( v_i \) is approximately proportional to \( \theta_i - D/(B + D) \) regardless of overall input intensity
\( I \). Grossberg (1980) has also shown that the region of maximal sensitivity of a cell obeying (1) shifts without compression
as the background intensity is parametrically increased. This is known as the shift property, which has been demonstrated
experimentally for certain classes of retinal cells (Werblin, 1971).
Similar results ensue when the center-surround mechanism consists of distinct, overlapping center and surround components, as shown in Fig. 1b. Equation (1) then becomes

\[
\frac{dv_i}{dt} = -A v_i + (B - v_i) \sum_k \mathcal{C}(k - i) I_k - (D + v_i) \sum_k \mathcal{S}(k - i) I_k.
\]  

(4)

The distance-dependent terms \(\mathcal{C}(k - i)\) and \(\mathcal{S}(k - i)\) represent the center and surround mechanisms of the receptive field (RF) profile of each cell in the network. Most of the general results reported here are insensitive to the exact shape of the center and surround mechanisms. Where explicit closed-form results are required these terms are described as Gaussians:

\[
\mathcal{C}(k - i) = C \exp \left[-\frac{(k - i)^2}{2\sigma_C^2}\right]; \quad \mathcal{S}(k - i) = S \exp \left[-\frac{(k - i)^2}{2\sigma_S^2}\right].
\]  

(5)

Each Gaussian is described by its peak amplitude \((C, S)\) and standard deviation \((\sigma_C, \sigma_S)\). The overall RF shape is composed of the antagonistic excitatory center and inhibitory surround Gaussians, and is thus referred to as a Difference-Of-Gaussians (DOG) profile. Throughout this article it is assumed that \(\sigma_C < \sigma_S\) and \(C > S\), which is representative of an on-center, off-surround anatomy, though all results generalize to off-center on-surround anatomies as well. Parameters can be found in the companion article (Gaudiano, 1992).

It should be noted that interactions between the RF center and surround mechanisms are mediated by the shunting dynamics, so that the “effective” RF profile—for example, an outline of the cell’s response to pointwise stimulation as a function of spatial location—depends on the inputs and on the cell’s activation, and may deviate from a true DOG. This stands in contrast with more typical DOG models (e.g., Rodieck, 1965), wherein the output of the cell results from linear operations on a DOG kernel, so that the response profile generally remains a DOG regardless of inputs or cell activation.

3 Spatiotemporal Analysis

The properties of feed-forward shunting networks outlined thus far are based on steady-state visual processing. However, most experimental data on visual system response are based on input functions that are modulated temporally as well as spatially. In order to assess the validity of the shunting network as a model of retinal circuitry, it is possible to study its properties for example sinusoidally or square-wave modulated spatial gratings, and drifting sinusoidal gratings.

Substitution of (7) into (6) and simplification leads to

\[
\frac{\partial v(x, t)}{\partial t} = -Av(x, t) + [B - v(x, t)] \int_{-\infty}^{\infty} \mathcal{C}(x - \xi) I(\xi, t) d\xi -
\]

\[
[D + v(x, t)] \int_{-\infty}^{\infty} \mathcal{S}(x - \xi) I(\xi, t) d\xi.
\]  

(6)

Analytical solutions can be found when the input function is space-time separable, i.e., when it can be expressed as the product of a spatial modulation function \(I(x)\) and a temporal modulation function \(m(t)\)

\[
I(x, t) = I(x) m(t).
\]  

(7)

Many of the stimuli used for experimental measurements can be expressed as space-time separable functions, including for example sinusoidally or square-wave modulated spatial gratings, and drifting sinusoidal gratings.

Substitution of (7) into (6) and simplification leads to

\[
\frac{\partial v(x, t)}{\partial t} = -Av(x, t) + [B - v(x, t)] m(t) I_C - [D + v(x, t)] m(t) I_S.
\]  

(8)

The terms

\[
I_C = \int_{-\infty}^{\infty} \mathcal{C}(x - \xi) I(\xi) d\xi \quad \text{and} \quad I_S = \int_{-\infty}^{\infty} \mathcal{S}(x - \xi) I(\xi) d\xi
\]  

(9)

represent the convolution of the spatial distribution \(I(x)\) with the RF center and surround mechanisms. The notation is meant to emphasize the time-independent nature of the spatial convolution, which allows (8) to be rewritten as a first-order, linear, nonhomogeneous, time-varying, ordinary differential equation:

\[
\frac{dv}{dt} + [A + m(t)(I_C + I_S)] v = m(t)(BI_C + DI_S).
\]  

(10)
Particular choices of \( m(t) \) yield closed-form solutions for specific initial conditions, although often the solution must be left in integral form.

4 Response to Whole-Field Sinusoidal Modulation

Standard techniques can be used to analyze the response of (10) to whole-field sinusoidal temporal modulation of an arbitrary spatial distribution function. The temporal modulation function is

\[
m(t) = c_1 + c_2 \sin \Omega t
\]

where \( c_1 \geq c_2 \) to ensure that inputs cannot become negative,\(^1\) and \( \Omega \) is the temporal frequency of modulation. This form of whole-field modulation differs from typical counterphase-flicker modulation because the lack of negative values prevents contrast reversals (see Fig. 3 below).

An exact closed-form solution can be found if \( A = 0 \):

\[
v(x, t) = \frac{B I_C - D I_S}{I_C + I_S} - K \exp \left( -at + \frac{b}{\Omega} \cos \Omega t \right).
\]

For an arbitrary initial value \( v(x, 0) = v_0 \), the integration constant is

\[
K = \left( \frac{B I_C - D I_S}{I_C + I_S} - v_0 \right) \exp \left( -\frac{b}{\Omega} \right).
\]

Equation (12) predicts that \( v(t) \) equilibrates to a constant value that only depends on the convolution terms \( I_C \) and \( I_S \). In other words, the asymptotic response of each cell depends only on the spatial distribution of the input pattern and the RF profile, while all temporal information is lost. This is due to the fact that the term \( K \exp \left( -at + \frac{b}{\Omega} \cos \Omega t \right) \) in (12) approaches zero as \( t \to \infty \), so that the entire right-hand side of (12) approaches the constant \( (B I_C - D I_S)/(I_C + I_S) \).

Because the shunting network (10) is a linear system, this observation can be extended to any whole-field temporal modulation that can be expressed as a sum of sinusoidal components.

Fig. 2 illustrates the activity \( v(t) \) of a cell in response to whole-field sinusoidal temporal modulation of a spatial sinusoidal grating. The top trace shows the temporally modulated input. The bottom trace shows that the cell integrates from its initial condition \( (v_0 = 0) \) to a new equilibrium level without subsequent temporal modulation. This follows figures are normalized to the activation range \([-D, B]\). Likewise, inputs are in non-dimensional form.

The spatiotemporal response of a 1-D population of shunting cells is shown in Fig. 3. The top surface represents the input distribution, while the bottom surface represents the population response. The population consists of 100 cells aligned along the \( x \) axis, with time evolving along the \( t \) axis, and activation level (or input) represented as magnitude along the \( v \) (or \( I \)) axis. The surfaces in Fig. 3 are thus equivalent to combining traces such as those of Fig. 2 for one-hundred adjacent cells. Equivalently, the lower surface can be thought of as the response of a single cell to inputs of varying spatial phase, in which case each line represents the response of the same cell as the grating is shifted in space relative to the cell’s RF center. This relationship is depicted by the thick lines drawn on each surface, representing the input (top) and response (bottom) of a cell aligned at zero spatial phase with the grating (equivalent to the individual traces of Fig. 2). The response profile of Fig. 3 shows that the network is able to process information on the spatial input distribution, even though each individual cell is unable to track whole-field temporal modulation.

The network’s inability to track temporal modulation is a direct result of the multiplicative terms that yield factorization of relative contrast from overall intensity: when the spatial distribution is whole-field modulated and \( A = 0 \), the temporal modulation function \( m(t) \) only affects the rate of integration in (10), leaving the relative contrast unchanged.

A possible solution to this problem is to increase the passive decay \( A \) to one order of magnitude greater than the excitatory saturation point \( B \) results in only a minor improvement in response to whole-field modulation (solid lines). Furthermore, the large passive decay has the undesired side effect of hampering the network’s ability to discount overall input intensity from spatial patterns (cf., Section 2): an increase in background intensity results in poorer temporal processing (Fig. 4, dashed lines).

\(^1\)Negative inputs would correspond to negative conductances, which do not appear in a passive membrane equation. Also, in the context of retinal luminance processing, the input should be nonnegative.
5 The Retina and Shunting Networks

How can a network be designed whose good spatial processing characteristics do not imply poor temporal characteristics? The problem outlined in the previous section is the result of an asymmetry in the function of the shunting network. Namely, the nonnegative nature of input signals makes the network much more sensitive to input increments than decrements. This asymmetry can be corrected by means of two complementary input pathways to the network, one signaling luminance increments, the other signaling luminance decrements. Together, these pathways allow the shunting network to faithfully track temporal modulation.

As mentioned in Section 1, such complementary pathways are known to exist in the parallel ON and OFF retinal circuits (Barlow, 1953; Kufler, 1953); thus one possibility is that signals from the ON and OFF pathways interact to generate a response to increasing as well as decreasing inputs. Although some evidence exists for feedback signals within the retina, this does not seem like a tenable solution, particularly since the ON and OFF pathways have been shown to be largely functionally independent (Schiller, 1982; Slaughter and Miller, 1981).

Alternatively, a mechanism that acts feed-forward within each pathway can force the shunting cells to actively decay at input signal cessation. In addition to having the desired functional properties, I will show in this and the companion article (Gaudiano, 1992) that a model based on such a mechanism is able to fit anatomical, morphological, and physiological data on the circuit connecting cones, bipolar cells, and X and Y retinal ganglion cells.

Fig. 5a schematically illustrates such a mechanism. Each photoreceptor gives rise to equal and opposite signals through two bipolar cells: one whose activation is maximal when photoreceptor activation is maximal, and one whose activation is maximal when photoreceptor activation is minimal. Such parallel, push-pull pathways were recently found and physiological data on the circuit connecting cones, bipolar cells, and X and Y retinal ganglion cells.

Each photoreceptor gives rise to equal and opposite signals through two bipolar cells: one whose activation is maximal when photoreceptor activation is maximal, and one whose activation is maximal when photoreceptor activation is minimal. Such parallel, push-pull pathways were recently found and physiological data on the circuit connecting cones, bipolar cells, and X and Y retinal ganglion cells.

Fig. 5a schematically illustrates such a mechanism. Each photoreceptor gives rise to equal and opposite signals through two bipolar cells: one whose activation is maximal when photoreceptor activation is maximal, and one whose activation is maximal when photoreceptor activation is minimal. Such parallel, push-pull pathways were recently found and physiological data on the circuit connecting cones, bipolar cells, and X and Y retinal ganglion cells.

Alternatively, a mechanism that acts feed-forward within each pathway can force the shunting cells to actively decay at input signal cessation. In addition to having the desired functional properties, I will show in this and the companion article (Gaudiano, 1992) that a model based on such a mechanism is able to fit anatomical, morphological, and physiological data on the circuit connecting cones, bipolar cells, and X and Y retinal ganglion cells.

I next introduce a simple mathematical equation that embodies the proposed push-pull mechanism, and analyze its spatiotemporal response. Detailed derivation of all results presented here and in the companion article can be found elsewhere (Gaudiano, 1991).

With reference to Fig. 5a, let the input impinging upon each element of the shunting network be bounded between zero and a maximum level $M$. This constraint is valid on assumption that inputs to the shunting network originate from a layer of preprocessing elements—such as photoreceptors and bipolar cells—which possess a finite dynamic range. Each photoreceptor gives rise to two signals, one that affects the excitatory mechanism of the target RGC through a depolarizing bipolar cell, and one that affects its inhibitory mechanism through a hyperpolarizing bipolar cell. It is assumed that the same push-pull mechanism exists for cells comprising both the center and the antagonistic surround mechanisms. This does not preclude the possibility that center and surround mechanisms be mediated by different cell classes: for example, amacrine cells may mediate the surround mechanism, in which case it is assumed that these cells also receive parallel opponent inputs from push-pull bipolar cells.

If the bipolars act on a fast time scale relative to the photoreceptors, they can be lumped into equation (4) by addition of two terms to the input. Specifically, the excitatory input through the RF center $\sum C_i(k-i)I_k$ is complemented by an opponent inhibitory input $\sum C_i(k-i)(M-I_k)$, also through the RF center; likewise, the inhibitory surround input $\sum S(k-i)I_k$ is complemented by an opponent excitatory term $\sum S(k-i)(M-I_k)$. The resulting push-pull shunting equation is

$$\frac{dv_i}{dt} = -Av_i + (B - v_i) \left[ \sum C_i(k-i)I_k + \sum S(k-i)(M-I_k) \right]$$

$\text{Fig. 5 here}$
\[-(D + v_i) \left[ \sum S(k - i)I_k + \sum S(k - i)(M - I_k) \right], \quad (14)\]

where \(I_k \leq M\) for all \(k\). The equation shows that the center and surround each act on both excitatory and inhibitory channels, but in an opposing fashion.

Equation (14) can be rewritten as an integro-differential equation in space and time:

\[ \frac{\partial v(x, t)}{\partial t} = -A v(x, t) + [B - v(x, t)] \left[ \int_{-\infty}^{\infty} S(x - \xi)I(\xi, t)d\xi + \int_{-\infty}^{\infty} S(x - \xi)(M - I(\xi, t))d\xi \right] \]

\[ - [D + v(x, t)] \left[ \int_{-\infty}^{\infty} S(x - \xi)I(\xi, t)d\xi + \int_{-\infty}^{\infty} S(x - \xi)(M - I(\xi, t))d\xi \right] \quad (15)\]

Simplifying and rewriting in standard form leads to

\[ \frac{dv}{dt} + v(A + MC + MS) = m(t)(B + D)(Ic - Is) + BM - DMc, \quad (16)\]

where, as before (9), \(Ic\) and \(Is\) represent the convolution of an arbitrary spatial input distribution with the RF center and surround mechanisms, and

\[ MC = M \cdot Vc = M \int_{-\infty}^{\infty} S(x)dx \quad \text{and} \quad MS = M \cdot Vs = M \int_{-\infty}^{\infty} S(x)dx, \quad (17)\]

with \(Vc\) and \(Vs\) representing the volume (area) of the center and surround components, respectively, of the RF profile.

Equation (16) shows that addition of the push-pull mechanism has caused cancellation of all time-varying terms multiplying the dependent variable \(v\), thus reducing the network to a linear time-invariant (LTI) system. An LTI system such as (16) is easier to analyze than its time-varying counterpart (10), and leads to different predictions.

Under the general assumption of an arbitrary initial condition \(v(x, 0) \equiv v_0\), a general solution is found:

\[ v(x, t) = \frac{1}{a}(BM - DMc) \left[ 1 - e^{-at} \right] + e^{-at} \left[ (B + D)(Ic - Is) \int_{0}^{t} e^{\alpha \tau} m(\tau)d\tau + v_0 \right], \quad (18)\]

where

\[ a = A + MC + MS = A + M(Vc + Vs). \quad (19)\]

Equation (18) is easily solved for many different choices of temporal modulation function \(m(t)\). Note that (18) holds regardless of the dimension of the underlying network. The RF profiles, and thus the spatial input function \(I\), can be one- or two-dimensional without affecting the general solution. Similarly, (18) is not dependent on the specific choice of RF profile: the quantities \(A, MS\) and \(MC\) are constant for any given choice of RF profile and maximum bipolar activation \(M\).

### 7 Push-Pull Response to Whole-Field Modulation

To illustrate the effectiveness of the push-pull mechanism, I first analyze network response to the same spatiotemporal sinusoidal function that yields a saturating temporal response in the simple shunting network (Section 4). Substitution of (11) into (18) leads to the solution

\[ v(x, t) = v_0e^{-at} + \frac{1}{a} \left[ c_1(B + D)(Ic - Is) + BM - DMc \right] \left[ 1 - e^{-at} \right] + \]

\[ \frac{c_2(B + D)(Ic - Is)}{a^2 + \Omega^2} \left[ \Omega e^{-at} + a \sin \Omega t - \Omega \cos \Omega t \right], \quad (20)\]

with \(a\) as in (19). As can be seen from inspection of each term of (20), the network’s response exhibits transient as well as sustained temporal modulation, and does not equilibrate to a constant value as long as \(Ic \neq Is\). A simple trigonometric identity shows that

\[ a \sin \Omega t - \Omega \cos \Omega t = \psi \sin (\Omega t \pm \phi), \quad (21)\]

where

\[ \psi = \sqrt{a^2 + \Omega^2} \quad \text{and} \quad \tan \phi = \frac{\Omega}{a}. \quad (22)\]
which means that the temporal modulation in response to sinusoidal stimulation is also a pure sinusoid \(^2\) of the same fundamental frequency \(\Omega\), but scaled in amplitude by a constant \(\psi\), and phase shifted, or delayed, by a constant \(\phi\). Both of these constants depend jointly on the RF volumes and on the temporal frequency \(\Omega\).

Fig. 6 (lowest trace, solid line) shows the result of numerical integration of the discrete push-pull equation (14) when passive decay is zero, and all parameters are the same as those used to generate Fig. 2. As predicted, addition of the push-pull mechanisms enables the shunting cell to accurately follow temporal modulation. Fig. 7 shows the same result for a population of 100 push-pull shunting cells. All parameters are the same as the ones used in Fig. 3, but unlike the simple shunting population of that figure, the push-pull network is able to process both spatial and temporal information contained in the input.

### 8 A Unified Mechanism for X and Y Cells

There is one property of the general solution (18) that deserves special consideration. If the integral on the right of (18) remains bounded in time, then the equation can be seen as the combination of two terms: an asymptotically constant, or sustained term multiplied by a constant gain \(1/a\), and a transient term which decays at the rate \(e^{-\alpha t}\). The behavior of both terms depends critically on the constant \(a\). However, from (19) we see that the constant \(a\) depends on the size (the area for a 1-D profile, or the volume for a 2-D profile) of the receptive field. In particular, an increase in the size of the RF, other things being equal, has the following effects: (1) The amplitude of the sustained response decreases. (2) The network integrates more rapidly, i.e., response transients are sharpened.

The transition from more sustained to more transient response with an increase in RF size is suggestive of the functional and morphological distinction between the two main classes (X and Y) of cat retinal ganglion cells as first reported by Enroth-Cugell and Robson (1966). Ten years after the first report of X and Y cells, Hochstein and Shapley (1976b) confirmed and extended the X/Y classification through a variety of tests. Based on their results, these authors proposed that X cells perform linear summation of input signals throughout their center and surround RF mechanisms, but concluded that Y cell receptive fields include “three types of spatial component ... the conventional (linear) centre, the conventional antagonistic (linear) surround, and nonlinear ... subunits” (p.279).

Although the original work of Enroth-Cugell and Robson (1966) sparked a wave of qualitative and quantitative experiments on RGC response, and although a number of qualitative and quantitative RGC RF models have been proposed, the qualitative nonlinear receptive field model of Hochstein and Shapley (1976b) is still considered to be “the authoritative account of the Y-cell’s receptive fields” (Lennie et al., 1990, p. 112). The push-pull, distance-dependent, feed-forward shunting network (14) may instead provide a framework for a unified quantitative explanation of the behavior of both X and Y cells in response to spatiotemporally modulated inputs. Specifically, the interactions of the push-pull bipolar mechanism with the RF center and surround components may give rise to X-like response characteristics for small RF profiles, or Y-like response characteristics for larger RF profiles, in agreement with experimental data linking X cells with beta cells, and Y cells with alpha cells (Cleland and Levick, 1974; Fukuda et al., 1984; Saito, 1983). I will show, however, that certain general properties of RGC response depend on interactions with a nonlinear preprocessing mechanism.

### 9 Response to Input Onset and Offset

As a first step to motivate the above propositions, I now analyze the model’s response to a stimulus waveform that is more typically found in the experimental literature: the temporal step function. Let

\[
m(t) = c_1 + c_2 u(t), \quad u(t) = \begin{cases} 
0 & \text{if } t < 0 \\
1 & \text{if } t \geq 0
\end{cases}
\]  

(23)

where, as before, \(c_1 \geq c_2\). The definite integral in (18) is evaluated to yield the response function

\[
v(x, t) = \frac{1}{a} \left[ (c_1 + c_2)(B + D)(I_C - I_S) + B M_S - D M_C \right] \left[ 1 - e^{-\alpha t} \right] + v_0 e^{-\alpha t}.
\]  

(24)

Fig. 8 illustrates the response for the push-pull network to sudden input onsets and offsets. The top trace depicts the input at a single spatial location \((x = 0)\) as time evolves. The middle trace shows the response of a simulated X cell, and the bottom trace shows the response of a simulated Y cell. The simulation is started with \(v_0 = 0\), and the input.

---

\(^2\)The fact that the response is a pure sinusoid should be expected because, as pointed out in Section 6, the push-pull shunting network is an LTI system.
immediately turns on. Both cell types respond by integrating to a new equilibrium value. After a while the input is abruptly decreased, and the cells respond by exponentially decreasing to a new equilibrium level.

The only difference between the simulated X and Y cells is that the RF center of the former is smaller than that of the latter by a factor of 0.3, i.e., \( \sigma_Y^X/\sigma_Y^X = 0.3 \). As predicted in Section 8, the X cell integrates more slowly, and reaches higher steady-state values than the Y cell.

On the basis of the response profiles shown in Fig. 8, I forego detailed term-by-term analysis of the closed-form solution, and instead point out a discrepancy between the response of the model and the experimentally observed response of RGCs to step functions: although different classes of RGC respond in different fashions to input onset and offset, both X and Y cells respond to stimulus onset (offset) with sharp overshoots (undershoots) followed by a decay (rise) to an approximately constant level (for some examples, see Gaudiano, 1992).

Rather than ascribe this discrepancy to a shortcoming of the present model, I use it as a motivational tool to again argue that some processing of the input signal must take place prior to the input level of the shunting network.

In Section 6, I introduced the push-pull shunting formalism under the assumption that inputs to the network arise from a preprocessing layer whose output is bounded between zero and a constant value \( M \). This is the main reason to require preprocessing of the incoming luminance signals, as otherwise no guarantee can be made that the input is bounded in the range \([0, M]\).

Based on the observations of this section, the next qualitative property that I wish to ascribe to the preprocessing layers is the ability to generate transient overshoots and undershoots in response to sudden input fluctuations. Such overshoots and undershoots are found for example in photoreceptor responses (e.g., Baylor and Hodgkin, 1974; Carpenter and Grossberg, 1981). In the remainder of the present article, I will present further evidence in support of the claim that nonlinear preprocessing in the photoreceptors and bipolar cells is necessary to explain important aspects of RGC behavior. A specific model of temporal adaptation that has these characteristics is presented and analyzed in the companion article (Gaudiano, 1992).

### 10 Response to Modulated Gratings

We now turn to another class of spatiotemporal modulation functions that is frequently encountered in this experimental literature: the temporally-modulated sinusoidal grating.

The stimuli in question differ from the temporally modulated stimuli of Sections 4 and 7 in that the temporal modulation is only affecting one component of the spatial input, which is superimposed on a steady background. Mathematically, this input distribution is represented as a sum of two terms:

\[
I(x, t) = I(x) + J(x)m(t). \tag{25}
\]

Based on this definition, the modulated component of the spatial input (the grating \( J \)) and the temporal modulation function need not be restricted to nonnegative values: as long as the maximum overall negative excursion of the modulated grating \( J(x)m(t) \) is smaller in amplitude than the steady background \( I(x) \), the net input distribution remains nonnegative.

The input function (25) only affects the right-hand side of (16), and direct substitution into (18) leads to the general solution

\[
v(x, t) = \frac{1}{a} \left\{ (B + D) (I_C - I_S) + BM_S - D M_C \right\} (1 - e^{-at}) + \\
\left[ (B + D) (J_C - J_S) \int_0^t e^{a\tau} m(\tau)d\tau + v_0 \right] e^{-at}, \tag{26}
\]

with \( I_C, I_S, M_C, M_S, \) and \( a \) defined as before in (9), (17), (19), and

\[
J_C = \int_{-\infty}^{\infty} c_C(x - \xi)J(\xi)d\xi \quad \text{and} \quad J_S = \int_{-\infty}^{\infty} S(x - \xi)J(\xi)d\xi, \tag{27}
\]

for an arbitrary spatial modulation function \( J(x) \).

This solution is almost identical to (18), with the exception that the two components of the input distribution give rise to two distinct response components: a temporally modulated component depending only \( J(x) \) and an asymptotically steady component depending only on \( I(x) \). Segregation of steady and modulated components is a known property of LTI systems.
Consider now the network response to stimuli that are sinusoidally modulated in both space and time against a steady uniform background, i.e., let \( I(x) \equiv I \) (constant); the temporal modulation function is

\[
m(t) = \sin \Omega t,
\]

and the spatial modulation is

\[
J(x) = \alpha \cos \omega x.
\]

The above equations describe a form of counterphase modulation such as is typically encountered in the experimental literature (e.g., Hochstein and Shapley, 1976a,b).

Substituting (5), (28), and (29) into (26) leads to the following solution:

\[
t(x,t) = t_0 e^{-\alpha t} + \frac{B + D}{a} \left[ I (V_C - V_S) - \frac{D V_C - B V_S}{B + D} \right] (1 - e^{-\alpha t}) + \frac{B + D}{a^2 + \Omega^2} \left[ V_C \exp \left( -\frac{\omega^2 \sigma_C^2}{2} \right) - V_S \exp \left( -\frac{\omega^2 \sigma_S^2}{2} \right) \right] \alpha \cos \omega x \left( \Omega e^{-\alpha t} + \alpha \sin \Omega t - \Omega \cos \Omega t \right).
\]

Once again I begin by making general observations about the form of this solution, which consists of three main components: a transient component, an asymptotically constant component, and a spatiotemporally modulated component.

The first, transient component represents the network’s dependence on its initial state \( t_0 \), and decays at a rate \( e^{-\alpha t} \).

The second, asymptotically constant component reaches steady-state as rapidly as the transient component vanishes, and its equilibrium value depends jointly on the steady background \( I \), the center and surround RF volumes \( V_C \) and \( V_S \), and the physiological parameters \( A, B, D \) and \( M \). However, this term is not affected by the superimposed, temporally modulated grating. The mathematical form of this component shows that interactions between input parameters \( I \), morphological parameters \( (V_C, V_S) \), and physiological parameters \( (A, B, D, M) \) can jointly modify sensitivity to steady input components, resting level, and integration rate of the cell.

The third, spatiotemporally modulated component exhibits a complex dependence on both the spatial and temporal modulation functions. However, this term is unaffected by the steady background \( I \). Thus this term represents the network’s ability to respond to spatiotemporal information in the modulated component of the input. Interaction of the modulated input with the morphological and physiological parameters simultaneously affects spatial frequency selectivity, temporal frequency selectivity, and integration rate. In conjunction with the asymptotically constant component, this shows that manipulation of the network’s parameters can selectively affect sustained or modulated response components, thus giving rise to distinct functional behaviors. Fig. 6 illustrates the network’s ability to segregate a uniform background from modulated input components, even when the passive decay rate \( \alpha \) is zero. This figure should be compared to the response of the plain shunting network of Fig. 4.

In addition, the form of the modulated term shows that for a temporally modulated sinusoidal grating there always exists a position in the network exhibiting a null response, i.e., there are network elements whose response is constant over time. This is due to the quantity \( \cos \omega x \) multiplying the entire right-hand expression in (30), which vanishes for \( \omega x = \pm (2n + 1) \pi / 2 \), \( n = 0, 1, 2 \ldots \).

This point deserves special attention, for reasons that are explained in the next section.

11 Linear vs. Nonlinear Responses

Enroth-Cugell and Robson (1966) made use of a number of criteria to determine whether a cell should be classified as an X or Y RGC. One of the primary criteria, which I will refer to as the null test, consisted of measuring a cell’s response to temporally modulated sinusoidal gratings—such as those used in the last section—as a function of the relative spatial phase between the cell’s RF center and the grating. These authors found that when X cells are stimulated by introduction and withdrawal of a sinusoidal grating, there usually exist two symmetrical values (±90°) of the relative spatial phase at which no response can be elicited. On the other hand, Y cells do not exhibit a null response at any value of relative spatial phase. Instead, at those relative phases where X cells exhibit a null response, Y cells exhibit an on-off, or frequency doubling response, i.e., they respond equally to introduction and withdrawal of the grating. Similar results hold for other types of spatial and temporal modulation functions (Hochstein and Shapley, 1976a,b). This result is in direct disaccord with the analytical finding of the last section: equation (30) predicts that a null response will obtain at ±90° regardless of parameter choice.

In Section 9, I suggested that the overshoots and undershoots seen in RGC responses are generated in preprocessing layers, such as the photoreceptor layer. I now claim that the same overshoots and undershoots are also responsible
for on-off responses in Y cells. In the companion article (Gaudiano, 1992), I introduce a mechanism for nonlinear preprocessing of the input to the push-pull network that resolves the discrepancies outlined thus far, including the existence of on-off responses in Y cells but not in X cells.

12 Summary

This article has introduced a model of spatiotemporal processing in retinal ganglion cells. The model was derived from the feed-forward shunting network (Grossberg, 1970) as a solution to the network’s inability to track temporal modulation of nonnegative input signals.

The resulting push-pull shunting network possesses a number of important analytical and functional properties. Mathematical analyses have shown that addition of push-pull bipolar preprocessing to the shunting network leads to a linear, time-invariant system (16). An LTI system such as (16) has the advantage of being analytically more tractable than its time-varying counterpart (10), and leads to distinct predictions. Perhaps most important is the observation that the response of an LTI system automatically segregates steady and modulated input components. This property of the push-pull shunting network has the advantage of allowing the network to selectively enhance either component by appropriate changes in morphological and physiological parameters. This last observation has led to the proposition that X and Y cells may consist of a single mechanism acting in different parameter ranges.

An initial analysis of the hypothesis that the push-pull shunting network may be a valid model of X and Y RGCs has revealed a number of discrepancies with the experimental data. Specifically, it was noted that (a) a compressive nonlinearity must precede the push-pull mechanism to ensure bounded inputs; (b) transient overshoots and undershoots in response to sudden input onset and offset must originate prior to the shunting interactions; (c) in the absence of a nonlinear preprocessing mechanism, the push-pull shunting network cannot generate the on-off responses typically found in Y cells.

However, it was suggested that all of these discrepancies could be resolved by inclusion of a nonlinear preprocessing mechanism whose properties resemble those of photoreceptors. In the companion article (Gaudiano, 1992) I will introduce a model of such a nonlinear mechanism, and show that the resulting network can indeed resolve these discrepancies and replicate many experimentally determined aspects of X and Y RGC behavior.

Acknowledgments

The author wishes to express his gratitude to Drs. E. Mingolla, M. Cohen, and S. Grossberg for useful comments and suggested revisions of this work in its present form and earlier versions. Thanks also to the anonymous referee who reviewed these articles for many suggestions and for pointing out some important references related to this work. This work was partly supported by the National Science Foundation (NSF IRI-87-16960).

References


**Figure Captions**

**Figure 1:** (a) Schematic of a 1-D feed-forward, distance-dependent shunting network. Inputs $I_i$ impinge upon the network in an on-center, off-surround fashion. (b) Typical Difference-Of-Gaussians (DOG) RF profile. Each cell is excited by inputs falling within a small central area, and is inhibited by inputs falling within a broader surround.

**Figure 2:** Response of a shunting cell to whole-field input modulation. In this example, the passive decay rate $A$ is set to zero. The cell (bottom trace) is unable to follow input modulation (top trace).

**Figure 3:** The behavior of a 1-D population of 100 cells obeying feed-forward shunting dynamics (bottom surface), in response to sinusoidal modulation in both space and time (top). See text.

**Figure 4:** The response of a shunting cell to whole-field sinusoidal modulation when the passive decay rate is nonzero ($A = 80$ and $B = 8$, see (1)). See text.

**Figure 5:** (a): Schematic of the push-pull mechanism. Light signals impinging upon each photoreceptor cause equal and opposite changes in membrane potential of two bipolar cells. These cells differentially activate excitatory and inhibitory membrane mechanisms in a single RGC. (b): The circuit giving rise to ON-center and OFF-center RGCs. Bipolar cells of opposite polarity diverge to generate different response types. The push-pull architecture of (a) co-exists with the standard ON-center/OFF-center dichotomy of (b); both ON-center and OFF-center RGCs receive push-pull inputs, but in opposite configurations.

**Figure 6:** The response of a push-pull shunting cell to sinusoidal temporal modulation of a spatial sinusoidal grating superimposed upon a steady, uniform background of varying intensity. The input modulation is the same for all curves, while the steady background increases in going from the solid line to the dotted and dashed lines. Correspondingly, the response consists of a modulated component whose amplitude does not change between curves, superimposed upon a sustained response whose amplitude is linearly related to the average intensity of the input. Model parameters are the same as in Fig. 2. Compare also to Fig. 4.

**Figure 7:** The response of a 1-D network of 100 push-pull shunting cells. All parameters are the same as in Fig. 3. Top trace: the input is sinusoidally modulated both in space and time. Bottom trace: the network’s response faithfully tracks temporal as well as spatial modulation. Dark lines represent the input and response for a cell located at zero-degree spatial phase relative to the grating, and are thus equivalent to the bottom traces (solid lines) of Fig. 6.

**Figure 8:** Response of a push-pull shunting cell to square-wave temporal modulation. See text. All parameters as in Fig. 6.
Instructions for the compositor

These two manuscripts were composed with the LaTeX document preparation system. An attempt was made to ensure that all symbols are easily distinguishable, including greek letters, subscripts and superscripts, and italicized text. The only exception is the presence of a few uppercase script variables, which are difficult to distinguish from regular uppercase variables in the font used. I have underlined all script letters in green, as suggested in the “Instructions to Authors.” Based on the sample fonts used in your journal I trust that there will be no confusion between the script and non-script letters in the finished format.

All figures are reproduced in final size, either 8.1 or 16.2cm wide. Marginal notes are used to indicate the approximate desired position of the figures, and figures which should be grouped together are indicated in a single marginal note. This grouping can be modified to suit publication constraints.