International Journal of Neural Systems, Vol. 3, No. 3 (1992) 237–248

DIFFERENT RETINAL GANGLION CELLS HAVE DIFFERENT FUNCTIONAL GOALS*

ZHAOPING LI
Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

Received 17 February 1992
Revised 14 August 1992

In mammalian retina, the Y (or M) ganglion cells are significantly more transient in response, more selective to stimuli of low spatial and high temporal frequencies and less selective to spectral information than the X (or P) cells. It is shown that these differences in cell properties can be explained by a model that assigns different functional goals to the different ganglion cell types. In this model, the goal of the Y cells is to extract as *fast* as possible the *minimum* amount of information necessary for quick responses. In contrast, the goal of the X cells is to extract as *much* information as possible. Temporal characteristics of the information extraction by the two cell groups are also derived.

1. Introduction

Natural visual signals are corrupted with noise and redundant due to strong spatio-temporal correlations. It has been proposed that the cat or monkey retina is a general purpose processor that encodes light signals for efficient transmission down the optic nerve.[2,14] Stated another way, the retinal ganglion cells are structured to transmit as much visual information as possible to the brain at a given transmission cost. Consequently, to remove the spatial redundancy in inputs, the ganglion cells require the center-surround structure in their receptive fields. To average out the noise and thus to extract more information, the centers of the receptive fields expand under low illumination or when stimulated at high temporal frequencies. These results agree with the physiological observations.[6]

These earlier works, however, discuss only the ganglion X cells in cats or parvocellular cells (P cells) in monkeys and do not explain the existence of another major type of ganglion cell — the Y cells in cats or magnocellular cells (M cells) in monkeys.[6,8,15,17,20,21] In the rest of the paper, statements on the X-Y cells will be also applicable to the P-M cells unless otherwise stated. Although the X-Y and P-M cells do differ in some aspects such as their contrast gains, we will leave such differences outside the paper. Compared to X cells, Y cells are significantly more sensitive to visual stimuli at lower spatial frequencies (lower spatial resolution) and higher temporal frequencies (transient responses). These cells have larger dendritic fields, a smaller cell density, thicker axons and thus a larger conduction velocity. The X cells respond approximately linearly to inputs while Y cells respond linearly only to stimuli of low spatial frequencies. In monkeys, where color vision exists, P cells are color-opponent while most M cells show little spectral selectivity. Both X and Y cells have center-surround receptive fields and send their outputs to the visual cortex. The Y cells also project to the Superior Culliculus.

This study introduces a model to show that the *different properties* of X and Y cells are designed to achieve *different functional goals*. It is generally believed that while the X channel does the analysis of the *fine* structures in the visual image, the Y channel does the *initial* analysis of the *gross* structure.[4,9,10,22] In the present study, such different goals for X and Y cells are formulated explicitly using the language of information theory — while the goal of the X cells is to extract as *much* visual information as possible, the goal of the Y cells is modeled to extract as *fast* as possible the *minimum* amount of information needed for quick responses. Using optimization technique and assuming that the cat or monkey retina is not involved in any special visual tasks such as "bug detection or recognition", it will be shown that the receptive field properties of the X and Y cells are optimal to achieve their respective goals.

* Work supported in part by a grant from the Seaver Institute.

This model is motivated by the following considerations. For cats and monkeys, the visual tasks are first to locate and then to identify visual objects. To quickly locate objects relevant for survival and to give visual feedbacks for motor tasks or body orientation, it is necessary to have fast visual responses with a minimum necessary extraction of information. The X cells do not serve this purpose. In order to extract accurate information, they integrate the signal over time to improve the signal-to-noise ratio using temporal correlations in visual inputs. Such temporal integration sacrifices speed. Therefore, the Y cells instead are modeled to extract visual information as fast as possible or within a shortest time after the information arrives. Furthermore, they are assumed to extract only a minimum amount of information needed to reach a fast decision in the brain. Denoting the X pathway as the "most" pathway (referring to extracting as much information as possible), then the Y pathway is the "fast" pathway. Roughly, the Y path detects "where" and then the X path recognizes "what" the visual objects are.

One should note that the *speed* of information *extraction* differs from the *rate* of information *transmission* which is more of a concern in the practice of communication engineering. Here, a slower *extraction* (for X cells) means a *longer latency* for any information to go from the photoreceptors to the ganglion cells whereas a slower *transmission* indicates that a *smaller amount* of information is transmitted from one to the other *per unit time*. A slower *extracting* ganglion cell receives the visual information later than a faster one but the *rate* or the amount of the information *transmitted* to it can still be higher. The "fast-most" model can also be stated as follows. The Y cells aim to *extract quickly* the minimum necessary information, while the X cells aim to *transmit* the visual information at a high *rate* to the brain. A cell has to *extract* information by a *larger amount*, although possibly with a longer latency, in order to *transmit* the information at a *higher rate* to the brain.

This "fast-most" model suggests that extracting as much information as possible is not necessarily the goal for every pathway in the sensory systems — even at the early processing stages. *Fast* extraction of *minimum necessary* information is an important goal as well. Note that the X-Y pathways remain segregated in the visual cortex.[21,24] Feeding the fast information back to the "most" path at lower visual levels can enhance the extraction of "more" information through attention and gain control. The "fast-most" model is consistent with the higher axonal conduction velocity of Y cells and the earlier arrival of their signals in the cortex[5,9] (see Ref. 13 for counter views).

The paper is organized as follows. In Sec. 2, it will be shown that the temporal course of information extraction by a cell depends on the form of its impulse response to visual inputs. Therefore, the "fast" and "most" goals, expressed as different conditions on the temporal structure of information extraction, require different impulse response forms. Section 3 shows that the goals of "fast" and "most" information extraction cause transient and sustained responses, respectively. In Sec. 4, the requirement on the Y cells to extract only the minimum information needed is shown to cause their selectivity to stimuli of low spatial frequencies and their color-insensitivity. These response properties required by the functional goals agree with those observed physiologically and thus support the model. The different goals are further illustrated by the respective temporal courses of information extraction from stimuli of various spatial frequencies. Discussion follows in Sec. 5. The main line of argument in this paper should be comprehensible without following all the mathematical details which are provided for interested readers.

2. Formulation

Before studying the consequences of the "fast-most" model, we first formulate the approach explicitly. To keep the study more tractable, only *linear* response properties of the cells are considered. We shall show that the linear model can already account for the qualitative differences between the two paths in the linear regions, which dominate the X responses and the low spatial frequency region of the Y responses.

Concentrating first on the temporal aspect of visual inputs, describe a visual scene at time t by a scalar signal $S(t)$. This signal is received by the photoreceptor with intrinsic noise $N(t)$ which includes quantum and transduction noises. The receptor output is $L(t) = S(t) + N(t)$. The ganglion cell with impulse response $A(t)$ has an output $O(t) = \sum_{t'} A(t-t')L(t') + N_\delta(t)$. Here the intrinsic ganglion cell noise $N_\delta(t)$ includes the effects of spike generation in the optic nerves. The function

$A(t)$, with $A(t) = 0$ for $t < 0$ can be viewed as the retinal temporal filter. Time t is discretized since the photoreceptor output $L(t)$ has a limited temporal resolution (about 20–30 ms for humans) i.e. a temporal frequency cutoff which is set to $\omega_{\text{cutoff}}=1$ for convenience.[11]

Without loss of generality, we assume that all the signals and noises have zero means. Both N and N_δ are assumed uncorrelated with S and are Gaussian white noise with variances N^2 and N_δ^2, respectively. The signal S is temporally correlated and has an autocorrelator $R_s(t-t') \equiv \langle S(t)S(t')\rangle$. Here $\langle\cdots\rangle$ indicates the average over the ensemble of visual scenes; time translation invariance of R_s is assumed because of the temporal homogeneity of the visual inputs. The receptor output L has a correlator $R_L(t-t') = R_s(t-t') + N^2\delta(t-t')$. Quantities $\mathbf{u} = S, N, N_\delta, O, L$ are approximated as Gaussian distributed with probabilities

$$P(\mathbf{u}) \propto e^{-\frac{1}{2}\sum_{i,j} u(i)R_u^{-1}(i,j)u(j)}, \qquad (1)$$

where $R_u^{-1}(i,j)$ is the $\{i,j\}$th element of the inverse of the matrix R_u whose elements[3] are $[R_u]_{ij} = R_u(i-j)$. The matrix R_u has dimension d which is the input duration. In usual visual activities, $d \to \infty$.

Let $I(a;b) = \sum_{a,b} P(a,b)\log\frac{P(a,b)}{P(a)P(b)}$ be the amount of information carried by a about b.[19] Then the total amount of information at the ganglion cell level about the visual inputs is $I(\mathbf{O};\mathbf{S}) = \frac{1}{2}\log\frac{\det(AR_LA^T+N_\delta^2)}{\det(AN^2A^T+N_\delta^2)}$ where A is a matrix with elements $A_{ij} = A(i-j)$ and T on the superscripts denotes the transpose of a matrix. Here $\mathbf{O}$ and $\mathbf{S}$ are the outputs and inputs received over all time. $I(\mathbf{O};\mathbf{S})$ provides no temporal characteristics of the information extraction speed. In principle, one needs to wait till $t \to \infty$ before all outputs $O(t)$ are received to get the total $I(\mathbf{O};\mathbf{S})$.

To access the speed of information extraction, the following question is asked. For new visual information brought in at time t' by $S(t')$, how much of this information is extracted by the ganglion cell outputs $O^t \equiv \{O(t), O(t-1), O(t-2), \ldots, O(-\infty)\}$ over all time up to $t \geq t'$.The answer is $I_{\text{out}}(t,t') \equiv I(O^t;S^{t'}) - I(O^t;S^{t'-1})$ where S^t is defined analogously to O^t and $I(O^t;S^{t'-1})$ is the total extracted information up to t about the input signals arrived before t'. Because of the temporal translation invariance, $I_{\text{out}}(t,t') = I_{\text{out}}(t-t')$ depends only on time differences. Therefore $I_{\text{out}}(t)$ is[a] the information extracted by the ganglion cell from the new input information that arrived exactly t time steps earlier.

The longer one waits after a stimulus, the more information $I_{\text{out}}(t)$ is accumulated about it. This is because of a finite convolution time from input to output and because of the fact that measuring a later input signal gives extra information about an earlier one due to their correlations. $\tilde{I}_{\text{out}}(t) \equiv I_{\text{out}}(t) - I_{\text{out}}(t-1) \geq 0$ is the additional information extracted at exactly t time steps after the arrival of the information. With finite durations of impulse responses and input correlation time, $I_{\text{out}}(t)$ saturates or $\tilde{I}_{\text{out}}(t) = 0$, as $t \to \infty$. A larger $I_{\text{out}}(t)$ or $\tilde{I}_{\text{out}}(t)$ at small t, particularly a larger $I_{\text{out}}(0) = \tilde{I}_{\text{out}}(0)$, implies faster extraction. On the other hand, a larger $I_{\text{out}}(\infty)$=*information rate*=$I(\mathbf{O};\mathbf{S})/d$ indicates a larger amount of total output information whether it is extracted quickly or slowly. A fast extraction can be recognized by a large $I_{\text{out}}(0)/I_{\text{out}}(\infty)$ or a short saturation time for $I_{\text{out}}(t)$.

The photoreceptor level counterpart of $I_{\text{out}}(t)$ is $I_{\text{in}}(t)$, which[b] has the same properties as $I_{\text{out}}(t)$ and $I_{\text{in}}(t) \geq I_{\text{out}}(t)$ since ganglion cells cannot extract more information than the photoreceptors (Fig. 1).

Let T be some time window for the animal's fast responses. Then a fast extracting Y cell should have large $I_{\text{out}}(t)$ or $\tilde{I}_{\text{out}}(t)$ for $t \not\gg T$. At time $t \gg T$ after the arrival of a stimulus, it is no longer the goal of Y cells to continue increasing $I_{\text{out}}(t)$ or having nonzero $\tilde{I}_{\text{out}}(t)$, since such a task is assigned to the X cells. Roughly, the Y cells are to enhance the fast information

$$F \equiv \sum_{t\geq 0} \tilde{I}_{\text{out}}(t)e^{-t/T},$$

which attributes less importance to the information extracted later via a weight $e^{-t/T}$. The weighting function (and T) depends on animal needs and may

[a] $$I_{\text{out}}(t) = \frac{1}{2}\log\frac{\det[A(\tilde{R}_s(t+1)+N^2)A^T+N_\delta^2]}{\det[A(\tilde{R}_s(t)+N^2)A^T+N_\delta^2]}$$
where $\tilde{R}_s(n)$ is a square matrix of same dimension d as R_s, $\sum_{k=1}^{n}[\tilde{R}_s]_{d-n+i,d-n+k}[R_s^{-1}]_{d-n+k,d-n+j} = \delta_{ij}$ and $[\tilde{R}_s]_{ij} = 0$ if i or $j < d-n$.

[b] $$I_{\text{in}}(t-t') = I_{\text{in}}(t,t') \equiv I(L^t;S^{t'}) - I(L^t;S^{t'-1})$$
where L^t is defined analogously to O^t.

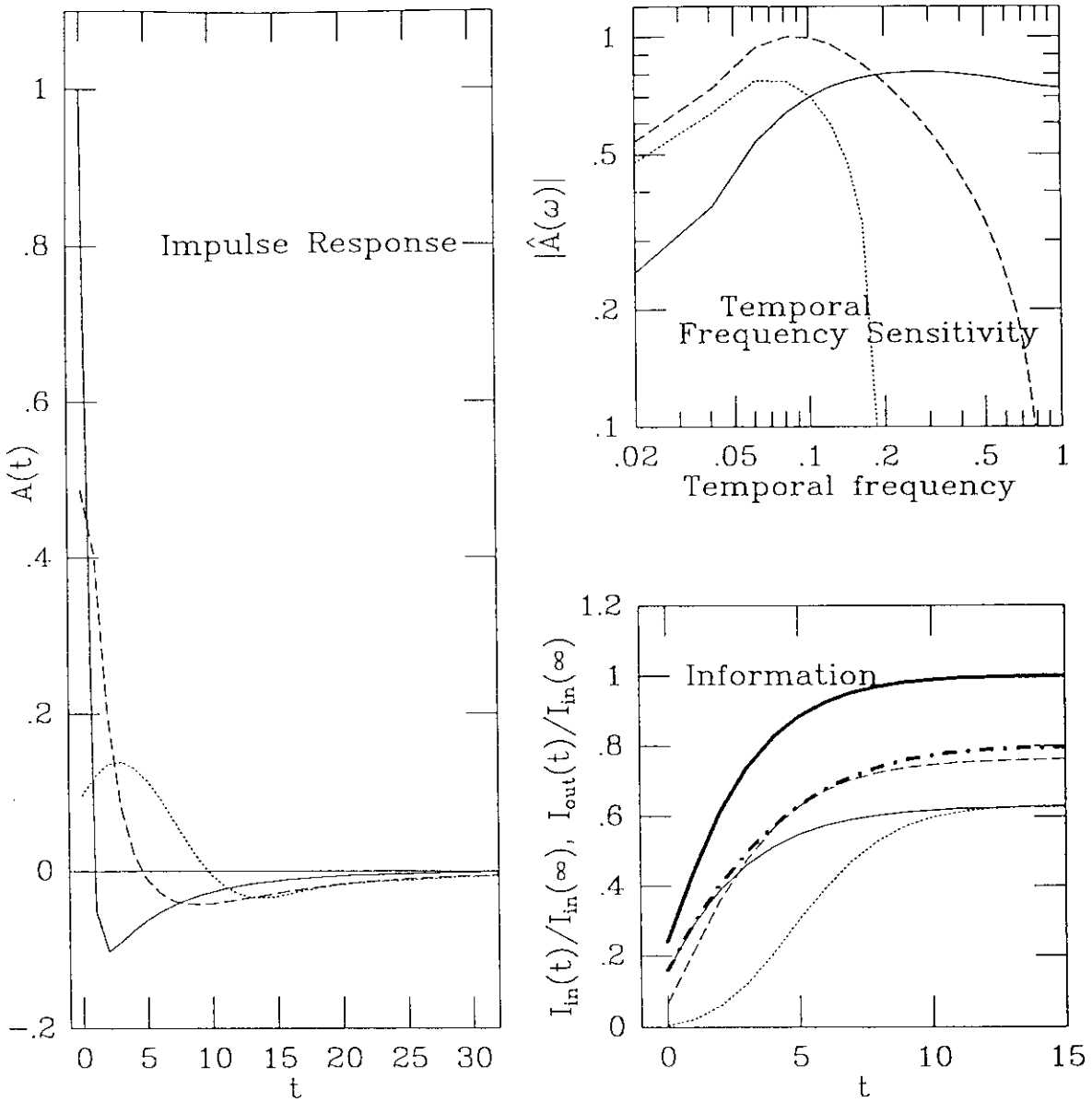

Fig. 1. Simulation results for a "fast" cell (with T=0.5, plotted with thin solid line), an uncooperating "most" cell (dashed line) and a cooperating "most" cell (dotted line). $R_s(t)/N^2 = 1.0e^{-t/40}$, $d = 100$. The lower right graph plots $I_{\text{in}}(t)$ (thick solid line), $I_{\text{out}}^{\text{fast}}(t)$ (thin solid line), $I_{\text{out}}^{\text{most}}(t)$ (dashed line), $I_{\text{out}}^{\text{most}_c}(t)$ (dotted line) and $I_{\text{out}}^{\text{total}}(t)$ (thick dot-dashed line). All of them are plotted by using $I_{\text{in}}(\infty)$ as a unit for convenience. They are respectively the temporal courses of information extraction by the photoreceptor, the "fast" cell, the uncooperative "most" cell, the cooperative "most" cell and the "fast" and cooperative "most" cells together. The left graph plots the impulse responses $A(t)$, obtained by gradient descent simulation. Note that all the impulse responses approach 0 as $t \to \infty$. The upper right graph plots the temporal frequency sensitivities $|\hat{A}(\omega)|$, calculated from Eq. 7 except for the "fast" cell. Both graphs are normalized such that the highest amplitude is 1. The cost C for the uncooperative "most" cell is 2.28, for the cooperative "most" cell 1.12 and for the "fast" cell 2.55. The two Lagrange multipliers are $\lambda^{\text{fast}} = 433$ and $\lambda^{\text{most}} = 150$.

take another form, provided that it decreases with t and diminishes for $t \gg T$. When $T \to 0$, $F = I_{\text{out}}(0)$ and the Y cells attend only to the information extracted instantaneously. In the opposite limit, $T \to \infty$ disregards the information extraction speed and gives $F = I_{\text{out}}(\infty)$ — the extraction goal of the X cells.

The temporal course of information extraction $I_{\text{out}}(t)$ is different from, although dependent on, the temporal course of cell impulse response $A(t)$. For a stimulus $S(0)$ at time 0, both its evoked response $A(t)$ and its transmitted information $I_{\text{out}}(t)$ are zero before $t = 0$ by causality. For $t \geq 0$, the response $A(t)$ can be positive or negative and decays to zero

as $t \to \infty$. In contrast, the accumulated information $I_{\text{out}}(t)$ is non-negative and always nondecreasing. The main aim of this study is to understand the cell response properties $A(t)$ from the information extraction goals expressed by the requirements on $I_{\text{out}}(t)$. Such requirements are stated by the values of F. The impulse response A from a Y cell should lead to a large F at its output for small T, i.e. fast information extraction. But F will also increase by raising the dynamical range or output power $\langle O^2(t)\rangle = \frac{1}{d}\sum_t [ARA^T]_{tt} + N_\delta^2$. For instance, passing inputs **S** directly to the brain without retinal filtering would extract the input information most quickly but the optic fiber would require high power or large channel capacity due to redundancies and noises in visual inputs. For convenience, let us name $C(A) \equiv \langle O^2(t)\rangle / N_\delta^2$ the cost to the animal.[3] Then an optimal A should maximize F at a given $C(A)$. Equivalently, the optimal A minimizes

$$E \equiv C(A) - \lambda^{\text{fast}} F\,, \tag{2}$$

where the constant λ^{fast} weighs the importance of maximizing F as opposed to minimizing the cost $C(A)$. The value of λ^{fast} can be adjusted such that only the minimum necessary information is extracted as we will see in Sec. 4. Therefore, impulse response $A(t)$ for the "fast" cell (or the "most" cell when $T = \infty$) should satisfy

$$\frac{\partial E}{\partial A(t)} = 0 \quad \text{for } t = 0,\, 1,\, 2,\, \ldots\, (d-1) \to \infty\,. \tag{3}$$

3. The Temporal Properties of the "Fast" and "Most" Cells

We examine here whether the impulse responses required by the "fast" and "most" cells agree with those observed physiologically. The responses from the solutions of Eq. 3 have the form $A = \frac{N_\delta}{N} \cdot g(R_s/N^2)$ where $g(\cdot)$ is some function. They depend on the ensemble of the visual scenes via the input correlator R_s. No available experimental data give R_s. Thus a simple choice is made by assuming that

$$\frac{R_s(t)}{N^2} = \frac{S^2}{N^2}\, e^{-t/\tau}\,, \tag{4}$$

where τ approximates the correlation time among visual inputs and S^2 is the average input signal power.

The optimal A are obtained approximately by gradient descents in a d-dimensional space $\{A(0), A(1), \ldots, A(d-1)\}$ to reach a minimum E for $d \gg \tau$ and T. Thus, the impulse response is modelled with d degrees of freedom denoting its value at d time steps. d should be large enough that at the last time step, the impulse response $A(d-1)$ should have reduced sufficiently to zero. The descents are smooth and the final results are robust and easily reproducible. The descents start at $A(0) = 1$ and $A(i \neq 0) = 0$, corresponding to ganglions passing the photoreceptor inputs with no temporal filtering. For comparison, the impulse responses for "most" cells are obtained by taking $T \to \infty$ for F in E or equivalently (and enacted for practical reasons) by minimizing

$$E^{\text{most}} = C(A) - \lambda^{\text{most}} I(O;\, S)/d\,. \tag{5}$$

Figure 1 shows an example of the impulse response solutions $A(t)$ of the "most" and "fast" cells (dashed and solid lines) respectively. Also shown are $I_{\text{in}}(t)$, $I_{\text{out}}(t)$ and the temporal frequency sensitivities of the cells — the Fourier transform $\hat{A}(\omega)$ of $A(t)$ at frequency ω. Both the "most" and "fast" cells are shown to be excited at short time delays and inhibited at long delays by inputs. (Such a temporal structure is analogous to spatial center-surround receptive fields, e.g. excitation from short distances and inhibition from long distances. The opposite polarity — inhibition at short delays and excitation at long delays — can arise from initializing the gradient descent with $A(0) = -1$ and $A(i \neq 0) = 0$). However, compared to the "most" cell, the "fast" cell has a shorter excitation followed rapidly by brief inhibition and thus a shorter response duration. Consequently, the "fast" cell is more sensitive to the higher frequencies than the "most" cells. These results indicate that the "fast" cell responds more transiently than the "most" cell, in agreement with the temporal response properties of the physiological Y and X cells. (One should note that $\hat{A}(\omega) \equiv 0$ for $\omega > 1$ by the definition of this discrete system. In real systems, $\hat{A}(\omega)$ quickly drops outside its temporal frequency band limit). The temporal course of information extraction $I_{\text{out}}(t)$ demonstrates the differences in the functional goals. At about the same cost, the "fast" cell extracts information faster ($I_{\text{out}}^{\text{fast}}(0) > I_{\text{out}}^{\text{most}}(0)$) while the "most" cell eventually extracts more information ($I_{\text{out}}^{\text{fast}}(\infty) < I_{\text{out}}^{\text{most}}(\infty)$).

Intuitively, the sustained response is a strategy to average out noise and integrate or enhance the signal using temporal correlations in the input. It ensures more information extraction, especially when the

signal is small. But obviously, such temporal integration sacrifices speed. In a transient response, a brief excitation followed by quick inhibition removes the average or the temporally redundant input and extracts any new features or information quickly. But it is susceptible to visual noise and leads to less or "inaccurate" information. Therefore, in a visual environment with redundancy and noise, the transient and sustained responses arise from the "fast" and "most" goals respectively.

This conclusion still holds when cooperation between the two paths are considered. For an efficient design, one expects the "most" path to utilize the information already extracted by the "fast" path. It should cooperate by simply attending to the information that is missing or requires slow extractions. Mathematically, the problem can be posed as follows. Given the temporal filter A^{fast} of the "fast" cell, find the filter A^{most_c} of the cooperating "most" cell such that the total output information from the two cells together is maximized at a given cost to the "most" cell. (The superscripts *fast*, *most* and *most*$_c$ are used to denote quantities corresponding to the "fast", the uncooperative "most" and the cooperative "most" cells, respectively.) Explicitly, one modifies Eq. (5) to minimize

$$E_{\text{most}} = C(A^{\text{most}_c}) - \lambda^{\text{most}} I(\{\mathbf{O}^{\text{fast}}, \mathbf{O}^{\text{most}_c}\}; \mathbf{S})/d, \tag{6}$$

where $I(\{\mathbf{O}^{\text{fast}}, \mathbf{O}^{\text{most}_c}\}; \mathbf{S}) = \frac{1}{2} \log \frac{\det(AR_L A^T + N_\delta^2)}{\det(AR_N A^T + N_\delta^2)}$ with $A = \binom{A^{\text{fast}}}{A^{\text{most}_c}}$ is the total output information extracted by the two pathways together. If A^{most} minimizes Eq. (5) and A^{most_c} minimizes (6), then[c]

$$|\hat{A}^{\text{most}_c}(\omega)| = \begin{cases} \sqrt{|\hat{A}^{\text{most}}(\omega)|^2 - |\hat{A}^{\text{fast}}(\omega)|^2} \\ \qquad \text{if } |\hat{A}^{\text{most}}(\omega)|^2 - |\hat{A}^{\text{fast}}(\omega)|^2 > 0, \\ 0 \qquad \text{otherwise}. \end{cases} \tag{7}$$

[c]

$$|\hat{A}^{\text{most}}(\omega)|^2 = \begin{cases} \frac{N_\delta^2}{N^2} \left\{ \frac{\hat{R}_s(\omega)}{2(\hat{R}_s(\omega)+N^2)} \left(1 + \sqrt{1 + \frac{2\lambda^{\text{most}} N^2}{\hat{R}_s(\omega)}}\right) - 1 \right\} \\ \qquad \text{if } |\hat{A}^{\text{most}}(\omega)|^2 > 0, \\ 0 \qquad \text{otherwise}, \end{cases}$$

where $\hat{A}^{\text{most}}(\omega)$ and $\hat{R}_s(\omega)$ are Fourier transforms of $A^{\text{most}}(t)$ and $R_s(t)$, respectively.

Here $\hat{A}^{\text{most}_c}(\omega)$ and $\hat{A}^{\text{fast}}(\omega)$ are the frequency sensitivities of the (cooperative) "most" and "fast" cells, respectively. They are the Fourier transforms of the corresponding impulse responses.

Cooperation reduces the sensitivity from $\hat{A}^{\text{most}}$ to $\hat{A}^{\text{most}_c}$ significantly at high ω where $\hat{A}^{\text{fast}}$ is large and negligibly at low ω where $\hat{A}^{\text{fast}}$ is small. It makes the "most" cell respond in a more sustained manner by expanding its excitation duration at the expense of its inhibition duration (Fig. 1). This causes longer temporal averages over inputs and thus slower information extraction. Such a change is affordable since a faster extraction is covered by the "fast" cell via cooperation. These complimentary roles can be seen in the I_{out} plot. The "fast" cell is almost solely responsible for the information output within the initial response phase whereas the cooperating "most" cell significantly picks up information only when the extraction by the "fast" cell is saturating. Although the cooperative "most" cell alone extracts less information than the "fast" cell in this example, together they extract more information than the uncooperative "most" cell at a smaller cost to the cooperative one.

Therefore, the impulse responses required by the "fast-most" model agree qualitatively with those observed physiologically, whether or not the cells cooperate in extracting information. This conclusion does not depend crucially on the exact form of the correlator R_s (Eq. (4)) and can be obtained from modified correlators such as $R_s(t) \propto e^{-(t/\tau)^\alpha}$ with $\alpha = 0.5, 1.5$. Here, by concentrating on the temporal domain, we have been ignoring the spatial aspects of the inputs and the response. If the temporal signal $S(t)$ (and $N(t)$ etc.) denotes the temporal evolution of a particular spatial Fourier component of the visual image, then the derived $A(t)$ are the impulse responses to the sin wave grating stimuli of the corresponding spatial frequency (see Sec. 4).

4. Ganglion Cell Properties in the Spatial and Spectral Domain

This section shows that the "fast-most" model also explains the different receptive field properties of X and Y cells observed in the spatial and spectral domain. Physiologically, the receptive fields of both cell types have the center-surround structure.[6] Their sensitivities increase with the spatial frequency f of the stimuli at small f, reach the peak sensitivity at f_{peak} and then decrease with f to zero at f_{max}. As the temporal frequency of the stimuli

increases, f_{peak} decreases and the cells adapt from spatial band-pass to low pass filters, i.e. the receptive field center expands and/or surround shrinks. However, compared to the X cells, the Y cells have smaller f_{peak} and $f_{\max}$ and have higher sensitivities at small f. In other words, the Y cells have larger receptive fields and a selectivity to stimuli of lower spatial frequencies. Furthermore, the P cells are color-opponent while the M cells show little color sensitivity. Here we extend the model to account for these observations.

Extending to the spatial domain, visual inputs are temporal signals arriving parallel at different spatial locations on the retina. These parallel signals cannot be treated independently due to their correlations in the natural visual environment. However, such spatial correlations roughly depend only on the distances between spatial locations. Consequently, if we view the visual signals as arriving simultaneously through channels of different *spatial frequencies*, they are uncorrelated with each other up to the second order. Each spatial frequency channel f has a particular input signal $S(f, t)$ with correlator $R_s(f, t)$ and requires particular impulse responses $A(f, t)$ for both the "fast" and the "most" paths. Therefore, the former optimization problem, Eqs. 2 and 6, should be applied simultaneously to minimize the following quantities for all f;

$$E^{\text{fast}}(f) \equiv C(A^{\text{fast}}(f)) - \lambda^{\text{fast}}(f)F(f) \qquad (8)$$

$$E^{\text{most}}(f) \equiv C(A^{\text{most}_c}(f)) - \lambda^{\text{most}}(f) \times I(\{\mathbf{O}^{\text{fast}}(f), \mathbf{O}^{\text{most}_c}(f)\}; \mathbf{S}(f))/d. \qquad (9)$$

The costs C, outputs $\mathbf{O}$ and extracted information, etc. consequently all depend on f. Cooperation between "fast" and "most" paths is assumed.

In Eqs. (8) and (9), $\lambda^{\text{most}}(f)$ and $\lambda^{\text{fast}}(f)$ are different functions of f. This is because the information from each f plays a different role in the "most" and "fast" pathways which are to *identify* and to *locate or detect* the visual objects, respectively. First, signals from the channel f provide visual features with spatial scale of order $1/f$. Therefore, a higher f channel is less important to the Y path since it adds a smaller correction to the location of the objects. In contrast, all the f channels are equally important to the X path since signals at each spatial scale can give new features to identify objects. Second, because the Y cells focus only on the minimum information needed to detect objects, it is reasonable to expect that there exist a $f_{\max}$ above which the information extracted is considered more than the minimum needed. Third, since a lower spatial resolution is needed to detect than to recognize objects, the $f_{\max}$ for the Y cells should be smaller than that for the X cells. All these differences arise from the different goals of the Y and X paths. The former is to extract minimum needed information and the latter is to extract most information. It is sufficient for the present understanding to model such effects by having e.g.

$$\lambda^{\text{fast}}(f) \propto e^{-f/f_c}, \qquad (10)$$

where f_c is a constant. Another choice could be, for example, $\lambda^{\text{fast}} \propto 1/f$. Both choices attribute less significance to higher f channels for the "fast" path. In contrast, λ^{most} is taken to be independent of f.

Guided by the available experimental[7] result $R_s(f) \propto \frac{1}{|f|^2}$, one models the correlation function of signals from each f channel as

$$R_s(f, t) = \frac{S^2}{|f|^2} e^{-c \cdot f t} \equiv \frac{S^2}{|f|^2} e^{-t/\tau(f)}, \qquad (11)$$

where c is a constant and $\tau(f) \equiv 1/(cf)$. Spatial and temporal correlations are coupled by having a shorter correlation time $\tau(f)$ for inputs of higher spatial frequency. This is expected since the details of visual images usually vary faster than the large scale features. The qualitative results shown below do not depend crucially on the exact form of $R_s(f, t)$.

From Eqs. (8) and (9), impulse responses to all spatial frequency f can be derived. One can thus obtain $\hat{A}^{\text{fast}}(f, \omega)$ and $\hat{A}^{\text{most}_c}(f, \omega)$, the sensitivities of the "fast" and "most" cells to stimuli of spatial frequency f presented at temporal frequency ω. We assume that, except translated from each other, the receptive fields of the cells are the same within the same class, at least within a small region on the retina. The spatial forms of the receptive field can thus be derived by noticing the following. At a particular temporal frequency ω, the receptive field is simply the cell output patten at the ganglion cell layer to a point source input at spatial origin. Such an input has spatial Fourier component $\hat{S}(f, \omega) = 1$ independent of f and the spatial Fourier component of output cells activities is thus $\hat{O}(f, \omega) = \hat{A}(f, \omega)\hat{S}(f, \omega) + noise = \hat{A}(f, \omega) + noise$. The receptive field at temporal frequency ω is thus the inverse Fourier transform of $\hat{A}(f, \omega)$ in the spatial variable f.

Figure 2 shows that as ω increases, both cell types change from spatial band-pass to low-pass filters and

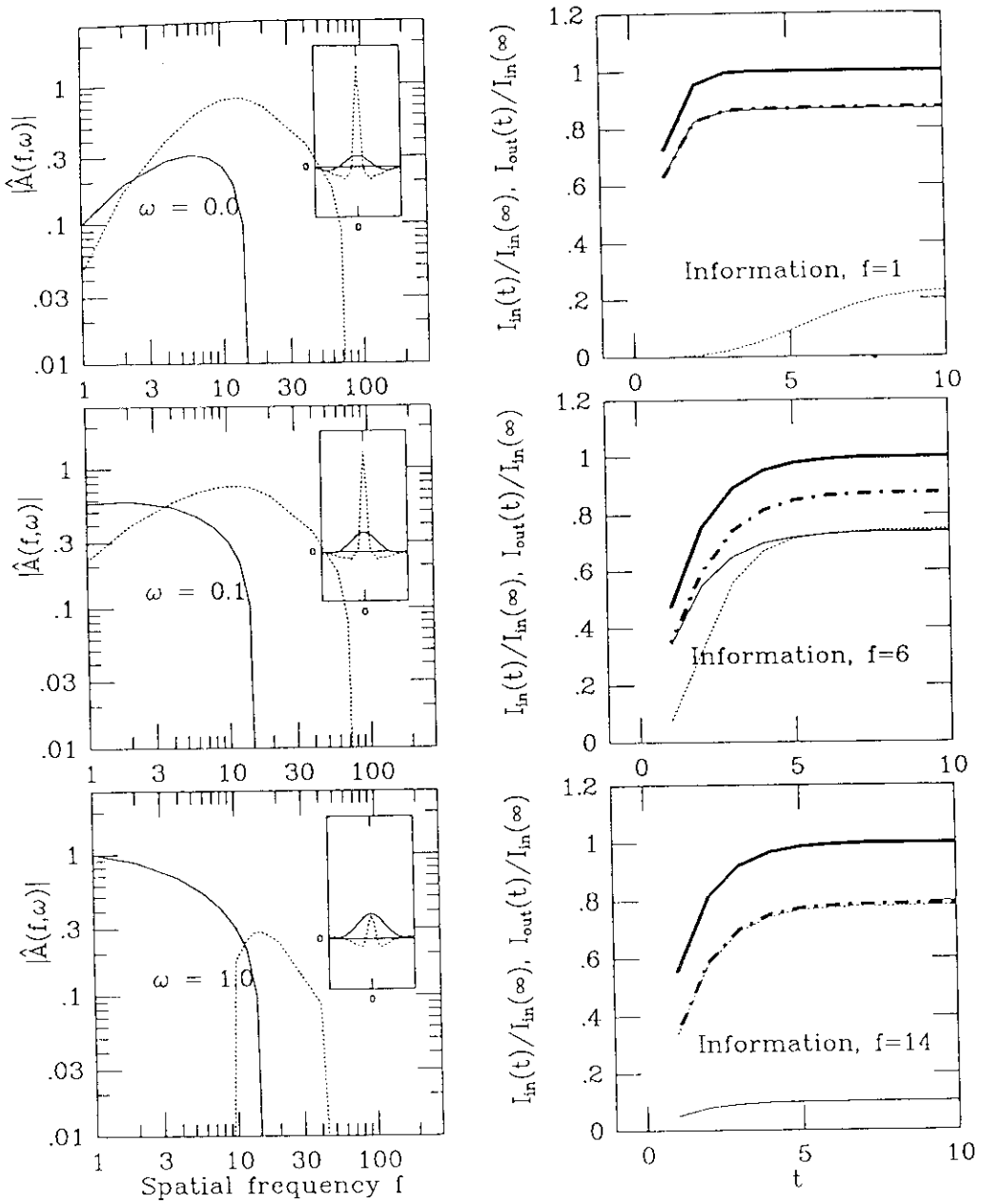

Fig. 2. Simulation results in the spatio-temporal domain. The left column shows the spatial frequency sensitivities of the "fast" cells ($T = 0.5$, solid line) and the cooperative "most" cells (dotted line) at three different temporal frequencies ω. The cell receptive fields at the respective temporal frequencies are shown in the insets. To make the center-surround structure more apparent, the receptive field is integrated in the second dimension. Note that Y cells have larger receptive fields and that the receptive field center expands and/or the surround shrinks, with increasing ω (except for the physiologically unrealistic X cells are high ω, see text). The expansion of receptive field center for the "most" cell from $\omega = 0$ to $\omega = 0.1$ is not significantly visible from the plot but is nevertheless present. The plots are normalized such that $\max_{\{f,\omega,a=\text{most}_c,\text{fast}\}}|\hat{A}^a(f,\omega)| = 1$. The temporal courses of the information extraction at three different spatial frequencies f are illustrated at the right column. In each graph, there are curves respectively for the photoreceptors $I_{\text{in}}(t)$ (thick solid line), the "fast" cells $I_{\text{out}}^{\text{fast}}(t)$ (thin solid line), the cooperative "most" cells $I_{\text{out}}^{\text{most}_c}(t)$ (dotted line) and the "fast" and cooperative "most" cells together $I_{\text{out}}^{\text{total}}(t)$ (thick dot-dashed line). All of them again plotted in the unit of $I_{\text{in}}(\infty)$. Note that at $f = 1$ the thin solid line and the dot-dashed line are almost superposed onto each other. This indicates that the "fast" cells are responsible for almost all output information. At $f = 14$ where the dotted and dot-dashed line almost superpose onto each other, the "most" cells are responsible for almost all output information. Other parameters are: $\lambda^{\text{fast}}(f) = 800e^{-f/5}$, $\lambda^{\text{most}}(f) = 400$, $d = 100$. $R_s/N^2 = (20/|f|^2)e^{-ft/40}$. The $\hat{A}^{\text{most}_c}(f,\omega)$ in left column is obtained by Eq. (7) while the information curves are calculated by using $A(f,t)$ from simulated data.

the receptive fields adapt from a center-surround structure to that of expanded center and/or faded surround. Cooperation reduces $\hat{A}^{\text{most}_c}(f, \omega)$ at low f where the "fast" cell is dominating while Eq. (10) leads to a smaller $\hat{A}^{\text{fast}}$ at higher f. Consequently, compared to the "most" cells, the "fast" cells have a smaller slope of sensitivity at low f, a lower $f_{\max}$ and a lower f_{peak} and thus a larger receptive fields. At very high ω where the "fast" cell dominates, $\hat{A}^{\text{most}_c}$ vanishes at low f by cooperation and increases only at higher f where the "fast" cell performs poorly. If, as in physiology, the "most" cells have a lower temporal frequency cutoff than the "fast" cells whose $\omega_{\text{cutoff}} \equiv 1$, then the area under the dotted curve in the lower left graph of Fig. 2 cannot be covered by the X cells. Nor can the Y cells cover it linearly with their smaller cell densities (lower $f_{\max}$). This is probably the reason for the nonlinearity in the Y cells to average the signals of high f values spatially.[1,8] Such averaging multiplexes the information at high f and high ω from the X pathway.

The different cell properties derived from the model agree qualitatively with those observed experimentally. The I_{out} plots in Fig. 2 further illustrate the different goals of the cells. The "fast" cells carry almost all the output information at low f. They are still faster at intermediate f where the "most" cells start to extract more total information. But at high f, the "most" cells carry almost all (i.e. both fast and slow) information if no multiplexing from X to Y happens. Therefore, the different cell properties can be understood from the "fast-most" model.

Similarly, this model can be extended to the spectral domain to account for the color insensitivity of the M cells. Here we ignore the spatial domain and consider luminance and chrominance channels giving parallel inputs to the retina. The luminance signal is defined as the integrated signals from all the input color cones while the chromatic signal measures the difference between inputs from the different cones. Both channels can provide information to detect visual objects for the magnocellular (or M) path. Since the M path extracts only a minimum amount of information needed for fast responses, one of the channels can be neglected when the other one alone can already detect the objects. It is shown below why the chrominance instead of the luminance channel is neglected.

The luminance and the chrominance signals are assumed to have the same spatiotemporal properties and to be uncorrelated with each other up to the second order. However, the signal power in the luminance channel is much larger.[3,24] It can be shown that for a temporal signal with correlator $R_s(t) = S^2 e^{-t/\tau}$ and $\tau > 1$, $I_{\text{in}}(\infty) \approx \frac{1}{2\tau} \log(1 + \frac{S^2}{N^2}\tau)$ and $I_{\text{in}}(0) \approx \frac{1}{2} \log(1 + \frac{S^2}{N^2} \cdot a)$ where $a \approx \frac{1}{\tau}$. Thus the luminance channel, characterized by a larger S^2, has more total information $I_{\text{in}}(\infty)$. Moreover, since the ratio between the information arriving instantly and accumulatively, $I_{\text{in}}(0)/I_{\text{in}}(\infty)$, increases with S^2, the luminance channel also carries *faster* information. Slower information in the chrominance channel makes a "fast" path less useful and is more suited for the "most" path which has no time limits for extraction. Figure 3 shows that the (uncooperative) "most" path compensates for the smaller signal in the chrominance channel by longer temporal averaging. But the "fast" path extracts a much smaller fraction of the fast information ($I_{\text{out}}(0)/I_{\text{in}}(0)$) in the chromatic than in the luminance channel. Thus for the M pathway, the chromatic channel is inefficient and negligible when the luminance channel already supplies the minimum information needed. This conclusion does not depend crucially on the exact form of R_s. In passing, let me remark that the results do not imply that the M path in low illumination is useless. Rather, a channel with a lower signal power is negligible for the M path when another channel of a higher signal power exists. In fact, the M path is believed to be the predominant conveyor of visual information in low illumination.[18]

5. Summary and Discussions

The present study shows that the differences in the receptive field properties of the retinal ganglion cells are due to their different functional goals. The goal of Y cells is to *quickly* extract a *minimum* amount of visual information needed for *fast* responses while the goal of the X cells is to extract as *much* visual information as possible. The "fast-most" model explains the physiologically observed properties of the two ganglion cell types. Explicitly, the faster extraction by the Y cells causes them to respond more transiently or selectively to stimuli of higher temporal frequencies than the X cells. Furthermore, the requirement of minimum information extraction makes the Y cells more sensitive to stimuli of larger sizes or lower spatial frequencies. The color insensitivity of the M cells can be similarly explained by their "fast" goal. Cooperation between the two cell groups enhances the sustained nature of the

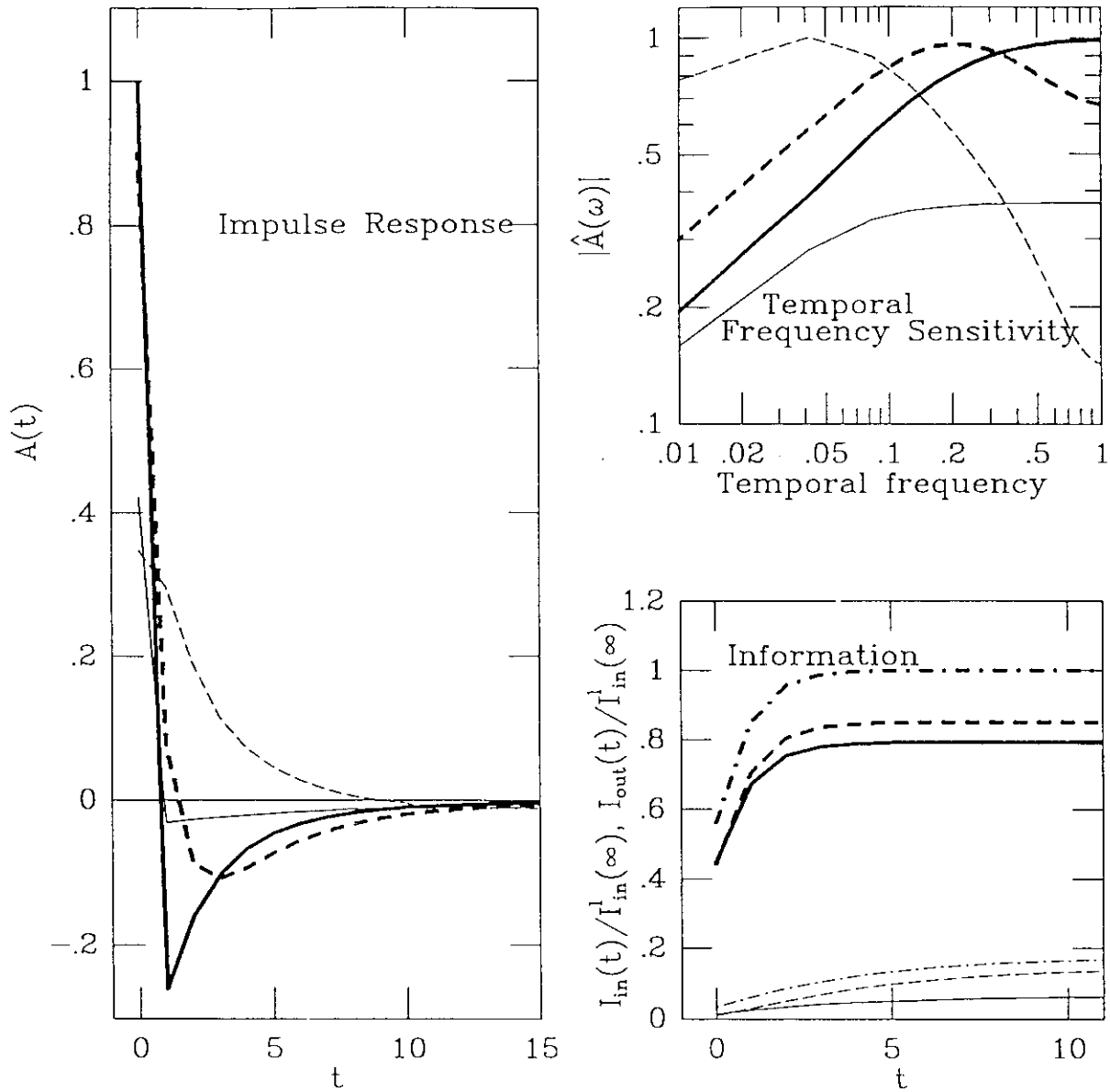

Fig. 3. Simulation results of the "fast" and "most" cells for luminance and chromatic channels. In all plots, the thicker lines are for luminance, the thinner lines for chrominance, the dashed lines for (uncooperating) "most" cells, the solid lines for "fast" cells ($T = 0$), and the dashed-dotted lines for $I_{\text{in}}(t)$. The information plot is in the unit of $I^l_{\text{in}}(\infty)$ of luminance input. The correlators of luminance and chromatic signals are R^l_s and R^c_s with $R^c_s/N^2 = 0.05R^l_s/N^2 = 0.2e^{-t/20}$. Other parameters are $d = 50$, $\lambda^{\text{most}} = 300$ and $\lambda^{\text{fast}} = 360$.

responses by the X cells and their selectivity to stimuli of higher spatial frequencies. (The degree of the cooperation in physiology is yet to be assessed.)

This model also provides the temporal characteristics of information extraction by the two ganglion classes in the visual input ensemble (Fig. 2). The Y cells are shown to extract input information faster especially from low spatial frequency (large) stimuli where they also extract more total information than the X cells. This is particularly true for monkey M cells which have much higher sensitivity than the P cells.[18] At higher spatial frequencies (smaller stimuli), the X cells extract more total information than the Y cells. However, they may still be slower in extraction, especially when they are band-limited in temporal frequency and when Y cells multiplex some information from them in regions of high spatio-temporal frequencies. To improve and test the model, quantitative statistical properties of the visual inputs (e.g. $R_s(f, t)$) and accurate knowledge

of the photoreceptor and ganglion noises are needed. While an experimental measurement of the output information under the whole ensemble of natural scenes would be formidable, such measurements can be performed in an ensemble of limited input stimuli. If the receptive fields adapt to the new ensemble quickly enough, then these measurements can provide a partial test of the present model.

An alternative view of X-Y division has been proposed by van Essen and Anderson.[24] They argued that X-Y division is to reduce the cost to animals by having different cells specializing on different spatio-temporal regions while achieving a unified goal of most information extraction. Here such a cell division is argued to be caused by the cooperation between cells of *different* functional goals.

It is conceivable that different animal needs of "fast" and "most" information may explain why X-Y cells in cats have comparable sensitivities while M cells are much more sensitive than P cells in monkeys. Also, more detailed modeling is needed to implement the nonlinearity in the Y cells in the high spatial frequency region.[1,8,25] Although nonlinearity plays an important role in the function of Y cells, the "fast" goal has been shown to be already manifest in the linear part of the cell properties.

The "fast-most" model suggests that extracting as much information as possible is not necessarily the goal for every pathway in the sensory system — even at the early processing stages. *Fast* extraction of *minimum necessary* information is an important goal as well. The projection of the two paths to higher visual centers provides a different perspective when looking at the visual cortex. This perspective can hopefully shed some light on the segregation, interaction and cooperation between the two paths at each visual level. For example, the "fast" goal of the Y path has potential implications for visual attention, gain control and visual feedbacks to motor tasks, etc. Such studies can in turn test and improve the present model and will lead to a better understanding of the information processing for visual tasks.

Acknowledgements

I wish to thank J. J. Atick, A. V. M. Herz and many others for useful discussions and suggestions and the Seaver Institute for its support.

References

1. C. H. Anderson, "The transmission of information in X and Y ganglion cells," *Invest. Ophthamol. Vis. Sci.* **27**, 242 (suppl.) (1986).
2. J. J. Atick and A. N. Redlich, "Towards a theory of early visual processing," *Neural Comput.* **2**, 308–320 and "What does the retina know about natural scenes?" *Neural Comput.* **4**, 196–210 (1992).
3. J. J. Atick, Z Li and A. N. Redlich, "Understanding retinal color coding from first Principles," *Neural Comput.* **4**, 559–572 (1992) and "Color coding and its interaction with spatio-temporal processing in the retina," IASSNS-HEP-90/75 (1990).
4. G. H. Bishop, "Fiber groups in the optic nerve," *Am. J. Physiol.* **106**, 460–470.
5. B. G. Cleland, W. R. Levick and K. J. Sanderson, "Properties of sustained and transient ganglion cells in the cat retina," *J. Physiol.* **228**, 649–680 (1973).
6. A. M. Derrington and P. Lennie, "The influence of temporal frequency and adaptation level on receptive field organization of retinal ganglion cells in cat," *J. Physiol.* **333**, 343–366 (1982) and "Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque," *J. Physiol.* **357**, 219–240 (1984).
7. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am.* **A4**, 2379–2394 (1987).
8. S. Hochstein and R. Shapley, "Quantitative analysis of retinal ganglion cell classifications," *J. Physiol.* **262**, 237–264 (1976).
9. H. Ikeda and M. J. Wright, "Receptive field organization of 'sustained' and 'transient' retinal ganglion cells which subserve different functional roles," *J. Physiol.* **227**, 769–800 (1972).
10. E. R. Kandel and J. H. Schwartz, *Principles of Neural Science* (Second Ed.) (Elsevier, 1985).
11. D. H. Kelly and H. R. Wilson, "Human flicker sensitivity: Two stages of retinal diffusion," *Science* **202**, 896–899 (1978).
12. J. J. Kulikowski and D. J. Tolhurst, "Psychophysical evidence for sustained and transient detectors in human vision," *J. Physiol.* **232**, 149–162 (1973).
13. P. Lennie, "Perceptual signs of parallel pathways," *Phil. Trans. R. Soc. Lond.* **290B**, 23–37 (1980).
14. R. Linsker, "Self-organization in a perceptual network," *Computer* **21:3**, 105–117 (1988) and "An application of the principle of maximum information preservation to linear systems," in *Advances in Neural Information Processing Systems I*, ed D. Touretzky (1989).
15. R. T. Marrocco and J. W. McClurkin and R. A. Young, "Spatial summation and conduction latency classification of cells of the lateral geniculate nucleus of macaques," *J. Neurosci.* **2:9**, 1275–1291 (1982).
16. F. M. De Monasterio, "Properties of concentrically organized X and Y ganglion cells of macaque retina," *J. Neurophys.* **41:6**, 1394 (1978).
17. F. M. De Monasterio and R. Gouras, "Functional properties of ganglion cells of the rhesus monkey retina," *J. Physiol.* **251**, 167–195 (1975).

18. K. Purpura, E. Kaplan and R. M. Shapley, *Proc. Natl. Acad. Sci. USA* **85**, 4534–4537 (1988).
19. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (The University of Illinois Press, Urbana, 1949).
20. R. Shapley and V. H. Perry, "Cat and monkey retinal ganglion cells and their visual functional roles," *Trends in Neurosci.* **9**, 229–235 (1986).
21. J. Stone, *Parallel Processing in the Visual System* (Plenum Press, New York, 1983).
22. U. Tulunay-Keesey, "Flicker and pattern detection: A comparison of thresholds," *J. Opt. Soc. Am.* **62**, 446–448 (1972).
23. D. J. Tolhurst, "Separate channels for the analysis of the shape and movement of a moving visual stimulus," *J. Physiol.* **231**, 385–402 (1973).
24. D. C. Van Essen and C. H. Anderson, "Information processing strategies and pathways in the primate retina and visual cortex," in *Introduction to Neural and Electronic Networks*, eds. S. F. Zornetzer, J. L. Davis and C. Lau (Academic Press, Florida, 1990).
25. J. D. Victor, "The dynamics of the cat retinal Y cell subunit," *J. of Physiol.* **405**, 289–320 (1988).