



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### The Optimal Synapse for Sparse, Binary Signals in the Rod Pathway

**Citation for published version:**

Clark, PT & van Rossum, MCW 2006, 'The Optimal Synapse for Sparse, Binary Signals in the Rod Pathway: Neural Computation' *Neural Computation*, vol 18, no. 1, pp. 26-44., 10.1162/089976606774841530

**Digital Object Identifier (DOI):**

[10.1162/089976606774841530](https://doi.org/10.1162/089976606774841530)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher final version (usually the publisher pdf)

**Published In:**

Neural Computation

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## The Optimal Synapse for Sparse, Binary Signals in the Rod Pathway

**Paul T. Clark**

*s9905130@sms.ed.ac.uk*

**Mark C. W. van Rossum**

*mvanross@inf.ed.ac.uk*

*Institute for Adaptive and Neural Computation,  
School of Informatics, Edinburgh, EH1 2QL, U.K.*

The sparsity of photons at very low light levels necessitates a nonlinear synaptic transfer function between the rod photoreceptors and the rod-bipolar cells. We examine different ways to characterize the performance of the pathway: the error rate, two variants of the mutual information, and the signal-to-noise ratio. Simulation of the pathway shows that these approaches yield substantially different performance at very low light levels and that maximizing the signal-to-noise ratio yields the best performance when judged from simulated images. The results are compared to recent data.

### 1 Introduction ---

In this letter, we study early visual processing at very low light levels. At these so-called scotopic light levels, the photon capture rate per rod-photoreceptor is on the order of one per minute. The rod cells can detect single photons (Baylor, Lamb, & Yau, 1979; Baylor, Nunn, & Schnapf, 1984; Schneeweis & Schnapf, 1995). Photon capture by the rod can lead to a response in the ganglion cells (Barlow, Levick, & Yoon, 1971; Mastronarde, 1983b). The rod is the most common cell type in the retina; there are 20 times more rods than cones (Sterling & Demb, 2004). The large number of rods serves to detect as many photons as possible, whereas the spatial resolution of the rod pathway is low. The scotopic rod pathway therefore has a large convergence. The rod-bipolar cell collects the signal from some 10..100 rods (Dacheux & Raviola, 1986; Grünert, Martin, & Wässle, 1994; Tsukamoto, Morigiwa, Ueda, & Sterling, 2001) while each rod connects to only two bipolar cells.

Even when no photon is absorbed, the rod response is corrupted with continuous noise. This poses a potential problem for the pathway: a sharp thresholding function before summing the rod responses is required to maintain the single photon response. Without the nonlinearity, the single

responses would drown in the noise (Baylor et al., 1984). An earlier biophysical model showed how such a nonlinearity can be implemented and demonstrated that the nonlinearity is indeed necessary to obtain observed performance (van Rossum & Smith, 1998). Recently, the transfer function of the synapse was measured, providing direct evidence for the existence of such a nonlinearity, and confirmed its synaptic mechanism (Field & Rieke, 2002b; Sampath & Rieke, 2004; Berntson, Smith, & Taylor, 2004).

The performance of the pathway is critically dependent on the synaptic transfer function and its threshold. This raises the question how the threshold should be set from first principles. Interestingly, this question in general has no straightforward answer (Basseville, 1989). In this study, we research ways to set the optimal synaptic transfer function for sparse, binary signals. Counterintuitively, we show that different performance criteria lead to different optimal thresholds. Simulation of the pathway suggests that these different threshold settings greatly influence the signal in the bipolar cell.

We first introduce our description of the pathway and then analyze different performance measures in the case of a sharp binary threshold. Next we extend to the more general case of smooth transfer functions, for which we rely on simulations. Finally, we discuss our predictions. We are not aware of any other studies comparing performance measures for sparse, binary detection—not in the bipolar pathway or in a general case.

## 2 Rod and Rod-Bipolar Pathway

**2.1 Model for the Rods.** The layout of the modeled rod-bipolar pathway is shown in Figure 1. At the lowest light levels, a rod photoreceptor might

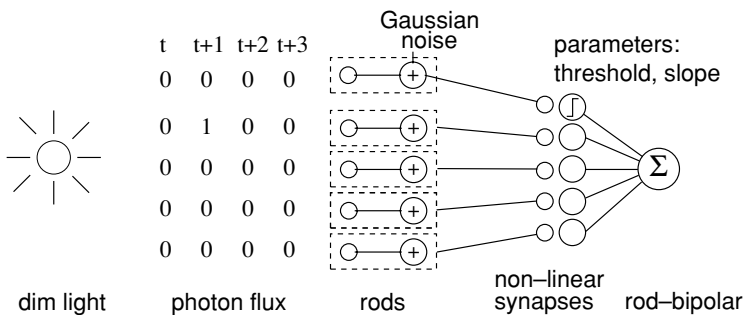


Figure 1: Diagram of our model of the rod-bipolar pathway. A dim light source causes a very sparse flux of photons (modeled in discrete time steps). The photons are detected by the rods. Intrinsic gaussian rod noise corrupts the response. After a nonlinear synapse, the rod-bipolar sums the rod responses. The question is how the synapse's threshold and slope should best be set to minimize signal loss.

detect a photon only every few minutes. Not every photon is absorbed and detected, but for simplicity we assume that each photon a rod receives is detected and leads to a response. This effectively yields an extra scale factor in the light level, the quantum efficiency, which is estimated between 3% and 50% (Baylor et al., 1979; Field, Sampath, & Rieke, 2005). The number of photons absorbed by the rod follows a Poisson distribution.

The full dynamics of the response and the noise are not taken into account. We discretize the time into bins with the duration of the pathway's integration time. The rod integration time is some 100..200 ms (Baylor et al., 1984; Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990). With the light level  $\rho$ , we denote the probability that a rod receives a photon per time bin. Because the power spectra are similar, it is unlikely that temporal integration by the synapse can strongly reduce the noise in favor of the signal (van Rossum & Smith, 1998). However, it is important to note that, precise data lacking, synaptic filtering could lower the noise somewhat; in addition, bandpass filtering could increase the temporal information (Bialek & Owen, 1990; Armstrong-Gold & Rieke, 2003).

At the low light levels we consider here,  $\rho \ll 1$  and the number of absorbed photons  $n$  is small: mostly zero and sometimes one. Thus, at the low light levels considered, a rod can essentially have two responses: it either detects a photon or does not. The probability that a particular rod detects two photons is negligible ( $\rho^2$ ), as is the probability that two out of  $N$  rods detect a photon simultaneously, namely,  $\rho^2 N(N-1)$ . The task of the bipolar cell is therefore to discriminate between the case that none of the rods absorbed a photon and the case that one rod absorbed a photon.

Importantly, the rod response is noisy, and its voltage distribution can be fitted to a gaussian with a standard deviation that increases with the number of photons absorbed. The probability distribution for a certain response amplitude  $x$  from a rod is (Field and Rieke, 2002a, 2002b)

$$\begin{aligned}
 P(x) &= \sum_{n=0}^{\infty} \frac{\rho^n \exp(-\rho)}{n!} \mathcal{G}(n\bar{x}, \sigma_D^2 + n\sigma_A^2) \\
 &= \sum_{n=0}^{\infty} \frac{\rho^n \exp(-\rho)}{n!} \frac{1}{\sqrt{2\pi(\sigma_D^2 + n\sigma_A^2)}} \exp\left[-\frac{1}{2} \frac{(x - n\bar{x})^2}{(\sigma_D^2 + n\sigma_A^2)}\right], \quad (2.1)
 \end{aligned}$$

where  $\mathcal{G}$  denotes the gaussian distribution. Without loss of generality, the mean response to a single event  $\bar{x}$  is normalized to 1. The empirical values for the noise in mouse rods are  $\sigma_D = 0.27$  and  $\sigma_A = 0.33$  (Field & Rieke, 2002b).

These values for  $\sigma$  are only approximate values for the noise seen by the bipolar. It should be noted that the signal as seen by the bipolar can be noisier than this, because stochastic vesicle release can corrupt the signal further (Rao, Buchsbaum, & Sterling, 1994; van Rossum & Smith, 1998);

there are no precise estimates on this. On the other hand, synaptic filtering might reduce the noise, as stated above.

Finally, the rod signal is corrupted with thermally driven spontaneous isomerization of rhodopsin (Baylor et al., 1984). This rate is about  $10^{-3}$  events per rod per integration time. These events introduce extra errors because they are indistinguishable from real photon captures and therefore cannot be filtered out. They are thought to be a major contribution to the so-called dark light (Copenhagen, Donner, & Reuter, 1987; Sterling, Freed, & Smith, 1988).

**2.2 The Bipolar Cell.** The rod provides input to both OFF and ON bipolar cells. The OFF-bipolar pathway does not seem to be tuned for low scotopic vision (Soucy, Wang, Nirenberg, Nathans, & Meister, 1998; Völgyi, Deans, Paul, & Bloomfield, 2004); hence, we ignore it here. The rod ON-bipolar cell pools the signal from some 10 to 100 rods. However, as the rod signal is noisy, the single photon signal would be lost in the noise if the bipolar were to sum the rod signals linearly. The reason is that the noise is pooled from all rods; thus, the standard deviation of the noise in the bipolar scales as  $\sqrt{N}$ , whereas only one rod carries the signal. It has been noted that therefore it is essential to threshold the rod signals before they are summed by the bipolar (Baylor et al., 1984). In a modeling study, it was proposed that this threshold is implemented using a second-messenger cascade synapse; such a threshold mechanism yielded performance consistent with the physiological and psychophysical data (van Rossum & Smith, 1998).

For now, we assume that the synaptic transfer function  $g(x)$  is a sharp step function with a threshold given by  $\theta$  so  $g(x) = 0$ , if  $x < \theta$  and  $g(x) = 1$  otherwise. The threshold  $\theta$  is the adjustable parameter. The simple transfer function is easy to study, and a binary function might seem to fit the binary input signal best. This second statement will turn out not to be fully correct, as is shown below, where other synapse models are discussed.

Consider first that there is just one rod connected to the bipolar cell. We introduce the *false-positive rate*  $\alpha$  (no photon, but erroneously detected) and the *false-negative rate*  $\beta$  (photon was received but not detected in the bipolar). For one rod, the  $n = 0$  and  $n = 1$  term in equation 2.1 yield

$$\alpha = \frac{1}{2} \left[ 1 - \operatorname{erf} \left( \frac{\theta}{\sqrt{2}\sigma_D} \right) \right]$$

$$\beta = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\theta - 1}{\sqrt{2(\sigma_D^2 + \sigma_A^2)}} \right) \right]$$

In case  $N$  rods are connected to the bipolar cell, the bipolar cell is assumed to sum the thresholded rod responses, that is,  $y = \sum_{i=1}^N g(x_i)$ . And after some combinatorics, one finds that the probability for the absorption of  $k$

photons and a bipolar response  $j$  equals

$$P(j, k) = \binom{N}{j} (1 - \rho)^{N-j} \rho^j \times \sum_{l=0}^j \binom{N-j}{k-l} \binom{j}{l} \alpha^{k-l} (1 - \alpha)^{N+l-j-k} \beta^{j-l} (1 - \beta)^l,$$

with the convention that  $\binom{j}{l} = 0$  if  $j < 0$ . In the limit of small  $\rho$  and small  $\rho N$ , again only two errors are important. First, none of the rods received a photon, but the output is unequal to zero. This probability is written as  $\alpha_N = 1 - (1 - \alpha)^N$  and can be interpreted as the generalized false-positive rate. Second, one of the rods received a photon, but the bipolar output is zero. This is written as  $\beta_N = \beta(1 - \alpha)^{N-1}$ . The false-positive and -negative rate characterize the pathway as a function of the threshold level  $\theta$ . The threshold can roughly be deduced from ganglion cell data that showed that the false-negative rate is about 50% (Mastrorarde, 1983a; van Rossum & Smith, 1998). This corresponds to a threshold setting of  $\theta \approx 1$ . In this letter, we examine the more fundamental question of how the optimal value for the threshold follows from the performance measure imposed.

### 3 Performance Measures

---

**3.1 Threshold from Minimizing the Detection Errors.** The problem of how to set the threshold can be analyzed with signal detection theory (Green & Swets, 1966; Van Trees, 1968). The setting of the threshold determines the trade-off between false positives and false negatives. A natural choice is to weigh both errors equally. In this case, the error equals the mean square error between input and output, or the Hamming distance. With one rod connected to the bipolar, the total error rate, denoted  $ER$ , is

$$ER(\theta) = (1 - \rho)\alpha(\theta) + \rho\beta(\theta),$$

where the first term is the false-positive rate and the second term the false-negative rate. Now the threshold  $\theta$  can be varied so that the ER is minimal.

If multiple rods are converging onto the bipolar, a simple counting argument gives for the error rate,

$$ER = (1 - \rho N)\alpha_N + \rho N\beta_N,$$

where we ignored terms of order  $\rho^2$ . In Figure 2A, the total error is plotted as a function of the threshold level when 10 rods are converging. For low thresholds, the false-positive rate is very high, and the error rate is close to 1. For very high thresholds, the error rate is quite low:  $ER = \rho N$ . Here the high threshold eliminates all photon events. The output is completely dark, which is not far from the truth but not very useful. For intermediate

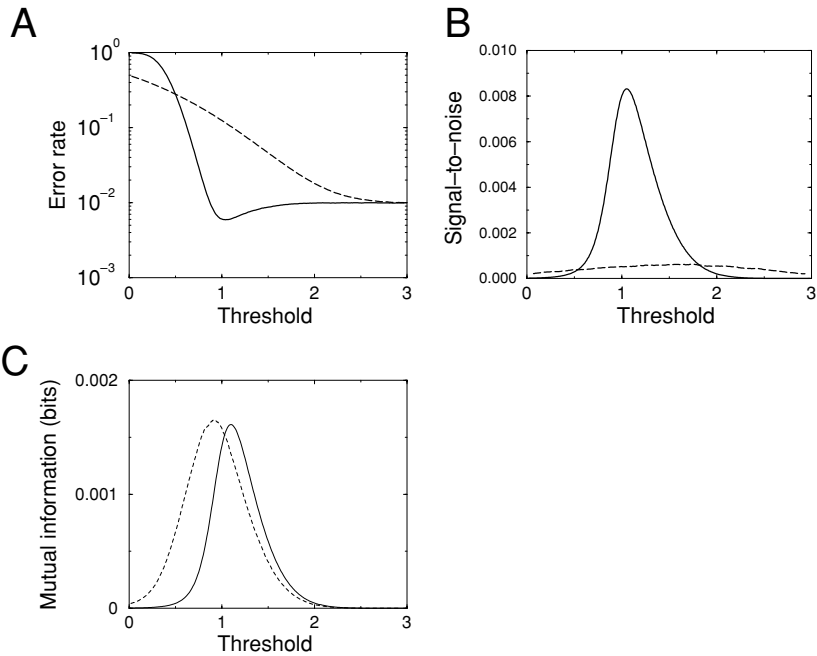


Figure 2: The behavior of the three performance measures of the bipolar pathway as a function of the threshold level. (A) The total error rate, which counts false positive and false negatives. The minimum in the error rate corresponds to best performance. The dashed line indicates the much worse performance when the synapses are linear and the threshold is done after the summation. (B) The signal-to-noise ratio for a contrast discrimination task. The thin dashed line indicates the performance when thresholding is done after the summation. (C) The mutual information between the light level and the bipolar output as a function of the threshold. The dashed line indicates the mutual information between the rod signal and bipolar signal ( $y$ -scale divided by 20 to aid visualization; on the same scale, the dashed line would be much larger than the solid line). Parameters for all panels: 10 rods, light-level: 0.001 photons/rod/time step; rod noise is according to Field and Rieke (2002b).

thresholds, the error rate has a minimum for which false-positive and false-negative rates are traded off. As the light level is lowered, the optimal threshold increases and can be larger than 1. This gives the somewhat counterintuitive result that if the signal is very sparse, a high threshold is beneficial, although this causes missing a large fraction of the events.

To show the benefit of the thresholding synapse, we also show the error rate when the synapses are linear and the signal is thresholded after

summing (see Figure 2A, dashed line). This error rate has no minimum, and performance is worse in this case.

**3.2 Threshold from Bayesian Inference.** The same threshold value also follows from probabilistic Bayesian inference. The probability that the rod absorbed a photon ( $k = 1$ ) versus that it did not ( $k = 0$ ), given a response  $y$ , is

$$\begin{aligned} g(y) &= \frac{P(k = 1|y)}{P(k = 0|y) + P(k = 1|y)} \\ &= \left[ 1 + \frac{P(y|k = 0) P(k = 0)}{P(y|k = 1) P(k = 1)} \right]^{-1} \\ &= \left[ 1 + \frac{(1 - \rho_{SP})\mathcal{G}(0, \sigma_D^2) + \rho_{SP}\mathcal{G}(1, \sigma_D^2 + \sigma_A^2)}{\mathcal{G}(1, \sigma_D^2 + \sigma_A^2)} \frac{1 - \rho}{\rho} \right]^{-1}, \end{aligned} \quad (3.1)$$

where for completeness, we introduced the spontaneous isomerization rate  $\rho_{SP}$ , measured in events per rod per time step. It mimics a photon event (see below).

Under the simplification that  $\sigma_A = 0$  and  $\rho_{SP} = 0$ , the probability that the rod absorbed a photon given the response is given by the well-known logistic function (Mackay, 2003),

$$g(y) = \frac{1}{1 + \exp[-(y - \theta)/\kappa]}, \quad (3.2)$$

with parameters  $\theta = \frac{1}{2} - \sigma_D^2 \ln \frac{\rho}{1-\rho}$  and  $\kappa = \sigma_D^2$ . If this probability is 50% or higher, a photon event was most likely, and the output is set to one; otherwise, there was likely no photon, and the output is set to zero. This threshold setting corresponds to the point where the rod probability distributions for the one photon and no photon signal intersect. The inference interpretation is equivalent to minimizing the number of errors done in the previous section and thus yields exactly the same optimal threshold.

When spontaneous events are taken into account, the threshold is approximated by

$$\theta = \frac{1}{2} - \sigma_D^2 \ln(\rho - \rho_{SP}). \quad (3.3)$$

This has no solution for  $\rho < \rho_{SP}$ ; the intuition is that any response was likely a spontaneous event rather than a real photon. When the assumption  $\sigma_A = 0$  is dropped,  $g(y)$  is no longer a monotonic function. Instead, a transition occurs at a negative value of  $y$ , which make  $g(y) = 1$  also for negative



$y$ . However, the probability of these small values of  $y$  is negligible. Furthermore, numerically the relevant upper threshold is virtually identical to the case that  $\sigma_A = 0$  (1.17 versus 1.19 when  $\rho = 10^{-4}$ ).

**3.3 Threshold from Signal-to-Noise Ratio.** Another performance measure of the pathway is the following: the discrimination should be clearest when the signal-to-noise ratio in the bipolar is maximal (Field & Rieke, 2002b). Therefore, the synapse should be tuned to maximize the signal-to-noise ratio.

Here we maximize the signal-to-noise ratio in a contrast discrimination task in which a dark patch has to be distinguished from a brighter one. For a given light level, the response distribution of the bipolar cell is  $Q(y; \rho) = [1 - q]\delta(y) + q\delta(y - 1)$ , where  $q(\rho) = \alpha_N + \rho N(1 - \alpha_N - \beta_N)$  is the average bipolar output, consisting of both correct and false responses. The variance of this distribution is  $q(1 - q)$ . The signal-to-noise ratio is

$$SNR(\rho_1, \rho_2) = \frac{2[q(\rho_1) - q(\rho_2)]^2}{q(\rho_1)[1 - q(\rho_1)] + q(\rho_2)[1 - q(\rho_2)]}.$$

The values of  $\rho_1$  and  $\rho_2$  are set as follows. When the discrimination is hardest, the discrimination between dark and highest light level is already difficult. Therefore, we will examine the case that  $\rho_1 = 0$  and  $\rho_2 = 2\rho$  (the factor 2 ensures that the mean light level is  $\rho$ ). However, we also consider the discrimination between two almost equal light levels:  $SNR(\rho - \delta\rho, \rho + \delta\rho)$  where  $\delta\rho \ll \rho$ . In practice, we found that this gave almost identical thresholds. In Figure 2B, the SNR is plotted for thresholding and for linear synapses followed by thresholding after summing. As the figure shows, the thresholding clearly improves the SNR.

**3.4 Threshold from Information Theory.** The detection problem and the need for a threshold in the synapse can also be studied using information theory. In general, the mutual information between an input variable  $x$  and an output  $y$  is  $IM = \int dx P(x) \int dy P(y|x) [\log_2 P(y|x) - \log_2 P(y)]$  (Cover & Thomas, 1991). We first calculate the mutual information between the light intensity and the bipolar signal as a function of the threshold. As above, we consider an input distribution with just two light intensities, 0 and  $2\rho$ , with each probability  $\frac{1}{2}$ .  $P(y, x)$  has four terms:  $P(0, 0) = \frac{1}{2}[1 - q(0)]$ ,  $P(1, 0) = \frac{1}{2}q(0)$ ,  $P(0, \rho) = \frac{1}{2}[1 - q(2\rho)]$ , and  $P(1, \rho) = \frac{1}{2}q(2\rho)$ . The mutual information therefore becomes a sum over four terms,

$$IMRHO = \frac{1}{2}[1 - q(0)] \left\{ \log_2[1 - q(0)] - \log_2 \left[ 1 - \frac{1}{2}q(0) - \frac{1}{2}q(2\rho) \right] \right\} \\ + \frac{1}{2}q(0) \left\{ \log_2 q(0) - \log_2 \left[ \frac{1}{2}q(0) + \frac{1}{2}q(2\rho) \right] \right\}$$

$$\begin{aligned}
& + \frac{1}{2}[1 - q(2\rho)] \left\{ \log_2[1 - q(2\rho)] - \log_2 \left[ 1 - \frac{1}{2}q(0) - \frac{1}{2}q(2\rho) \right] \right\} \\
& + \frac{1}{2}q(2\rho) \left\{ \log_2 q(2\rho) - \log_2 \left[ \frac{1}{2}q(0) + \frac{1}{2}q(2\rho) \right] \right\}.
\end{aligned}$$

This measure, labeled IMRHO, is our third performance criterion to set the threshold. The mutual information has a maximum as a function of the threshold level (see Figure 2C, solid line). Like the other criteria, the mutual information deteriorates when the signal is thresholded only after the rod signals have been summed (not shown). For sharp thresholds, the IMRHO is very similar to the SNR. In the case that the discrimination is done between  $\rho$  and  $\rho + \delta\rho$  with small  $\delta\rho$ , one can show by expansion in  $\delta\rho$  that they are identical.

Above, the information between light level and bipolar was used. Alternatively, one can optimize the mutual information between the actual photon signal and the bipolar signal. The photon signal is given by a Poisson process dependent on the light level. After all, one can argue that threshold should care only about the photons that are absorbed by the rod. We term this criterion IMROD. If just a single rod is connected to the bipolar,  $x$  describes the photon signal and  $y$  the bipolar output. Both  $x$  and  $y$  take values zero and one only. This does not mean the noise in the rod is ignored; it is captured in the  $\alpha$  and  $\beta$ .

$P(y, x)$  now has the terms  $P(0, 0) = (1 - \rho)(1 - \alpha)$ ,  $P(1, 0) = (1 - \rho)\alpha$ ,  $P(0, 1) = \rho\beta$ ,  $P(1, 1) = \rho(1 - \beta)$ . This yields

$$\begin{aligned}
IMROD = & (1 - \rho)\alpha \{ \log_2 \alpha - \log_2 [\alpha + \rho(1 - \alpha - \beta)] \} \\
& + (1 - \rho)(1 - \alpha) \{ \log_2(1 - \alpha) - \log_2 [1 - \alpha - \rho(1 - \alpha - \beta)] \} \\
& + \rho\beta \{ \log_2 \beta - \log_2 [1 - \alpha - \rho(1 - \alpha - \beta)] \} \\
& + \rho(1 - \beta) \{ \log_2(1 - \beta) - \log_2 [\alpha + \rho(1 - \alpha - \beta)] \}.
\end{aligned}$$

When instead of one rod,  $N$  rods are converging onto the bipolar cell,  $\alpha_N$  and  $\beta_N$  should be used and  $\rho$  should be replaced by  $\rho N$ .

This second mutual information measure reaches much higher values. This is understandable because unlike IMRHO, it lacks the Poisson process, which links the light level to actual photons. In the Poisson process, a lot of information is lost. Because the threshold setting does not affect the transformation of light level into absorbed photons, one could expect that both information measures have a similar dependence on the threshold. But this variant predicts consistently a lower threshold value (see Figure 2C, dashed curve).

#### 4 Optimal Threshold Levels

---

We have seen that the different performance criteria can yield different optimal threshold values. To gain a better understanding, we examined the optimal threshold as stimulus parameters are varied. The first observation is that for the binary transfer function, the SNR and IMRHO predict very similar thresholds. Figure 3A shows the optimal threshold for all criteria as the light level is varied. For high light levels, all approaches yield an optimal threshold close to 0.5 (although the approximations are expected to break down when  $\rho N \approx 1$ ). In practice, the minimal light level is limited by the dark-light to some  $10^{-3}$  events/rod/integration time, although behavioral responses can persist at even lower light levels. At these light levels, the different approaches are still quite similar.

To expose the differences more clearly, we purposefully neglected the spontaneous events and considered unrealistically low light levels. The threshold according to the SNR and ER is roughly linear in the log of the light level. The threshold value from IMROD is lower than for the SNR or ER. The intuition is that the mutual information approach prefers lower threshold values, because a low threshold yields a richer output distribution, although this increases the error rate.

Next, we examined the dependence on the number of rods converging on the bipolar cell. The threshold values depend only weakly on the number of rods (see Figure 3B). With increasing the number of rods, the thresholds come closer. Finally, the thresholds depend on the noise in the rods (see Figure 3C). The lower the noise, the smaller the threshold. This is easily understood as in the zero noise limit, where the discrimination is easy; a threshold of 1/2 would be best according to all criteria. For high noise, the ER threshold is proportional to  $\sigma^2$  (as shown above), whereas the IMROD threshold increases linearly with  $\sigma$ . Interestingly, for high noise, the optimal SNR threshold decreases after an initial increase.

As stated above, rod responses contain spontaneous rhodopsin isomerization events that have not been included so far. Effectively, this introduces additional false positives. The false-positive rate becomes  $\alpha_{SP} = (1 - \rho_{SP})\alpha + \rho_{SP}(1 - \beta)$ , where  $\rho_{SP}$  is the spontaneous isomerization rate measured in events per rod per time step. These events affect the various performance criteria differently. The ER predicts a higher threshold when the spontaneous events are included (see Figure 3D). In fact, the optimum threshold diverges when the mean light level approaches the spontaneous event level (see also equation 3.3). For light levels less than the spontaneous rate, there is no optimal threshold; the curve in Figure 2A has no minimum. Indeed, the fewest errors in that case are made when the output is always zero. In contrast, the other measures have a finite optimal threshold for light levels lower than the spontaneous rate. For the SNR, this is easily understood: in the presence of spontaneous rate, a discrimination task has to discriminate between  $\rho_1 + \rho_{SP}$  and  $\rho_2 + \rho_{SP}$ . Hence, the optimal

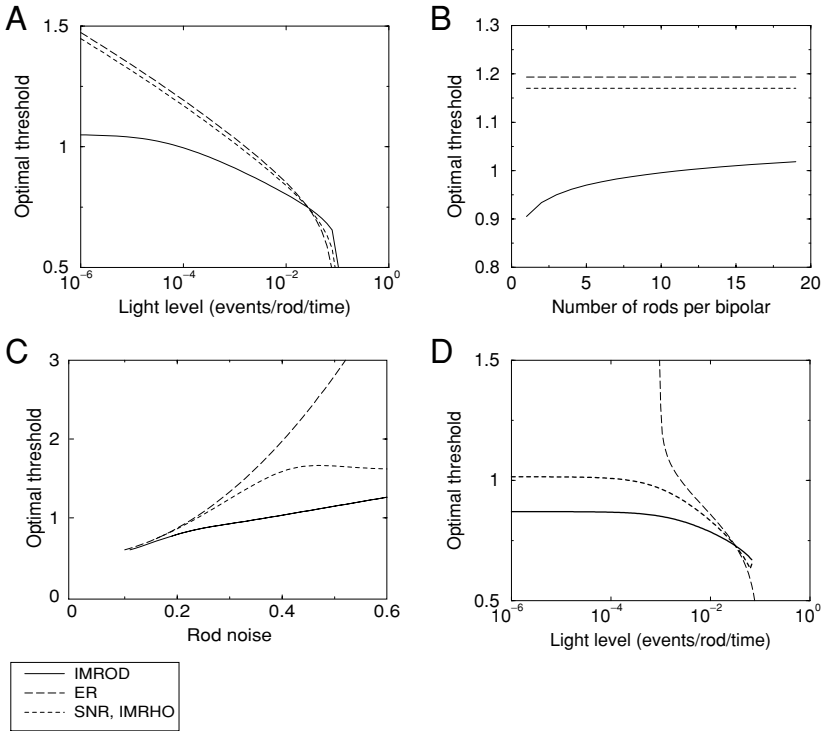


Figure 3: Dependence of the optimal threshold according to different performance criteria versus various parameters. Dashed line: the number of errors criterion (ER); dotted line: the signal-to-noise ratio (SNR) and the mutual information between light level and bipolar response (IMRHO) (overlapping); solid line: mutual information between rod and bipolar response (IMROD). (A) Optimal thresholds for as a function of the light level. Ten rods; noise as in Field and Rieke. (B) Optimal threshold value as a function of the number of rods. Light level of  $10^{-4}$  events per rod; noise as in Field and Rieke (2002b). (C) Dependence of the threshold value on the noise level in the rod. In this simulation, a simplified noise model was used, where the rod noise was independent of photon absorption (i.e.,  $\sigma_A = 0$ ). 10 rods,  $\rho = 10^{-4}$ . (D) Effect of spontaneous events on the optimal threshold. The spontaneous event rate was  $10^{-3}$  events per rod per time step. Other parameters as in A. Notably, the threshold based on the number of errors diverges when the light level is less than the spontaneous rate.

threshold shifts as if the light level were higher and equal to  $\rho + \rho_{SP}$  rather than  $\rho$ .

We tested whether the precise value of the time bin is important. In particular, the mutual information and its optimal threshold could depend

on the resolution of the sampling. We doubled the duration of the time bin. This manipulation doubles  $\rho$ , but also leads to a different value of  $\alpha$  and  $\beta$ . For light levels below 0.01 events per second there is no noticeable difference in the threshold for both ER and SNR. Only for the mutual information IMROD did we see a slightly higher threshold (1.02, compared to 0.99 for the original time bin;  $\rho = 10^{-4}$ ,  $N = 10$ ).

## 5 Simulated Rod-Bipolar Pathway

---

A priori it is not obvious which performance criterion should be used to set the synaptic threshold; all presented methods seem valid. To tackle this question, we simulated how the threshold setting would change the output of the bipolar system. It is likely that the different thresholds would lead to different visual percepts. These simulations allow us to examine the bipolar pathway as a function of the threshold level. The simulations consist of the following steps:

1. An image was split in rectangular regions, each corresponding to the receptive field of a single bipolar cell. The gray-scale of a certain pixel was extracted and multiplied with the mean light level to obtain the pixel's light level.
2. A Poisson process with a rate given by the light level determined if a rod absorbed a photon.
3. Gaussian noise was added to the rod response.
4. The nonlinear transfer function was applied to the rod response to mimic the synapse.
5. The transformed rod responses were summed in the bipolar.

We repeated this procedure for each bipolar cell, and the final output picture was averaged over many trials. This averaging mimics the pooling by the amacrine cells. Finally, we applied histogram equalization to the output picture using image processing software. This smooth, monotonic transformation improved visibility and gave the images a similar appearance despite very different mean output levels. Without it, the images can either easily saturate or become very dark. In the retina, such transformations can be performed by the circuitry of the amacrine cells and further downstream.

It is important to note that this simulated pathway is just an approximation of the real one, as the number of trials used here is much higher than the number of bipolars connected to the amacrine cell, and the bipolar-to-amacrine synapse might also contain a threshold; as in the bipolar, the signals are still quite sparse. These effects could change the results. Unfortunately, they are hard to study given our limited knowledge of processing by these circuits at the lowest light levels.

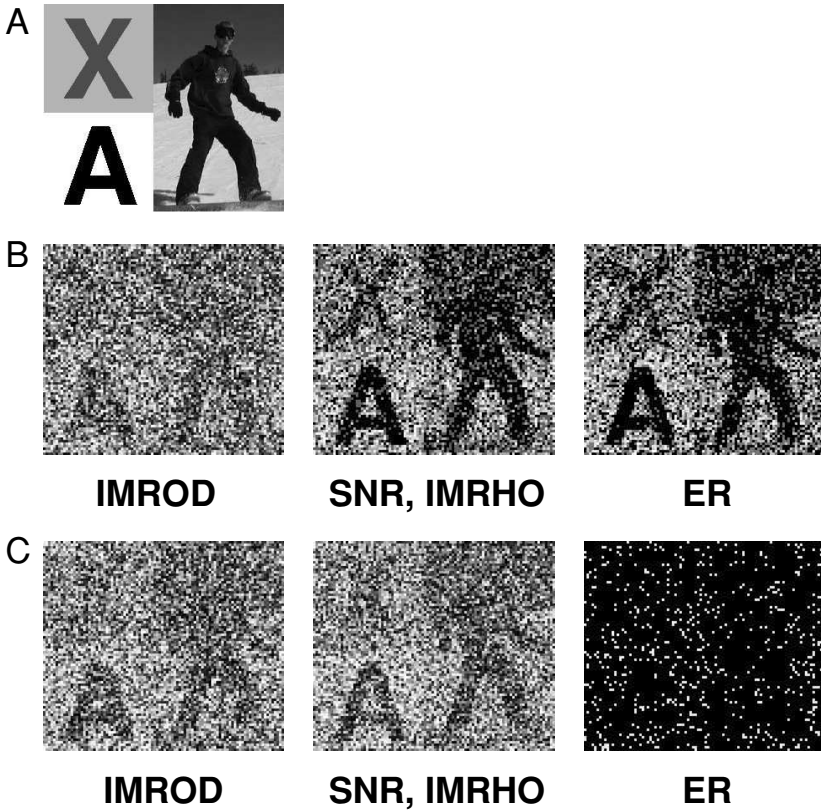


Figure 4: Simulated rod-bipolar image processing. (A) Original image. (B) Simulated images with a threshold level of the synapse 1.03 (optimal to maximize mutual information, IMROD), 1.33 (SNR and IMRHO), and 1.38 (ER). The threshold according to SNR gives a better-quality image than the image with the threshold according to IMROD. Average over 50,000 samples; mean light level  $10^{-5}$ ; 10 rods; noise according to Field & Rieke (2002b). (C) Same as in B except that the rod noise is higher. Now, minimizing the error rate (right-most figure) does not lead to a clear picture. On the other hand the IMROD criterion performs decently for these parameters. The threshold settings were  $\theta = 1.12$  (IMROD), 1.66 (SNR), and 2.78 (ER). In combination with A, the SNR and IMRHO yield consistently the clearest image.  $\sigma_D = 0.5$ ;  $\sigma_A = 0$ ;  $\rho = 10^{-4}$ ; 10 rods; average over 50,000 samples.

We applied the simulation on an input image with a high- and a low-contrast letter and a somewhat natural scene (see Figures 4A and 4B). The light level was deliberately chosen very low to emphasize the differences between the criteria. The low threshold level as predicted by IMROD leads to a high false-positive rate. As a result, the image is not very clear, and

low-contrast boundaries are hard to see. Setting the threshold according to the SNR (and the similar IMRHO) yields a clearer picture. Also, the ER yields a good output in this case, which is expected, as the predicted thresholds are very close.

In the above situation, the SNR and the ER criteria predict a very similar threshold level (see also Figure 3A). To further distinguish between the thresholds predicted by the SNR and the ER criteria, we simulated a situation with high rod noise,  $\sigma_D = 0.5$ ,  $\sigma_A = 0$ . In this case, the threshold according to the SNR is lower (see Figure 3C). Now the ER method performs worse, but the SNR still yields a good image (see Figure 4C). Combined, these results indicate that for the sharp threshold synapse, maximizing the SNR, or the almost identical IMRHO, consistently gives the best images in our simulation.

## 6 Performance with a Sigmoidal Synaptic Transfer Function

So far, a hard threshold function has been imposed. We wondered if a smoother threshold function would yield different results. We have not tried to derive the best possible transfer function, but a variable-slope parameter  $\kappa$  was added to the transfer function to make a logistic function  $g(x) = [1 + \exp(-(x - \theta)/\kappa)]^{-1}$ . When  $\kappa = 0$ , the sharp threshold is recovered.

When the transfer function is the smooth logistic function, the output of the bipolar becomes continuously valued. Both the SNR and the mutual information measures are easily calculated numerically when the transfer function is soft. It simply requires a discretization of the bipolar output in sufficiently small bins. The total number of errors ER is slightly ambiguous. We define it as follows: when no photon was absorbed but the bipolar voltage was larger than  $1/2$ , a false positive was counted, whereas in the opposite case, a false negative was counted.

We optimized the synapse by varying the slope of the transfer function in addition to the threshold value. Analytical treatment becomes intractable in this case, so we rely on numerical evaluation. The results are shown in Figure 5. The ER depends only very weakly on the smoothness (not shown). But the SNR and both mutual information measures improve when a smooth rather than a sharp transfer function is used. This can be seen by comparing the values at  $\kappa = 0$  (the hard threshold) with nonzero  $\kappa$ . On the other hand, the transfer function should not be taken as too smooth,  $\kappa \gtrsim 1$ ; otherwise, the performance decreases. In this limit of large  $\kappa$ , the synaptic transfer function becomes smooth and mimics a linear one, which, as was shown above, has poor performance. The optimal value for  $\kappa$  is close to  $\sigma^2$ , as expected from equation 3.2.

The optimal value for the threshold  $\theta$ , which now describes the transfer function's midpoint, increases for smoother transfer functions. Both IMRHO and IMROD have a broad plateau with a very shallow maximum.

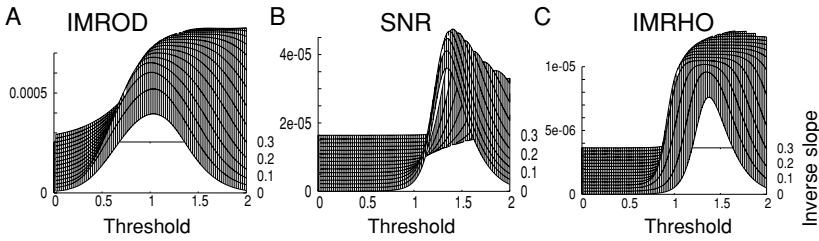


Figure 5: Performance criteria as a function of both the threshold  $\theta$  and the inverse slope  $\kappa$  of the synaptic transfer function. (A, C) IMROD and IMRHO increase with a smoother synapse and have broad maxima. (B) The SNR has a sharp maximum and decreases as the transfer function is made much smoother (higher inverse slope). Parameters as in Figure 4A.

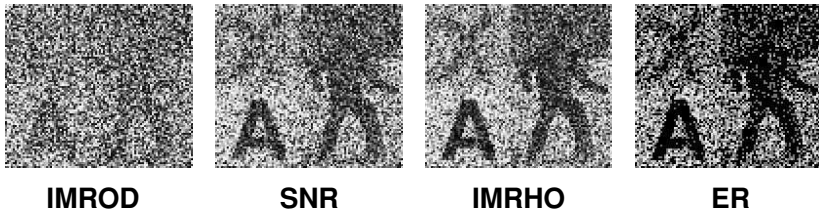


Figure 6: Simulated images when the synaptic transfer function is smooth. The optimal synapse settings were IMROD:  $\theta = 1.17$ ,  $\kappa = 0.14$ ; SNR:  $\theta = 1.37$ ,  $\kappa = 0.06$ ; IMRHO:  $\theta = 1.36$ ,  $\kappa = 0.11$ ; and ER:  $\theta = 1.38$ ,  $\kappa = 0$ . The other parameters as in Figure 4A.

The SNR has a sharper profile and has a maximum at rather low  $\kappa$ . Its optimal threshold shifts about 0.05 upward. The values of the optima are given in Figure 6.

Although the smoother synaptic transfer function increases the performance, inspection of the simulated images shows no obvious improvement (see Figure 6). This is not surprising as the increase is only some 20%, which is too small to yield significantly improved images. Likewise, the IMROD criterion still provides an image with many false positives. Another effect that occurs with the smooth transfer function is that the SNR and IMRHO no longer predict the same optimal threshold. However, this difference is too small to be visible in the images (compare SNR to IMRHO in Figure 6).

## 7 Discussion

We have studied the signal transfer in the first synapse of the visual system at low light levels. The presence of noise in the rods necessitates a strong nonlinearity in the synapse, as otherwise the continuously present noise



from the other rods would swamp the signal from a rod receiving a photon. We considered a variety of performance criteria that could be used to tune the synapse. At higher light levels, when the signal is not extremely sparse, the predicted thresholds are similar. But at low light levels, the performance is sensitive to the choice of criterion. There is no principal choice on which criterion to use a priori (Basseville, 1989).

Our results show that the predictions for the threshold are quite different at the lowest light levels. Which threshold, then, does the pathway use, and which threshold setting is the best? One possibility is to use the results from the simulated pathway, although these should be interpreted with care given the uncertainties in the circuitry. In the images, the signal-to-noise ratio (SNR) and the mutual information (IMRHO) consistently yield the best performance. The error rate (ER) (equivalent to Bayesian inference) is a reasonable criterion when rod noise is close to the measurements in Field and Rieke (2002b), but it predicts too high a threshold when the rod noise is larger. Interestingly, for good performance of the synapse, the mutual information needs to be calculated between the light level and the output (IMRHO), not between the photon signal and the output (IMROD). Otherwise, the threshold predicted is too low, and the pathway can have a very poor performance. Although this is not in conflict with information theory, it is a somewhat unexpected effect.

We have also explicitly included the effect of spontaneous events and the rod noise on the optimal synapse parameters. Finally, we considered a smooth transfer function and found that it is slightly better than a sharp one, but the difference is small.

An interesting question is how biology tunes and adapts the synapse according to the light level; we discussed some candidates earlier (van Rossum & Smith, 1998). This remains an outstanding issue experimentally and theoretically, as it emphasizes that the biology would need to optimize a quite noisy cost function. In the experiments, the threshold did not seem to adapt to the light level of the flash (Field & Rieke, 2002b), but the nonlinearity became weaker at higher background levels, increasing the response to flashes (Sampath & Rieke, 2004).

**7.1 Comparison to Earlier Work and Data.** We can compare our results to the data. Given the desired performance criterion, the current study predicts which threshold to expect at a given light level and noise level. Experimentally, however, the threshold and the other parameters (noise and convergence ratio) are hard to access. In the experiments, the transfer function and its threshold were not measured directly, but were inferred from the dependence of the mean and variance of the bipolar flash responses at higher light levels of about  $\rho \approx 1$ .

In Field and Rieke (2002b) the experimental data were described well when the transfer function was assumed to be a linear function with a step, that is,  $g(x) = x/[1 + \exp(-(x - \theta)/\kappa)]$ . The stimulus was a flash (about 1Hz)

at a flash intensity of  $10^{-4}\text{Rh}^*$ ; the mean light level was therefore some  $10^{-5}$  events per integration time. (We think it is more likely that the threshold adapts to the mean light level, not to the flash strength.) The experimentally observed threshold level found was 1.3. The optimal threshold for these parameters according to the SNR is 1.37 and a slope of 0.06 (see Figure 5). The observed inverse slope,  $\kappa = 0.1$ , in the data was close to the prediction from maximizing the SNR. As was already noted in Field and Rieke (2002b), the SNR gives a good prediction of the threshold. However, caution should be used; the vesicle release noise could effectively increase the rod noise, necessitating a higher threshold (see Figure 3C), whereas temporal filtering might reduce the noise. It is also not clear how the presumably omnipresent spontaneous events are consistent with these findings; including them would reduce the optimal threshold level (see Figure 3D).

However, in Berntson et al. (2004), the synaptic transfer was found to saturate when more than one photon was absorbed, as was assumed in this study. The reason for the discrepancy between the two experiments is not clear. For the current study, the difference in the actual shape of the transfer function is negligible, as the chance for multiple photon absorption in the rod is very small in the considered regime. However, it is likely that the different assumptions of the transfer function change the estimate for the threshold. This second experimental study found a threshold of 0.85 (Berntson et al., 2004). This seems to fit this study better when realistic spontaneous event rates are included.

Although in principle this study gives explicit predictions for the synaptic transfer function and its dependence on light level, the experiments and noise measurements are not sensitive enough to decide which performance criterion the biological synapse follows.

**7.2 Relevance to Other Systems.** This study seems to deal with a particular circuit and circumstances: the bipolar pathway at very low light levels. In order to have a beneficial effect of the threshold, the following conditions should occur: (1) the signal should be sparse (i.e., only one or a few inputs out of many are simultaneously active), (2) the signal should be discrete, (3) and all inputs carry noise in the absence of a signal.

However, the problem of detecting a sparse binary signal amid gaussian noise could be of much more general consideration. Consider, for instance, a view-invariant face cell in the cortex receiving many inputs, each of them active only when the face is seen from a particular angle. Given the ongoing spontaneous activity in neurons, such a system could need similar thresholding as the rod-bipolar pathway, in particular when the receptive field of the invariant cell is much wider than the tuning of the cells providing inputs.

It is not clear how far this analogy holds. Nonlinearities in the synaptic transfer, such as synaptic facilitation (Varela et al., 1997), seem too weak to provide the required nonlinearities. Another possibility is that by using

population coding, the system spreads the signal out over many inputs, relieving this problem.

### Acknowledgments

---

We thank Alexander Heimel, Fred Rieke, Chris Williams, and Robert Smith for insightful comments. Efstathios Politis helped in the initial phase of this project. The method of simulating low-light-level images was inspired by work of Andrew Hsu.

### References

---

- Armstrong-Gold, C. E., & Rieke, F. (2003). Bandpass filtering at the rod to second-order cell synapse in salamander (*ambystoma tigrinum*) retina. *J. Neurosci.*, *23*, 2796–2806.
- Barlow, H. B., Levick, W. R., & Yoon, M. (1971). Responses to single quanta of light in retinal ganglion cells of the cat. *Vision Research Supplement*, *3*, 87–101.
- Basseville, M. (1989). Distance measures for signal processing and pattern recognition. *Signal Processing*, *18*, 349–369.
- Baylor, D. A., Lamb, T., & Yau, K. W. (1979). Responses of retinal rods to single photons. *J. Physiol.*, *288*, 237–253.
- Baylor, D. A., Nunn, B. J., & Schnapf, J. L. (1984). The photocurrent, noise and spectral sensitivity of rods of the monkey *macaca fascicularis*. *J. Physiol.*, *357*, 575–607.
- Berntson, A., Smith, R. G., & Taylor, W. R. (2004). Transmission of single photon signals through a binary synapse in the mammalian retina. *Vis. Neurosci.*, *21*, 693–702.
- Bialek, W., & Owen, W. G. (1990). Temporal filtering in retinal bipolar cells: Elements of an optimal computation? *Biophys. J.*, *58*, 1227–1233.
- Copenhagen, D. R., Donner, K., & Reuter, T. (1987). Ganglion cell performance at absolute threshold in toad retina: Effect of dark events in rods. *J. Physiol.*, *393*, 667–680.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.
- Dacheux, R. F., & Raviola, E. (1986). The rod pathway in the rabbit retina: A depolarizing bipolar and amacrine cell. *J. Neurosci.*, *6*, 331–345.
- Field, G. D., & Rieke, F. (2002a). Mechanisms regulating variability of the single photon response of mammalian rod photoreceptors. *Neuron*, *35*, 733–747.
- Field, G. D., & Rieke, F. (2002b). Nonlinear signal transfer from mouse rods to bipolar cells in implications for visual sensitivity. *Neuron*, *34*, 773–785.
- Field, G. D., Sampath, A. P., & Rieke, F. (2005). Retinal processing near absolute threshold: From behavior to mechanism. *Ann. Rev. Physiol.*, *67*, 491–514.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Grünert, U., Martin, P. R., & Wässle, H. (1994). Immunocytochemical analysis of bipolar cells in the macaque monkey retina. *Journal of Comparative Neurology*, *348*, 607–627.

- Mackay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge: Cambridge University Press.
- Mastrorarde, D. N. (1983a). Correlated firing of cat retinal ganglions cells: I. Spontaneously active inputs to X- and Y-cells. *J. Neurophysiol.*, *49*, 303–324.
- Mastrorarde, D. N. (1983b). Correlated firing of cat retinal ganglions cells: II. Responses of X- and Y-cell to single quantal events. *J. Neurophysiol.*, *49*, 325–349.
- Rao, R., Buchsbaum, G., & Sterling, P. (1994). Rate of quantal transmitter release at the mammalian rod synapse. *Biophys. J.*, *67*, 57–64.
- Sampath, A. P., & Rieke, F. (2004). Selective transmission of single photon responses by saturation at the rod-to-rod bipolar synapse. *Neuron*, *41*, 431–443.
- Schneeweis, D. M., & Schnapf, J. L. (1995). Photovoltage in rods and cones in the macaque retina. *Science*, *268*, 1053–1056.
- Soucy, E., Wang, Y., Nirenberg, S., Nathans, J., & Meister, M. (1998). A novel signaling pathway from rod photoreceptors to ganglion cells in mammalian retina. *Neuron*, *21*, 481–493.
- Sterling, P., & Demb, J. B. (2004). Retina. In G. M. Shepherd (Ed.), *Synaptic organization of the brain*. New York: Oxford University Press.
- Sterling, P., Freed, M., & Smith, R. G. (1988). Architecture of rod and cone circuits to the on-beta ganglion cell. *J. Neurosci.*, *8*, 623–642.
- Tsukamoto, Y., Morigiwa, K., Ueda, K., & Sterling, P. (2001). Microcircuits for night vision in the mouse retina. *J. Neurosci.*, *21*, 8616–8623.
- van Rossum, M. C. W., & Smith, R. G. (1998). Noise removal at the rod synapse of mammalian retina. *Vis. Neurosci.*, *15*, 809–821.
- Van Trees, H. L. (1968). *Detection, estimation, and modulation theory I*. New York: Wiley.
- Varela, J. A., Sen, K., Gibson, J., Fost, J., Abbott, L. F., & Nelson, S. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J. Neurosci.*, *17*, 7926–7940.
- Völgyi, B., Deans, M. R., Paul, D. L., & Bloomfield, S. A. (2004). Convergence and segregation of the multiple rod pathways in mammalian retina. *J. Neurosci.*, *24*, 11182–11192.
- Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A., & Schnapf, J. L. (1990). The control of visual sensitivity. In L. Spillmann and S. J. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 53–101). San Diego, CA: Academic Press.