

Benefits of Pathway Splitting in Sensory Coding

Julijana Gjorgjieva,¹ Haim Sompolinsky,^{1,2} and Markus Meister³

¹Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, ²Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem, 91904, Israel, and ³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125

In many sensory systems, the neural signal splits into multiple parallel pathways. For example, in the mammalian retina, ~20 types of retinal ganglion cells transmit information about the visual scene to the brain. The purpose of this profuse and early pathway splitting remains unknown. We examine a common instance of splitting into ON and OFF neurons excited by increments and decrements of light intensity in the visual scene, respectively. We test the hypothesis that pathway splitting enables more efficient encoding of sensory stimuli. Specifically, we compare a model system with an ON and an OFF neuron to one with two ON neurons. Surprisingly, the optimal ON–OFF system transmits the same information as the optimal ON–ON system, if one constrains the maximal firing rate of the neurons. However, the ON–OFF system uses fewer spikes on average to transmit this information. This superiority of the ON–OFF system is also observed when the two systems are optimized while constraining their mean firing rate. The efficiency gain for the ON–OFF split is comparable with that derived from decorrelation, a well known processing strategy of early sensory systems. The gain can be orders of magnitude larger when the ecologically important stimuli are rare but large events of either polarity. The ON–OFF system also provides a better code for extracting information by a linear downstream decoder. The results suggest that the evolution of ON–OFF diversification in sensory systems may be driven by the benefits of lowering average metabolic cost, especially in a world in which the relevant stimuli are sparse.

Key words: efficient coding; ON–OFF; optimality; parallel pathways; retina; sensory processing

Introduction

Retinal ganglion cells are the output neurons of the retina and convey all visual information from the eye to the brain. Just two synapses separate these neurons from the photoreceptors, and yet the population of ganglion cells already splits into ~20 different types. Each type covers the entire visual field and sends a separate neural image to higher brain areas for additional processing (Wässle, 2004; Masland, 2012). What factors drove the evolution of such an early and elaborate pathway split remains mysterious. Here we test some explanations based on the principles of efficient coding.

We focus on a specific prominent instance of sensory splitting: the emergence of ON and OFF pathways. Among retinal ganglion cells, several of the well known cell types come in matched pairs that share similar spatiotemporal response properties, except that one type is excited by light increments (ON) and the other type by light decrements (OFF) (Wässle, 2004; Wässle et al., 1981a,b). A similar ON–OFF dichotomy is observed in other sensory modal-

ities, including insect vision (Joesch et al., 2010), thermosensation (Gallio et al., 2011), chemosensation (Chalasanani et al., 2007), audition (Scholl et al., 2010), and electrolocation in electric fish (Bennett, 1971). The broad prevalence of this organizing principle suggests that it provides an evolutionary fitness benefit of a very general nature. In the present work, we treat the problem in the context of vision but seek an explanation that generalizes beyond this specific sensory system while ignoring details that are often specific to one neural system or another.

Previous investigations of this question suggested that the ON–OFF split evolved for the rapid and metabolically efficient signaling of opposite changes in light intensity, because both increments and decrements are prominent in natural scenes (Schiller et al., 1986; Schiller, 1992; Westheimer, 2007). Here we formalize such an argument by testing different notions of efficiency, by examining different models of the neural response and different assumptions about the relevant stimuli to be encoded. Throughout, we compare side by side the performance of a model system consisting of an ON and an OFF cell with that of a system with two ON cells. The results support the notion that ON–OFF splitting enables more efficient neural coding; the benefits for ON–OFF coding already appear in the simplest possible coding model and become more pronounced as we make the model more realistic. If the dominant constraint on neural signaling comes from a limit on the maximal firing rate of cells, then pathway splitting into ON and OFF channels does not deliver greater information, but it uses fewer spikes than the alternate coding scheme with two ON cells. As a result, if the time-average firing rate is constrained, ON–OFF coding offers an improvement in information of ~15%; this holds under a broad range of stimuli

Received March 14, 2014; revised July 3, 2014; accepted July 22, 2014.

Author contributions: J.G., H.S., and M.M. designed research; J.G., H.S., and M.M. performed research; J.G. analyzed data; J.G., H.S., and M.M. wrote the paper.

This work was supported by a grant from the National Institutes of Health (J.G., H.S. and M.M.) and grants from the Gatsby Charitable Foundation and the Swartz Foundation (H.S.). J.G. thanks Xaq Pitkow for providing help with the sigmoidal nonlinearities from retinal data.

The authors declare no competing financial interests.

Correspondence should be addressed to either of the following: Julijana Gjorgjieva, Center for Brain Science, Harvard University, Cambridge, MA 02138, E-mail: gjorgjieva@fas.harvard.edu; or Markus Meister, Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, E-mail: meister@caltech.edu.

DOI:10.1523/JNEUROSCI.1032-14.2014

Copyright © 2014 the authors 0270-6474/14/3412127-18\$15.00/0

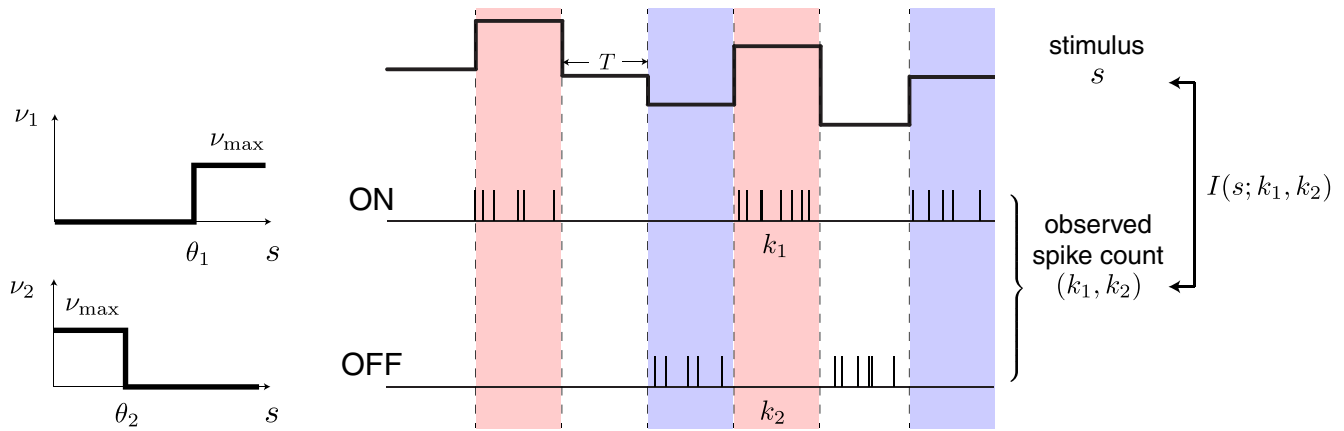


Figure 1. Modeling framework. A stimulus s is encoded by a system of two cell types: one ON and OFF (as shown) or two ON cells. Each cell has a binary response nonlinearity with thresholds θ_1 or θ_2 . During each coding window of duration T , the stimulus is constant and the spike count is drawn from a Poisson distribution. A measure of coding efficiency here is the mutual information between the stimulus s and the spike count responses of the two cells k_1 and k_2 , $I(s; k_1, k_2)$.

and response models, including linear decoding by downstream neurons, when the benefits become more pronounced. The largest benefit for ON–OFF splitting emerges if ecologically important stimuli consist of rare events with large ON or OFF signals embedded in a sea of events with low variance. Our results show that ON–OFF coding improves information transmission at low average metabolic costs under a broad set of conditions, with the most pronounced benefits appearing for a sparse stimulus ensemble.

Materials and Methods

Mutual information with binary nonlinearities and Poisson output noise. We first studied how static stimuli are encoded by ON–OFF and ON–ON systems. Each ganglion cell fires action potentials in response to a common scalar stimulus s drawn from a Gaussian distribution $P_s(s)$. We model each ganglion cell with a firing rate $\nu = g(s)$, a binary function with threshold θ , and a maximal value of ν_{\max} (Fig. 1). For instance, $\nu = \nu_{\max} \Theta(s - \theta)$ for ON type cells and $\nu = \nu_{\max} \Theta(\theta - s)$ for OFF type cells, where $\Theta(x)$ is the Heaviside function that is 0 if $x < 0$ and 1 if $x \geq 0$. In a system with two cells, each cell can have a different threshold, θ_1 and θ_2 .

The observable variable in our model is the number of spikes k in a time period T . We assume Poisson output noise; thus, the probability of observing k spikes given a firing rate ν can be written as

$$P(k|\nu) = \frac{[\nu T]^k}{k!} e^{-\nu T}. \tag{1}$$

The mutual information between the stimulus s and the observed spike count k for a single cell can be obtained from the equation

$$I(s;k) = \sum_{k=0}^{\infty} \int_s P_s(s) P(k|s) \log_2 \frac{P(k|s)}{P_k(k)} ds, \tag{2}$$

where $P_k(k)$ is the output distribution of the observed spike count, which can be calculated from $P_k(k) = \int_s P_s(s) P(k|s) ds$. In the absence of input noise, knowing the stimulus s unambiguously determines the response firing rate ν ; for instance, for an ON cell, if $s < \theta$, $\nu = 0$, and if $s \geq \theta$, $\nu = \nu_{\max}$. Therefore, in Equation 2, we can replace $P(k|s)$ with $P(k|\nu)$ from Equation 1.

For a binary response function with two firing rate levels, 0 and ν_{\max} , we can lump together all states with nonzero spike counts into a single state, which we denote as **1**. Correspondingly, the state with zero spikes is **0**. The reason for this simplification is that, if the ON (OFF) neuron emits at least one spike, then the stimulus must necessarily be above (below) its threshold, and the number of spikes does not yield any additional infor-

mation about the stimulus. Hence, we can evaluate the mutual information between stimulus and spiking response using the following expressions for the spike count probabilities:

$$\begin{aligned} P(0|\nu = 0) &= 1, & P(1|\nu = 0) &= 0, \\ P(0|\nu = \nu_{\max}) &= 1 - r, & P(1|\nu = \nu_{\max}) &= r, \end{aligned} \tag{3}$$

where $r = 1 - e^{-N_{\max}}$ and $N_{\max} = \nu_{\max} T$ is the expected spike count in a time period T given the maximal firing rate ν_{\max} of the cell.

The mutual information between the stimulus and the spiking response of the two cells can be written as

$$I(s;k_1,k_2) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \int_s P_s(s) P(k_1,k_2|s) \log_2 \frac{P(k_1,k_2|s)}{\int_{s'} P(s') P(k_1,k_2|s') ds'} ds. \tag{4}$$

When we compute the mutual information for two cells with only output noise, we assume that stimulus encoding by the two neurons is statistically independent conditional on the stimulus s ; Pitkow and Meister (2012) found that noise correlations arising from overlapping receptive fields in salamander ganglion cells were very small. Therefore, $P(k_1,k_2|s) = P(k_1|s)P(k_2|s)$, and as above, in the absence of input noise, we can replace $P(k_j|s)$ with $P(k_j|\nu_j)$ for $j = 1, 2$ from Equation 3.

With a binary step-shaped rate function, the mutual information depends on the stimulus distribution only through the probability of having a firing rate of ν_{\max} , which for an ON cell is $P_{\text{ON}}(\theta) = \int_{\theta}^{\infty} ds' P_s(s')$

and for an OFF cell is $P_{\text{OFF}}(\theta) = \int_{-\infty}^{\theta} ds' P_s(s')$. We can thus replace θ by

the corresponding cumulative probability, which essentially maps the stimulus distribution into a uniform distribution from 0 to 1. The resulting states for the ON–OFF and ON–ON systems are shown schematically in Figure 2. In this work, we determine the thresholds by maximizing the mutual information as a function of these variables, and because the stimulus dependence enters only through these values, the maximal mutual information is independent of the stimulus distribution, provided that the stimulus cumulative distribution is continuous.

We first consider the case where the value of N_{\max} is fixed. Maximizing the mutual information with respect to the thresholds, we obtain a symmetric solution for the ON–OFF system, with $\theta_1 > \theta_2$ and $\theta_2 = 1 - \theta_1$, where θ_1 is the threshold for the ON cell and θ_2 for the OFF cell. In fact, for any N_{\max} , the optimal solution is for the ON and OFF cells to have cells with non-overlapping response ranges, i.e., $\theta_1 > \theta_2$ (Fig. 3). Thus, for the ON–OFF system, the mutual information only depends on one of the thresholds:

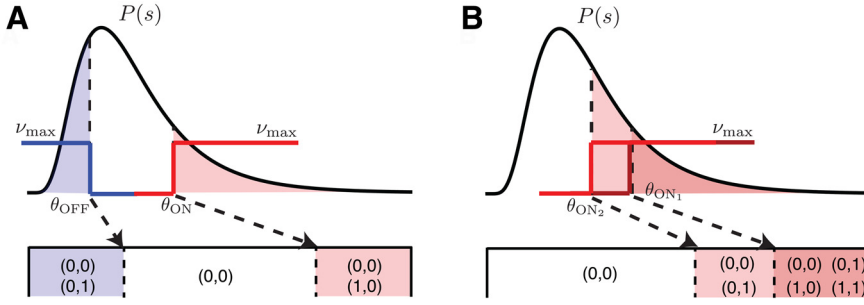


Figure 2. Output states in the binary neuron model. An arbitrary stimulus distribution $P(s)$ can be mapped into a uniform distribution so that the optimal thresholds can be depicted in terms of the fraction of stimuli below threshold, ranging from 0 to 1. **A**, Possible output states for the two neurons in the ON–OFF system, depending on the stimulus value. 0 denotes silence, 1 denotes one or more spikes in the coding window. **B**, Same as **A** but for two ON cells.

$$I_{\text{ON-OFF}} = -[1 - 2r(1 - \theta_1)] \log_2[1 - 2r(1 - \theta_1)] + 2(1 - \theta_1)(1 - r) \log_2[1 - r] - 2(1 - \theta_1)r \log_2[1 - \theta_1]. \quad (5)$$

This quantity is maximized for thresholds θ_1 for the ON cell and $\theta_2 = 1 - \theta_1$ for the OFF cell:

$$\theta_1 = 1 - \frac{1}{2r + (1 - r)^{(r-1)/r}} \quad \text{and} \quad \theta_2 = \frac{1}{2r + (1 - r)^{(r-1)/r}}. \quad (6)$$

Similarly, we obtain the same maximized mutual information transmitted by the ON–ON system but achieved with thresholds:

$$\theta_1 = 1 - \frac{1}{2r + (1 - r)^{(r-1)/r}} \quad \text{and} \quad \theta_2 = 1 - \frac{1 + r}{2r + (1 - r)^{(r-1)/r}}. \quad (7)$$

For the system with two identical ON cells (2ON) with threshold θ , the mutual information is equal to

$$I_{2\text{ON};\text{ident}} = -[\theta + (1 - \theta)(1 - r)^2] \log_2[\theta + (1 - \theta)(1 - r)^2]. \quad (8)$$

The information is maximized for an optimal threshold equal to

$$\theta_{\text{ident}} = \frac{1 - (1 - r)^{\frac{2}{r(2-r)}}}{1 + (1 - r)^{\frac{2}{r(2-r)}} r(2 - r)}. \quad (9)$$

The maximal mutual information for the system with a non-monotonic nonlinearity for the so-called U-cell, $\nu = \nu_{\text{max}} \Theta(s - \theta_1) \Theta(\theta_2 - s)$, was computed using Equation 4 by numerically iterating through a grid of thresholds θ_1 and θ_2 for the U-cell and θ_3 for the ON cell.

Instead of the maximal firing rate of the cells, ν_{max} , we can alternatively constrain the overall mean firing rate of each system, ν_{mean} , or the total mean spike count of each system, $N_{\text{mean}} = \nu_{\text{mean}} T$. In addition to the thresholds, here we also need to optimize the allocation of N_{mean} to each of the cells in the system. As in the case of constraining N_{max} , we find that the optimal solution for the ON–OFF system is symmetric, such that $\theta_2 = 1 - \theta_1$, and the two cells have response ranges that do not overlap for the entire range of N_{mean} , i.e., $\theta_1 > \theta_2$. This implies that the two thresholds divide stimulus space into three regions: one region coded only by the ON cell, another coded only by the OFF cell, and a region that is not coded by either cell. The mean spike count for each cell is one-half of the total mean spike count. For the ON–ON system, the fraction, f , of the total mean spike count to be assigned to the cell with the lower threshold was generally greater than 1/2 (Fig. 3C, inset).

To compute the mutual information from Equation 2, we expressed N_{max} as a function of the total mean spike count N_{mean} . For an ON cell with a larger threshold θ_1 , we can write $N_{\text{max}} = (1 - f)N_{\text{mean}}/(1 - \theta_1)$,

and for an ON cell with a smaller threshold θ_2 , we can write $N_{\text{max}} = fN_{\text{mean}}/(1 - \theta_2)$, where $0 \leq f \leq 1$.

Using sigmoidal nonlinearities and sub-Poisson noise from retinal data. We examined sigmoidal nonlinearities fitted from recorded retinal ganglion cells (Pitkow and Meister, 2012) of the form (see Fig. 6)

$$g_{\text{ON}}(s) = \nu_{\text{max}} \frac{1}{1 + e^{-\beta(s - \theta_{\text{ON}})}} \quad (10)$$

to describe the firing rate of an ON cell with threshold θ_{ON} as a function of the stimulus s , and

$$g_{\text{OFF}}(s) = \nu_{\text{max}} \frac{1}{1 + e^{\beta(s - \theta_{\text{OFF}})}} \quad (11)$$

to describe the firing rate of an OFF cell with threshold θ_{OFF} . The sigmoids have gain β and a maximal firing rate of ν_{max} . We studied nonlinearities with the same gain β for both ON and OFF cells.

Here we assumed an output noise distribution that matches the observed sub-Poisson noise in spike counts of salamander retinal ganglion cells. For a given mean spike count c , the measured spike count distribution $P(k|c)$ had a width that stayed constant with c after an initial Poisson-like growth (Pitkow and Meister, 2012, their Supplemental Fig. 3). The distribution is well described by the heuristic formula

$$P(k|c) \propto \exp\left[-\frac{(k - k_0(c))^2}{2\sigma^2}\right], \quad k = 0, 1, 2, \dots \quad (12)$$

where

$$k_0(c) = a \log(-1 + e^{c/a}), \quad (13)$$

with $a = 0.5$ and $\sigma = 0.75$ (Pitkow and Meister, 2012). We computed the mutual information by numerically evaluating the integrals arising in Equation 4, where the spike counts k_1 and k_2 of the two cells were summed up to some cutoff.

Figure 6 shows the results for two nonlinearities that were fit using experimental data from salamander (Pitkow and Meister, 2012) and macaque (Uzzell and Chichilnisky, 2004) ganglion cells. The average gain for salamander ganglion cells is $\beta = 5.8$ (expressed in inverse SD of the stimulus) and for macaque ganglion cells $\beta = 2.3$. Pitkow and Meister (2012) cite a median peak firing rate of 48 Hz for salamander ganglion cells, which using a coding time window of 50 ms translates into a maximum expected spike count $N_{\text{max}} = 2.4$. The median peak firing rate for macaque ganglion cells is 220 Hz, but these have a shorter coding time window of 10 ms, producing $N_{\text{max}} = 2.2$. The measured mean firing rates for each species are 1.1 Hz for salamander and 30 Hz for macaque (Uzzell and Chichilnisky, 2004; Pitkow and Meister, 2012).

Input noise. In addition to output Poisson noise, in a subset of our results, we also studied the effect of common input Gaussian noise with variance σ^2 (see Fig. 7). Then the firing rate for an ON neuron can be written as

$$\nu_1(s) = \nu_{\text{max}} \Theta(s + z - \theta_1) \quad (14)$$

and for an OFF neuron

$$\nu_2(s) = \nu_{\text{max}} \Theta(\theta_2 - s - z) \quad (15)$$

where $z \sim N(0, \sigma^2)$. The information was computed using Equation 4 with the conditional probability distribution of the two cells given by

$$P(k_1, k_2 | s) = \sum_{\nu_1 \in \{0, \nu_{\max}\}} \sum_{\nu_2 \in \{0, \nu_{\max}\}} P(k_1 | \nu_1) P(k_2 | \nu_2) P(\nu_1, \nu_2 | s) \quad (16)$$

because now the two cells are correlated conditioned on the stimulus s . In this expression, $P(k_j | \nu_j)$ can be computed from Equation 3 for $j = 1, 2$. Furthermore, $P(\nu_1, \nu_2 | s)$ can be computed by integrating the firing rates of the two cells over the Gaussian distribution with variance σ^2 .

In the text, we studied the effects of stimulus sparseness by using two distributions: a Gaussian and a Laplace stimulus distribution with a mean of 0 and SD of 1. The Laplace distribution is consistent with fits produced by filtering natural images with difference-of-Gaussian linear filters, corresponding to center-surround receptive fields (Field, 1994; Bell and Sejnowski, 1997).

To implement the sigmoidal nonlinearities with Poisson output noise in Figure 5, we used the same input noise model but assumed that the input noise for each cell was independent, i.e., we added a noise term z_1 to the stimulus for one cell and z_2 to the stimulus for the other cell, with $z_1, z_2 \sim N(0, \sigma^2)$. Now, for the mutual information in Equation 4, the conditional probability distribution of the two cells is (because the firing rates are independent conditioned on s)

$$P(k_1, k_2 | s) = P(k_1 | s) P(k_2 | s) = \left(\sum_{\nu_1 \in \{0, \nu_{\max}\}} P(k_1 | \nu_1) P(\nu_1 | s) \right) \left(\sum_{\nu_2 \in \{0, \nu_{\max}\}} P(k_2 | \nu_2) P(\nu_2 | s) \right) \quad (17)$$

where $P(k_j | \nu_j)$ can be computed from Equation 3 for $j = 1, 2$, and $P(\nu_j | s)$ was computed by integrating the firing rate of the cell over the Gaussian distribution with variance σ^2 ; for instance, for an ON cell with threshold θ_1 ,

$$P(\nu_1 = \nu_{\max} | s) = H\left(\frac{\theta_1 - s}{\sigma}\right), \quad (18)$$

and $P(\nu_1 = 0 | s) = 1 - P(\nu_1 = \nu_{\max} | s)$, where H is the complementary error function

$$H(\theta) = \frac{1}{2} \operatorname{erfc}\left(\frac{\theta}{\sqrt{2}}\right) = \int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz. \quad (19)$$

Now the effective sigmoidal nonlinearity of the cells is given by Equation 18. The gain of these sigmoids can be mapped to the gain of the sigmoidal nonlinearities fitted from the retinal data (Eq. 10) by $\beta = 1/(\sigma\sqrt{2\pi})$.

Linear estimator: static. As an alternative measure of coding efficiency to the mutual information, we minimized the mean square error (MSE) between the static stimulus s and a reconstructed version of the stimulus s_{est} (see Fig. 8):

$$E = \langle (s_{\text{est}} - s)^2 \rangle, \quad (20)$$

where the angle brackets denote average over stimulus values. The stimulus estimate was obtained from the observed ganglion cell responses, $x_1 - \langle x_1 \rangle$ and $x_2 - \langle x_2 \rangle$ (normalized to have zero mean), using appropriate decoding weights, w_1 and w_2 :

$$s_{\text{est}} = \sum_{i=1,2} w_i (x_i - \langle x_i \rangle). \quad (21)$$

The MSE can be written as:

$$E = \langle s^2 \rangle + w^T C w - 2w^T U, \quad (22)$$

where C denotes the matrix of pairwise correlations between the responses of the cells, and U denotes the correlation between response and stimulus, i.e.,

$$C_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle, \quad \text{and} \quad U_i = \langle s x_i \rangle - \langle s \rangle \langle x_i \rangle. \quad (23)$$

Differentiating the MSE in Equation 22 with respect to the weights w , we obtained the expressions for the optimal weights:

$$w = C^{-1} U, \quad (24)$$

so that the error can also be written as

$$E = \langle s^2 \rangle - U^T C^{-1} U. \quad (25)$$

For the ON–OFF system, we derived the following expression for the optimal weights:

$$w_1 = -w_2 = \frac{\langle x_1 s \rangle}{\langle x_1 x_1 \rangle} = \frac{\int_{\theta}^{\infty} s P_s(s) ds}{(1 + N_{\max}) \int_{\theta}^{\infty} P_s(s) ds} \quad (26)$$

and for this value, the optimal MSE is

$$E_{\text{ON-OFF}} = \langle s^2 \rangle - 2 \frac{N_{\max}}{1 + N_{\max}} \frac{\left(\int_{\theta}^{\infty} s P_s(s) ds \right)^2}{\int_{\theta}^{\infty} P_s(s) ds}. \quad (27)$$

For the ON–ON system, we derived the optimal weights w_1 (ON cell with threshold θ_1) and w_2 (ON cell with threshold θ_2) as

$$w_1 = \frac{J_1 H_2 (1 + N_2 - N_2 H_2) - J_2 N_2 (H_3 - H_1 H_2)}{H_1 (1 + N_1 - N_1 H_1) H_2 (1 + N_2 - N_2 H_2) - N_1 N_2 (H_3 - H_1 H_2)^2}, \quad (28)$$

$$w_2 = \frac{J_2 H_1 (1 + N_1 - N_1 H_1) - J_1 N_1 (H_3 - H_1 H_2)}{H_1 (1 + N_1 - N_1 H_1) H_2 (1 + N_2 - N_2 H_2) - N_1 N_2 (H_3 - H_1 H_2)^2} \quad (29)$$

where

$$H_1 = \int_{\theta_1}^{\infty} P_s(s) ds, \quad H_2 = \int_{\theta_2}^{\infty} P_s(s) ds \quad \text{and} \quad H_3 = \int_{\max(\theta_1, \theta_2)}^{\infty} P_s(s) ds$$

and

$$J_1 = \int_{\theta_1}^{\infty} s P_s(s) ds \quad \text{and} \quad J_2 = \int_{\theta_2}^{\infty} s P_s(s) ds$$

and for this value, the optimal MSE is

$$E_{\text{ON-ON}} = \langle s^2 \rangle - \frac{J_2^2 H_1 N_2 (1 + N_1 - N_1 H_1) + J_1^2 H_2 N_1 (1 + N_2 - N_2 H_2) - 2N_1 N_2 J_1 J_2 (H_3 - H_1 H_2)}{H_1 (1 + N_1 - N_1 H_1) H_2 (1 + N_2 - N_2 H_2) - (H_3 - H_1 H_2)^2 N_1 N_2}. \quad (30)$$

Note that here N_1 and N_2 are the maximal firing rates of cell 1 and cell 2, respectively. In the case of constraining the maximal spike count to N_{max} , then $N_1 = N_2 = N_{\text{max}}$. When the total mean spike count is constrained, then N_1 and N_2 are optimized. Given the MSE, the signal-to-noise ratio (SNR) of the reconstruction was defined as

$$\text{SNR} = \frac{\langle s^2 \rangle}{\langle E \rangle} = \frac{\langle s^2 \rangle}{\langle (s - s_{\text{est}})^2 \rangle}. \quad (31)$$

Again, we studied the effect of stimulus sparseness by using two static stimulus distributions, Gaussian and Laplace, with mean 0 and SD 1 (see Fig. 8).

Linear estimator: dynamic. We also studied how dynamic stimuli are encoded by ON-OFF and ON-ON systems (see Figs. 9, 10). Dong and Atick (1995) have shown that, in the regime of low spatial frequencies, the temporal part of the power spectrum of natural movies falls off as a power law with exponent -2 . Therefore, we modeled temporal stimuli as white noise low-pass filtered with a time constant of $\tau = 100$ ms, longer than all other time constants in the system (i.e., the temporal biphasic filter below). The input into the ganglion cell nonlinearity is a linearly filtered version, $g(t)$, of the temporally correlated stimulus, $s(t)$, with a temporally biphasic filter:

$$f(t) = \frac{1}{\tau_1^4} t^3 e^{-t/\tau_1} - \rho \frac{1}{\tau_2^4} t^3 e^{-t/\tau_2} \text{ for } t \geq 0, \quad (32)$$

and 0 otherwise, where $\tau_1 = 5$ ms, $\tau_2 = 15$ ms, and $\rho = 0.8$ were taken from fits to macaque ganglion cells (Chichilnisky and Kalmar, 2002; Pitkow et al., 2007). This filter represents the temporal processing by bipolar and amacrine cell circuits preceding ganglion cells in the retina (see Fig. 9).

The MSE and the optimal decoding filters (which in this case have temporal structure) were derived using Equations 22 and 24 (Fig. 10). For simplicity, we provide expressions for these filters and the MSE in the Fourier domain, but results in the main text are presented in the time domain (see Figs. 9, 10). Note that we define the Fourier transform of a temporal signal $x(t)$ as $X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt$ and the inverse Fourier transform as $x(t) = 1/(2\pi) \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega$. We found that the optimal ON and OFF filters were symmetric:

$$F_{\text{ON}}(\omega) = -F_{\text{OFF}}(\omega) = \frac{Q(\omega) \frac{e^{-\theta^2/2}}{\sqrt{2\pi\sigma^2}}}{H(\theta) + N_{\text{max}}(G_1(\omega) - G_3(\omega))} \quad (33)$$

and the optimal MSE

$$E_{\text{ON-ON}} = \langle s^2(t) \rangle \left(1 - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|Q(\omega)|^2 \left(N_1 [H(\theta_2) + N_2 G_2(\omega)] \frac{e^{-\theta_1^2}}{2\pi\sigma^2/\tau_1} + N_2 [H(\theta_1) + N_1 G_1(\omega)] \frac{e^{-\theta_2^2}}{2\pi\sigma^2/\tau_1} - 2N_1 N_2 G_3(\omega) \frac{e^{-\theta_1^2/2 - \theta_2^2/2}}{2\pi\sigma^2/\tau_1} \right)}{[H(\theta_1) + N_1 G_1(\omega)][H(\theta_2) + N_2 G_2(\omega)] - N_1 N_2 [G_3(\omega)]^2} d\omega \right), \quad (40)$$

$$E_{\text{ON-OFF}} = \langle s^2(t) \rangle \left(1 - \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{N_{\text{max}} |Q(\omega)|^2 \frac{e^{-\theta^2}}{2\pi\sigma^2/\tau_1}}{H(\theta) + N_{\text{max}}(G_1(\omega) - G_3(\omega))} d\omega \right), \quad (34)$$

where H is defined as in Equation 19, the function $Q(\omega)$ is the Fourier transform of the cross-correlation between the stimulus, $s(t)$, and the filtered stimulus, $g(t) = (f(t)s(t))$,

$$Q(\omega) = \frac{1}{1 + \tau^2 \omega^2} \left(\frac{1}{(1 + i\omega\tau_1)^4} - \rho \frac{1}{(1 + i\omega\tau_2)^4} \right), \quad (35)$$

and $G_1(\omega)$ and $G_3(\omega)$ are the Fourier transforms of

$$g_1(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} H\left(\frac{\theta - \sqrt{|\kappa(t)|} z}{\sqrt{1 - |\kappa(t)|}}\right) H\left(\frac{\theta - \sqrt{|\kappa(t)|} \text{sgn}(\kappa(t)) z}{\sqrt{1 - |\kappa(t)|}}\right) dz, \quad (36)$$

$$g_3(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} H\left(\frac{\theta - \sqrt{|\kappa(t)|} z}{\sqrt{1 - |\kappa(t)|}}\right) H\left(\frac{\theta + \sqrt{|\kappa(t)|} \text{sgn}(\kappa(t)) z}{\sqrt{1 - |\kappa(t)|}}\right) dz, \quad (37)$$

where $\kappa(t) = c(\tau)/\sigma^2$ is the normalized autocorrelation function of the filtered stimulus $g(t)$, with $\sigma^2 = c(0)$. The power spectrum of the filtered stimulus $g(t)$ is given by (note that this is a real number)

$$C(\omega) = \frac{1}{1 + \tau^2 \omega^2} \left| \frac{1}{(1 + i\tau_1 \omega)^4} - \rho \frac{1}{(1 + i\tau_2 \omega)^4} \right|^2. \quad (38)$$

For each N_{max} , we determined the thresholds for the ON and OFF cells at which the minimum MSE was achieved (see Fig. 10).

Similarly, we can write the expressions for the two filters and the minimum MSE for the system of two ON cells:

$$F_{\text{ON1}}(\omega) = \frac{Q(\omega) \left([H(\theta_2) + N_2 G_2(\omega)] \frac{e^{-\theta_1^2/2}}{\sqrt{2\pi\sigma^2}} - N_2 G_3(\omega) \frac{e^{-\theta_2^2/2}}{\sqrt{2\pi\sigma^2}} \right)}{[H(\theta_1) + N_1 G_1(\omega)][H(\theta_2) + N_2 G_2(\omega)] - N_1 N_2 [G_3(\omega)]^2} \quad (39)$$

with a similar expression for F_{ON2} by swapping the subscripts 1 and 2, and the optimal MSE is

where $G_1(\omega)$ is the Fourier transform of Equation 36 with θ_1 instead of θ , $G_2(\omega)$ is the Fourier transform of Equation 36 with θ_2 instead of θ , and $G_3(\omega)$ is the Fourier transform of:

$$g_3(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} H\left(\frac{\theta_1 - \sqrt{|\kappa(t)|} z}{\sqrt{1 - |\kappa(t)|}}\right) H\left(\frac{\theta_2 - \sqrt{|\kappa(t)|} \operatorname{sgn}(\kappa(t)) z}{\sqrt{1 - |\kappa(t)|}}\right) dz. \quad (41)$$

Results

Our aim is to inspect the phenomenon of sensory pathway splitting by the criteria of efficient coding. The efficient coding hypothesis holds that sensory systems are designed to optimally transmit information about the natural world given limitations of their biophysical components and constraints on energy use (Attneave, 1954; Barlow, 1961). This theory has been remarkably successful in the retina, where it predicts correctly a generic shape for ganglion cell receptive fields in both space and time (Atick and Redlich, 1990, 1992; van Hateren, 1992a,b). However, the retinal ganglion cells are not a homogeneous population, but split into ~ 20 different types (Wässle, 2004; Masland, 2012) with diverse response properties. In many cases, these responses are highly nonlinear and thus cannot be characterized entirely by a receptive field (Gollisch and Meister, 2010; Pitkow and Meister, 2014). Thus, an understanding of what drives the early pathway split will likely require a treatment that takes the nonlinearities into account.

The simplest and most prominent pathway split—in the retina and other sensory systems—involves ON and OFF response polarities: some neurons are excited by an increase in the stimulus and others by a decrease (Kuffler, 1953; Bennett, 1971; Wässle, 2004; Chalasani et al., 2007; Joesch et al., 2010; Gallio et al., 2011). Therefore, we begin by considering a simple sensory system of two neurons, with ON-type and OFF-type polarity encoding a common stimulus variable. We will compare the efficiency of this code to a system of two neurons with the same response polarity, say both ON-type. This way, one can measure the specific benefits of nonlinear pathway splitting. Of course, the specifics of this comparison will depend on detailed assumptions of the model. We will evaluate different versions of the stimulus–response function, starting with the simplest nonlinear form and progressing to more realistic versions. Efficient coding theory predicts that the optimal neuronal response properties depend on the natural stimulus statistics and the constraints on the system. Therefore, we consider different stimulus distributions and different constraints on the neural system. Finally, we will measure coding efficiency by three different criteria: (1) the mutual information between the response and stimulus; (2) the MSE achieved by a linear downstream decoder; and (3) the mean spike count needed to achieve a given performance (information or decoding error). We focus on results that emerge as robust under many different conditions.

Information transmission in a binary response model with Poisson output noise

Throughout this study, we assume that a neuron transmits information about a stimulus via the spike count observed over short time windows. The duration, T , of this coding window is chosen based on the observed dynamics of responses of retinal ganglion cells, typically 10–50 ms (Uzzell and Chichilnisky, 2004; Pitkow and Meister, 2012).

To build intuition for the benefits of ON–OFF splitting, we first studied a binary model of the response, which assumes that the firing rate of a neuron can only take two values: 0 and maximal, ν_{\max} (Fig. 1). An ON cell is silent when the stimulus lies below a threshold θ_1 and fires at the maximal rate above threshold. An OFF cell fires when the stimulus is below its threshold θ_2 and remains silent above. For a given firing rate and time window, the spike counts follow Poisson statistics. In this model system, one can derive analytic solutions for the coding efficiency (see Materials and Methods). Although this set of assumptions may seem grossly simplified, the binary response function is provably optimal under certain constraints (Stein, 1967; Shammai, 1990; Bethge et al., 2003; Nikitin et al., 2009) and even offers a reasonable approximation of retinal ganglion cell behavior (Pitkow and Meister, 2012). In later sections, we will relax these assumptions.

We begin by considering the mutual information transmitted by the responses about the stimulus. This information depends on the maximum expected spike count $N_{\max} = \nu_{\max} T$. Biophysically, such a constraint on the maximal firing rate arises naturally from refractoriness of the spike-generating membrane. Given a limit on the maximal firing rate, one can derive the optimal placement of the two thresholds θ_1 and θ_2 and compute the resulting information. This efficiency measure is independent of the stimulus distribution as long as its cumulative distribution is continuous (see Materials and Methods and Fig. 2). As expected, information increases monotonically as one raises N_{\max} (Fig. 3A). Interestingly, the ON–OFF and ON–ON systems encode exactly the same amount of information (Fig. 3A) over the entire range of N_{\max} .

The optimal thresholds can be measured in terms of the fraction of stimuli below threshold, ranging from 0 to 1 (Fig. 2). For the ON–OFF system, the optimal thresholds depend only weakly on N_{\max} (Fig. 3B). They are mirror symmetric about the stimulus median ($\theta_2 = 1 - \theta_1$), such that each of the two neurons reports the same fraction of stimulus values. Moreover, the threshold of the OFF cell is always below the threshold of the ON cell, resulting in an optimal threshold placement where the central stimulus region is coded by the silence of both cells. In contrast, although the thresholds of the optimal ON–ON system of the two ON cells begin identical for low N_{\max} , as N_{\max} increases, the separation between the two thresholds also increases.

To understand these behaviors, it helps to look more closely at how the stimulus is encoded. The two thresholds divide the stimulus range into three regions (Fig. 3B). When the ON cell fires, the stimulus lies above the ON threshold with certainty; likewise, when the OFF cell fires, the stimulus lies below the OFF threshold. When neither cell fires, the stimulus may lie in the center region or not, because zero spike counts can happen occasionally even at the maximal firing rate. This ambiguity gets smaller the larger the maximal firing rate, and in the limit of very large firing rates, the stimulus can be assigned uniquely to each of the three regions. Thus, in the low noise limit (large N_{\max}), the system maximizes information by maximizing the output entropy, i.e., by making those three regions equal in size, which explains the limiting values of the thresholds at $1/3$ and $2/3$ for large N_{\max} (Fig. 3B), resulting in $\log_2(3)$ bits of information (Fig. 3A). A similar argument holds for the ON–ON system, leading to the same threshold choices at high N_{\max} . However, at low N_{\max} , the two systems act differently: whereas the ON–OFF system responds to both high and low stimulus values, the ON–ON system responds only to high stimulus values but uses both neurons to do so, thus increasing the SNR. Thus, the representational benefit of the

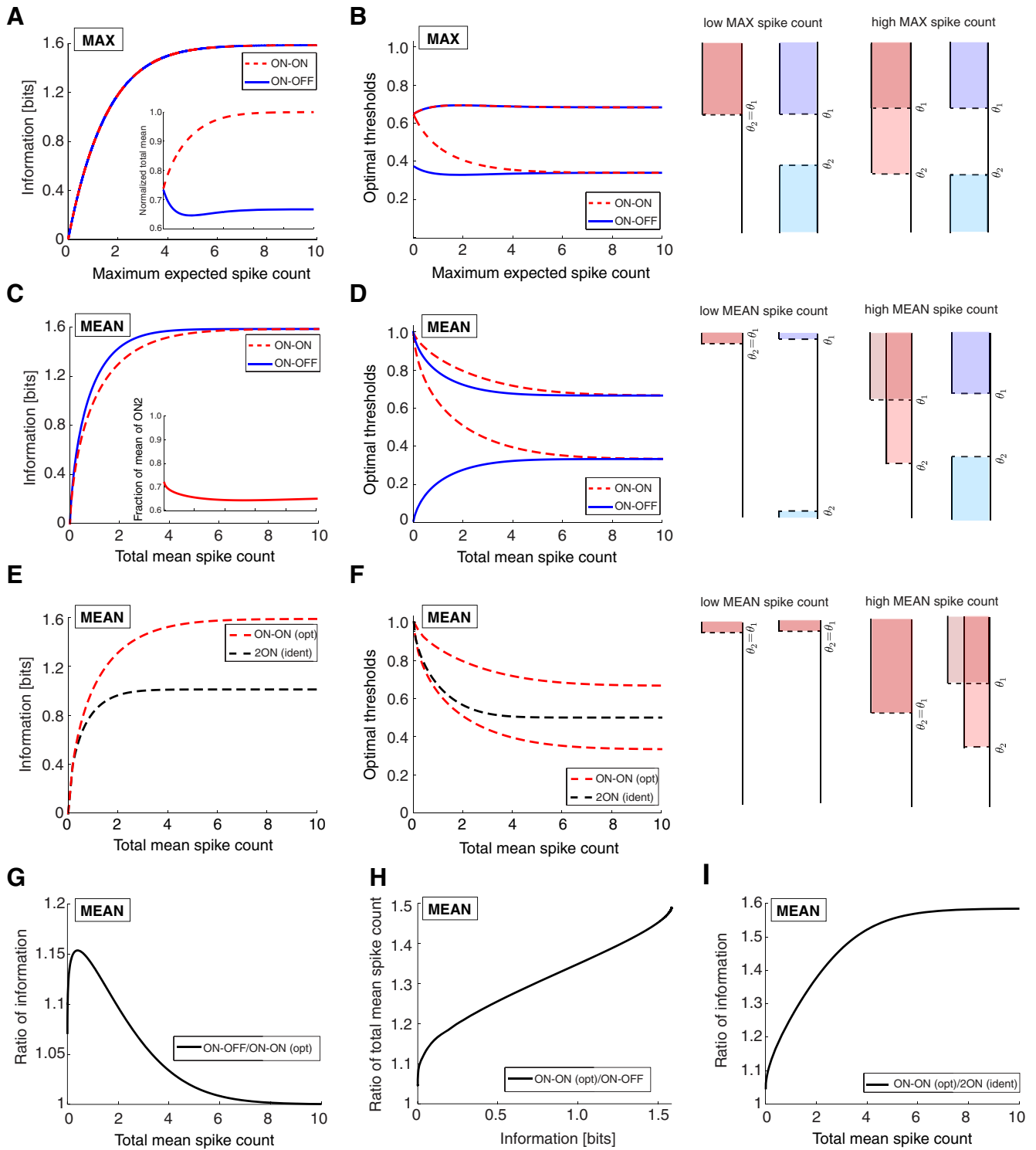


Figure 3. Mutual information with binary neurons and Poisson output noise. **A**, The mutual information for an ON–OFF and an ON–ON system is identical when constraining the maximum expected spike count, N_{\max} . Inset, The ON–ON system uses more spikes to convey the same information. The total mean is normalized by N_{\max} . **B**, The thresholds of the response functions for the optimal ON–OFF and ON–ON systems when constraining N_{\max} . Right, Optimal response functions in the limits of low and high N_{\max} . **C**, The ON–OFF system transmits more information than the ON–ON system when the total mean spike count is constrained. Inset, the fraction of total mean spike count assigned to the ON cell with the smaller threshold. **D**, The thresholds of the response functions for the optimal ON–OFF and ON–ON systems when constraining the total mean spike count. Right, Optimal response functions in the limits of low and high mean spike count (compare with **B**). **E**, The optimal ON–ON system transmits more information than a system of identical ON cells. **F**, The thresholds of the response functions for the optimal ON–ON and identical 2ON systems as a function of the total mean spike count. **G**, The ratio of mutual information transmitted by the optimal ON–OFF versus ON–ON systems in **C**. **H**, The ratio of mean spike counts required to convey the same information by the ON–ON versus ON–OFF systems in **B**. **I**, The ratio of mutual information from the optimal ON–ON versus identical 2ON systems in **E**.

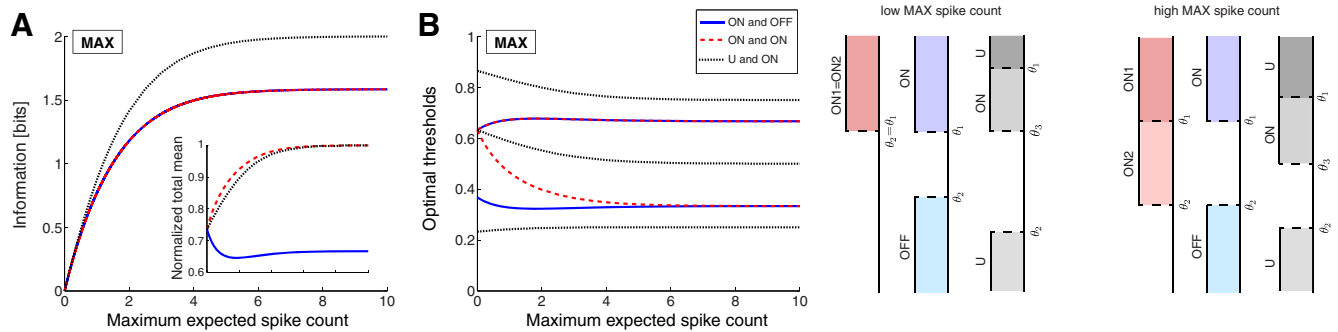


Figure 4. Mutual information with a non-monotonic response function. **A**, The mutual information for three different systems: (1) one ON and one OFF cell; (2) two ON cells; and (3) one U-cell with a U-shaped nonlinearity and an ON cell. Although the first two systems divide the stimulus range into three regions, the last divides it into a total of four regions and transmits a maximum of $\log_2(4) = 2$ bits of information. Inset, The total mean spike count (normalized by N_{\max}). **B**, The optimal thresholds of the response functions for the three systems in **A** when constraining N_{\max} . **Right**, Optimal response functions in the limits of low and high N_{\max} .

ON–OFF system is exactly compensated in the ON–ON system by its increased reliability, leading to the same information. Remarkably, the information achieved by the two systems remains identical throughout the range of N_{\max} . We have found no intuitive explanation for this analytical result.

Despite the equality in the information conveyed by the two systems, their efficiency differs. The ON–ON system requires up to 50% more spikes on average than the ON–OFF system (Fig. 3A, inset). Therefore, the ON–OFF system conveys higher information per spike than the ON–ON system. Next, we asked how the two systems behave if one constrains the total firing rate averaged over all stimuli rather than the maximal firing rate. Biophysically, such a constraint may arise from limits on the overall energy use for spiking neurons. Now each neuron can choose to use a high maximal firing rate for a small stimulus region or a low maximal firing rate over a larger region (Fig. 3C, inset). Also, the firing may be distributed unequally between the two neurons. Under this constraint on the mean spike count and allowing the two ON cells to adjust their maximal firing rates, we found that the ON–OFF system transmits more information than the ON–ON system (Fig. 3C). The largest performance ratio (ON–OFF/ON–ON) was 1.15, achieved at a total mean spike count of 0.4 (Fig. 3G). Interestingly, in both the low and high mean firing rate limits, the information ratio approaches 1. The ability of the ON–ON system to achieve similar information per spike as the ON–OFF system in these limits is attributable to the freedom to optimize not only the thresholds but also the maximal firing rates of the two ON cells.

Alternatively, one may ask how many more spikes the ON–ON system needs to match the information transmission of the ON–OFF system: this excess varies from ~ 0 to 50% (Fig. 3H). Again, we can understand the optimal threshold choices: at very low N_{mean} , the constraint on the mean spike count forces both systems to signal only an infinitesimally small stimulus region, to maximize the SNR. The ON–OFF system chooses small regions on either end of the range, whereas the ON–ON system uses both neurons redundantly for the same region. In the limit of large N_{mean} , the optimal thresholds again divide the stimulus range into equal thirds and maximize output entropy (Fig. 3D). For the entire range of N_{mean} , the ON–ON system assigned $\sim 2/3$ of the spikes to the ON cell with a lower threshold (Fig. 3C, inset).

Under both constraints analyzed so far, N_{\max} or N_{mean} , the optimal system of two ON cells consists of two neurons that differ in their thresholds or firing rates. Because this is another instance of nonlinear pathway splitting, we inspected the benefits of the optimal ON–ON system by comparing it to a suboptimal

ON–ON system constrained to use two identical response functions. The optimal system of ON cells transmits higher information when the total mean spike count N_{mean} is constrained (Fig. 3E, I; a similar result was obtained while constraining N_{\max} ; data not shown). The largest difference appears in the limit of high spike counts, where the suboptimal ON–ON scheme with identical response functions can distinguish only two equal stimulus regions [$\log_2(2) = 1$ bit] compared with three regions for the optimal ON–ON system [$\log_2(3)$ bits; Fig. 3F].

Although ON–OFF splitting yields a coding benefit among neurons encoding the same stimulus variable, the results of this section demonstrate that the details of the benefits of ON–OFF splitting depend on the assumed constraints. Two cells of opposite polarity have the same information capacity as two cells of the same polarity that use different response functions, when no additional cost than a constraint on their maximal firing rates is applied. What makes the ON–OFF system superior over ON–ON is the cost in terms of the mean firing rate; the ON–OFF system transmits higher information per spike than the ON–ON system when the maximal firing rate is constrained. However, when the total mean spike cost is constrained while allowing the two ON cells to adjust their maximal firing rates, the two systems achieve the same information per spike in the regime of high and low spike counts. The largest information benefit for the ON–OFF system of $\sim 15\%$ emerges in the intermediate firing rate regime, which is most biologically relevant.

Information transmission with non-monotonic nonlinearities

We also considered non-monotonic response functions, for example, a U-shaped nonlinearity, where a cell fires for both high and low stimulus values but not in between. We will call this a U-cell. An additional ON cell with monotonic nonlinearity can resolve the ambiguity between high and low stimuli.

A system with a U-cell and an ON cell encodes even higher information than the ON–OFF and ON–ON systems when the maximal spike count is constrained (Fig. 4A; the largest ratio is 1.26). This is because the U–ON system allows the stimulus range to be split into a total of four regions (as opposed to three in Fig. 3). Therefore, the maximum information transmitted by this system is $\log_2(4) = 2$ bits. From the optimal thresholds derived for this system (Fig. 4B), we compute the resulting total mean firing rate at which this higher information is achieved (Fig. 4A, inset). The total mean of the U–ON system is larger than the mean used by the ON–OFF system; therefore, the information per spike still remains the highest for the ON–OFF system. Similarly, a pair of

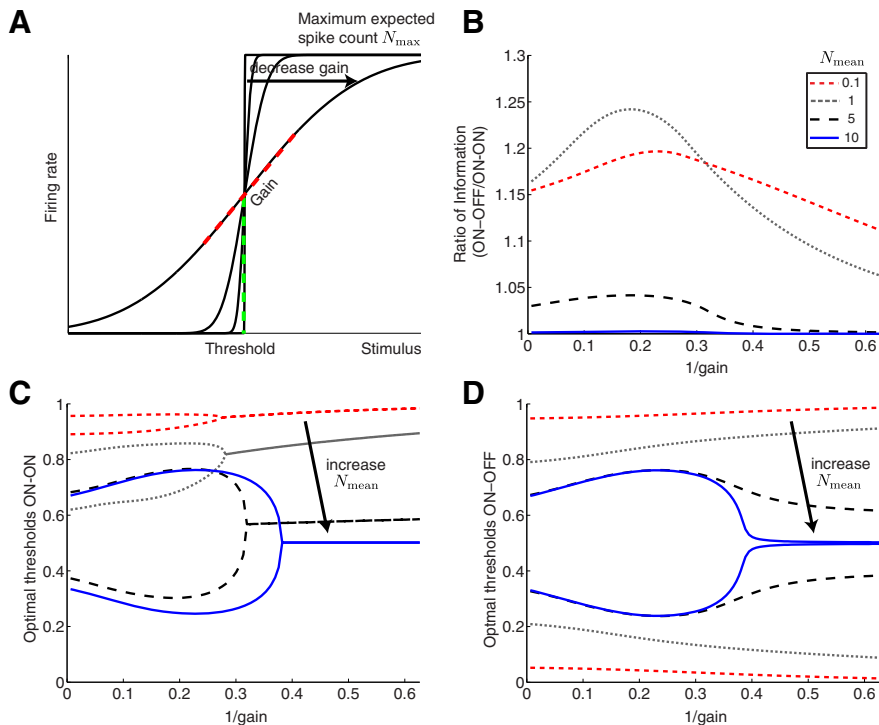


Figure 5. Mutual information with sigmoidal response functions and Poisson output noise. **A**, The sigmoidal response model (see Materials and Methods) is illustrated for an ON cell showing the maximum expected spike count N_{\max} , the gain (red dashed), and the threshold (green dashed) of the cell. The schematic shows the effect of varying the gain for a fixed threshold. **B**, The ratio of mutual information transmitted by the optimal ON–OFF versus ON–ON systems about a stimulus drawn from a Laplace distribution. The total mean spike count of each system, N_{mean} , was constrained. **C**, The optimal thresholds of the sigmoidal response functions for the ON–ON system when constraining the total mean spike count. Below a critical value of the gain, the two thresholds are identical. **D**, The optimal thresholds of the sigmoidal response functions for the ON–OFF system when constraining the total mean spike count. The two thresholds are different except in the limit of small gain and large mean spike count.

U-cells produces identical results to using one U-cell and one ON cell (data not shown). Therefore, it appears that two pathways of opposite polarity, one ON and one OFF, provide a more cost-effective way of coding a single stimulus variable, and for the remainder of the report, we proceed using ON or OFF cells rather than U-cells.

Information transmission with sigmoidal response nonlinearities

The stimulus–response functions of many neurons are not as sharp as a binary step (Chichilnisky and Kalmar, 2002; Uzzell and Chichilnisky, 2004). Therefore, we next considered sigmoidal response nonlinearities with limited gain (see Materials and Methods; Fig. 5A).

Information transmission by the ON–OFF and ON–ON systems behaved similarly to the case of step nonlinearities. For instance, when constraining the maximum expected spike count, the information was very similar (although not identical; the largest observed difference was 3%) for the two systems, whereas the total mean spike count was higher for the ON–ON system (data not shown). Similarly, when constraining the total mean spike count and optimizing the maximal firing rates of the cells, the ON–OFF system achieved higher mutual information about the stimulus than the ON–ON system (Fig. 5B).

A qualitative difference from the case of step nonlinearities (infinite gain) occurs when one examines the optimal thresholds for the two systems. In the ON–ON system, the two cells adopt different thresholds at high gain values (as in the infinite gain binary model; Fig. 3D), but at low gain, their thresholds collapse

to the same value (Fig. 5C; Kastner et al., 2014). The critical gain value at which this occurs depends on the mean spike count (Fig. 5C). For the ON–OFF system, the two thresholds were different for all gain values but approached each other in the limit of low gain and high mean spike count (Fig. 5D). Just as observed with step nonlinearities, the optimal thresholds for the ON and OFF cells were mirror symmetric about the stimulus median. This is expected based on the symmetries of the problem.

We also considered the effect of an asymmetric distribution of stimuli. For example, the distribution of spatial contrasts in natural images has a heavier tail below the median than above (Tadmor and Tolhurst, 2000; Ratliff et al., 2010). Under such a distribution, the ON and OFF cells adopt asymmetric response functions: the threshold of the OFF cell moves away from the stimulus median and that of the ON cell moves closer (data not shown). Such an asymmetry has indeed been observed for the nonlinearities of ON and OFF macaque ganglion cells (Chichilnisky and Kalmar, 2002). Higher maximal firing rates of the ON cell (Uzzell and Chichilnisky, 2004) or a steeper gain of the OFF cell (Chichilnisky and Kalmar, 2002) can also produce asymmetries in the response functions (data not shown).

Realistic encoding and noise models from retinal data

Given the intuition acquired from studying the effects of variable gain on the response functions, we next explored two-neuron systems with realistic choices for the response gain, firing rates, noise, and stimulus distributions. For the stimulus–response function, we used a sigmoidal nonlinearity derived from measured responses of primate (Chichilnisky and Kalmar, 2002; Uzzell and Chichilnisky, 2004) and salamander (Pitkow and Meister, 2012) retinal ganglion cells. Because the spiking responses of ganglion cells are more regular than expected from Poisson statistics (Berry et al., 1997; Uzzell and Chichilnisky, 2004), we used a noise distribution that was measured empirically (Pitkow and Meister, 2012). With this encoding model, the shape of the stimulus distribution matters for the efficiency of information transmission. In choosing a stimulus distribution, we followed previous work on image processing of natural scenes (Field, 1994; Bell and Sejnowski, 1997), which suggests that the signal encoded by a typical retinal ganglion cell follows a Laplace distribution. Similar distributions with positive kurtosis also arise from temporal filtering of optic flow movies during human locomotion and simulated fixational eye movements on static images (Ratliff et al., 2010). However, the main conclusions we present here hold also for a Gaussian distribution.

Three parameters characterize the sigmoidal nonlinearity (Eqs. 10 and 11): (1) the maximum expected spike count N_{\max} ; (2) the gain β ; and (3) the threshold θ (Fig. 6A). These parameters were set to measured values from salamander (Pitkow and Meister, 2012) and macaque (Chichilnisky and Kalmar, 2002) ganglion cells (see Materials and Methods). Under these conditions,

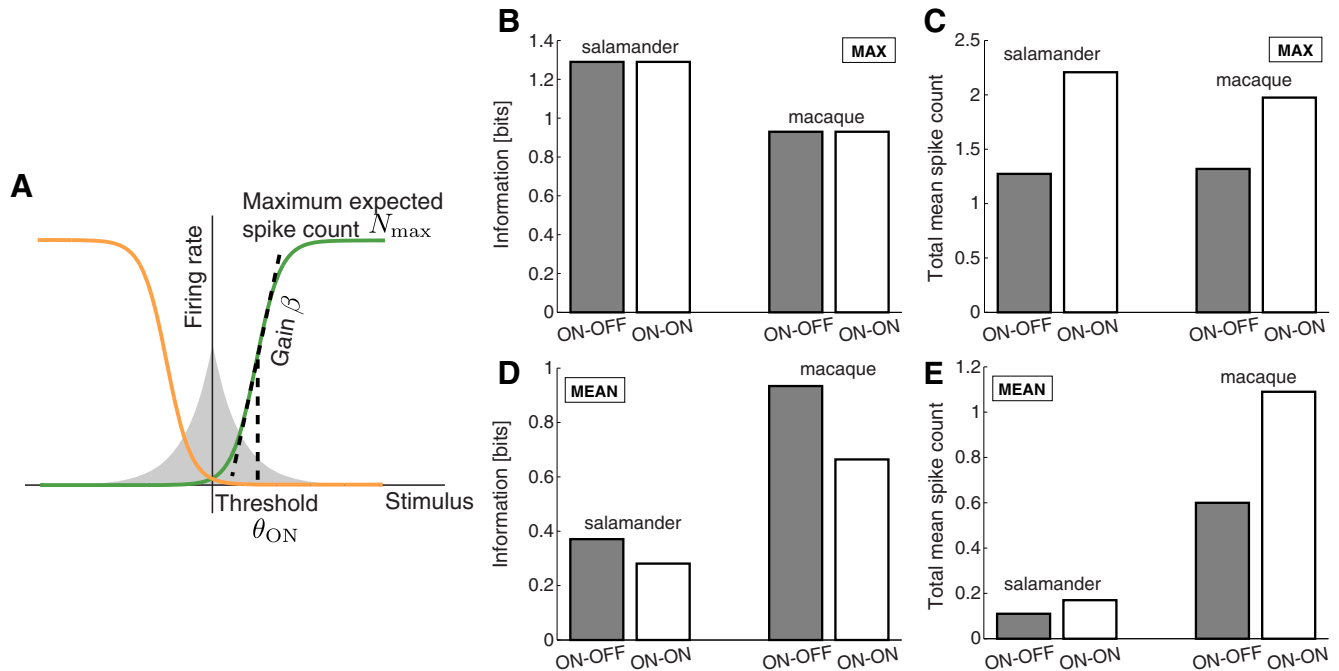


Figure 6. Mutual information with response functions and noise from retinal data. **A**, Two sigmoidal nonlinearities for an ON cell (green) and an OFF cell (orange), describing the firing rate as a function of stimulus with the maximum expected spike count N_{\max} , the gain β , and the threshold θ . The shaded curve denotes the Laplace stimulus probability distribution. **B**, The information transmitted by the ON–OFF and ON–ON systems, using nonlinearities measured from salamander (left) and macaque (right) ganglion cells (Pitkow and Meister, 2012). The maximum spike count was constrained. **C**, The total mean spike count used by the ON–OFF relative to the ON–ON system to transmit the maximal information for the two types of nonlinearities in **B**. **D**, The information transmitted by the two systems using a constraint on the total mean spike count. **E**, The total mean spike count that would be needed by the ON–OFF and the ON–ON systems to transmit the same information as the ON–OFF system shown in **D**.

the ON–OFF and ON–ON systems transmitted almost identical information (Fig. 6B), just as observed with the simple binary response model (Fig. 3A). Again, the ON–ON system required considerably more spikes on average (Fig. 6C; ratio of total mean spike count for ON–ON/ON–OFF, 1.73 for salamander and 1.50 for macaque). However, the mean firing rates produced by this model greatly exceed those observed in real ganglion cells: for example, the model salamander ganglion cells fire at average rates of 12–32 Hz compared with 1.1 Hz for the average real salamander ganglion cells (Pitkow and Meister, 2012). This discrepancy arises because the optimal model thresholds are closer to the stimulus median (ranging from 0.3 to 0.5 in units of the stimulus SD) than the experimentally measured value of 2. For the macaque parameters, the mean firing rates and thresholds follow similar patterns: model neurons have mean firing rates in the range of 65–132 Hz compared with 30 Hz for real neurons (Uzzell and Chichilnisky, 2004). Again, this discrepancy results from mismatched thresholds (0.2–0.5 in the model vs 1.3 in experiment, again in units of the stimulus SD).

This mismatch disappears if one constrains the total mean firing rate instead of the maximum rate. After setting the mean firing rate to the values observed experimentally, the ON–OFF system outperformed the ON–ON system for both salamander and macaque parameters (Fig. 6D; information ratio ON–OFF/ON–ON, 1.3 for salamander and 1.4 for macaque), just as observed under the simpler response model in the intermediate signal-to-noise regime (Fig. 3G,H). Conversely, the ON–ON system requires a considerably greater firing rate to match the performance of the ON–OFF system: almost twice the rate measured experimentally in the case of macaque data (Fig. 6E). In this model—with a constraint on the mean firing rate—the optimal thresholds of the response functions are much closer to the ex-

perimentally measured values (2.0 for both analysis and experiment for salamander and 1.1–1.9 for analysis vs 1.3 for experiment for macaque, in units of the stimulus SD).

Along with the thresholds, the optimization procedure also adjusted the maximal firing rates, and their values are again much closer to the experimental values than the mean firing rates obtained previously under a constraint on the maximum rate (salamander, 12–38 Hz in the model and 48 Hz in the data; macaque, 200–580 Hz in the model and 220 Hz in the data). Therefore, although we did not constrain the maximal firing rates in this scenario, the optimal solution converged on reasonable values. Note that these maximal rates were rarely encountered: because of the high firing thresholds, the model retinal ganglion cells spent only 3–7% of the time in the upper half of the sigmoid nonlinearity.

In summary, this section shows that the main results observed with a binary response model and Poisson firing hold also under more realistic assumptions about the encoding function and output noise. For the remainder of the analysis, we will continue using binary nonlinearities and Poisson output noise, because these choices are conducive to analytical solutions and thus allow us to explore in greater detail other factors that might influence information transmission. We also continue constraining the mean firing rate, because our results show that it produces more realistic predictions for retinal nonlinearities than a constraint on the maximal firing rate.

The effect of common input noise

Until now, we made the assumption that variability in the spiking output of each individual ganglion cell is the dominant source of noise, consistent with the small shared variability among these neurons observed in some studies (Pitkow and Meister, 2012). However, under different experimental conditions, nearby ganglion cells ex-

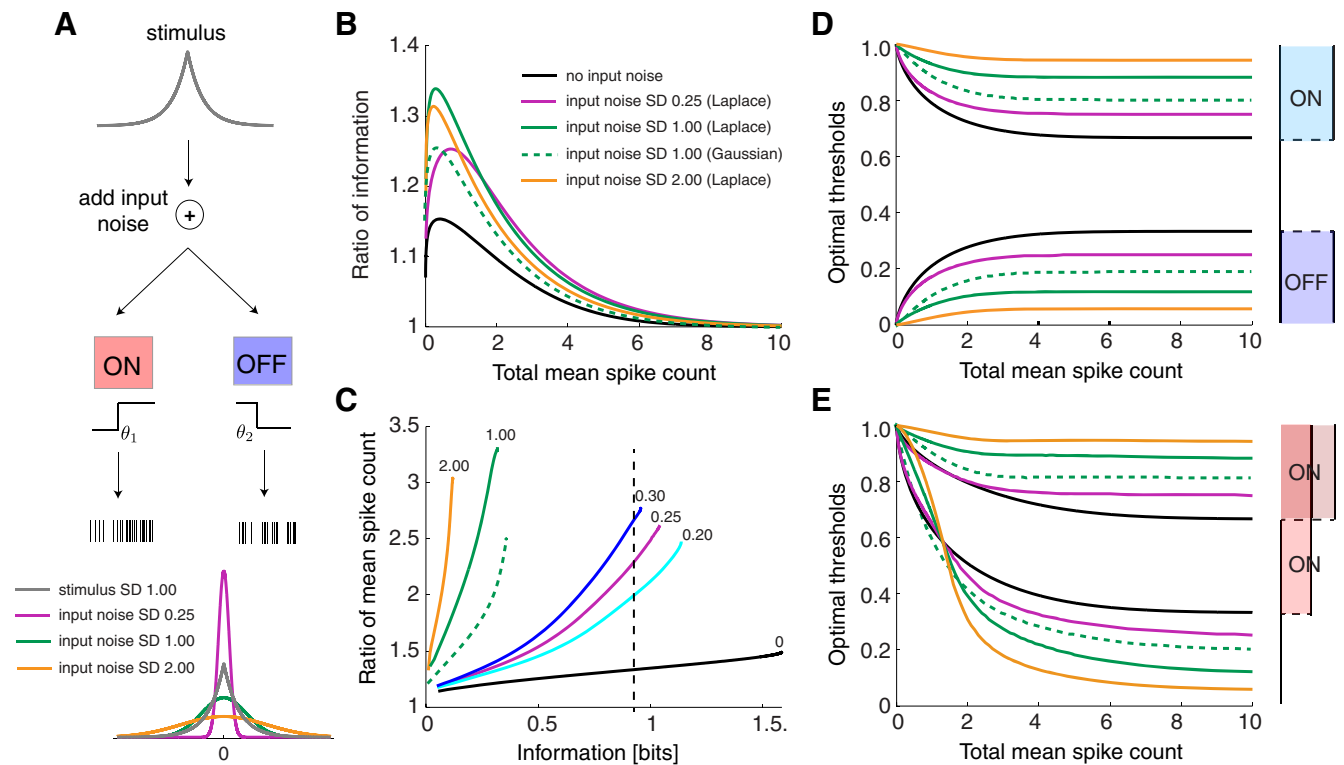


Figure 7. Input noise increases the benefits of ON–OFF splitting. **A**, Top, Input noise is added to the stimulus before the nonlinearities. Bottom, Distribution of stimulus (Laplace) and input noise (Gaussian with different SDs). **B**, Ratio of information transmitted by the ON–OFF versus ON–ON systems when constraining the total mean spike count. Input noise increases the benefit of the ON–OFF system but non-monotonically. In this and subsequent panels, the dashed green line demonstrates the results for a Gaussian stimulus distribution with input noise SD 1 (compare to the full green line for the Laplace stimulus distribution with SD 1). **C**, The ratio of total mean spike count needed to transmit a given information by the ON–ON versus the ON–OFF system, computed for different values of the noise SD. For instance, to transmit 0.9 bits (dashed black line), the ON–ON scheme needs twice more spikes at input noise SD 0.2 and 2.75 times more spikes at input noise SD 0.3. **D**, The optimal thresholds in the ON–OFF system get pushed toward the tails of the distribution as the input noise increases. **E**, The optimal thresholds in the ON–ON system are also distributed at the two tails of the distribution for large spike counts but cluster at one end of the stimulus distribution for low spike counts.

hibit strong noise correlations, presumably arising from noise in shared presynaptic neurons, as early as the photoreceptors (Ala-Laurila et al., 2011). To capture these cases of neural encoding, we extended the existing model by adding a source of common additive noise at the input to the two neurons; the combined input (stimulus plus noise) is passed through the ON and OFF nonlinearities and converted into spike counts by a Poisson process (Fig. 7A). The input noise was drawn from a Gaussian distribution. Again, we computed the information about the stimulus transmitted by ON–OFF and ON–ON systems with binary response functions and optimally chosen thresholds, while holding a constraint on the total mean firing rate but allowing the cells to adjust their maximum firing rates. Specifically, we examined the effect on information transmission by varying the level of input noise and the sparseness of the stimulus distribution.

We first studied the effect of input noise using a Laplace stimulus distribution. As expected, addition of input noise degrades the information transmission in both systems and more so the stronger the noise (data not shown). In terms of information capacity, the ON–OFF system outperforms the ON–ON system most significantly for an intermediate constraint on the total mean spike count, and the addition of input noise further enhances those benefits (Fig. 7B). The relative advantage of the ON–OFF system is maximal when the noise power is approximately equal to the stimulus power. Under optimal conditions, the ON–OFF information rate is up to 34% higher than for the ON–ON system. Alternatively, if the goal is to transmit a fixed amount of information about the stimulus, the ON–OFF scheme achieves it with a much lower mean firing rate, by a factor of up to 3.4

(Fig. 7C). This far exceeds the maximal benefit without input noise (Fig. 3H). If the noise power exceeds stimulus power, these benefits decrease again (Fig. 7B, C).

One can trace the effects of input noise to the optimal threshold settings in the encoding model. As the input noise increases, the ON–OFF system places its thresholds closer to the two ends of the stimulus range, where the stimulus values are less confounded with noise (Fig. 7D). The ON–ON system does the same in the regime of high spike counts (Fig. 7E). This implies that one of the ON cells fires almost all the time, which produces the great difference in coding efficiency between the two models. At low mean spike count, the ON–ON system again opts for redundant coding of high stimulus values, as in the model without input noise (Fig. 3D).

In the presence of input noise, the coding efficiency and the resulting optimal thresholds do depend on the shape of the stimulus distribution, unlike the situation without input noise (Fig. 3). Besides the empirically founded Laplace distribution, we also examined a Gaussian distribution that has been popular in many neural coding models. The two distributions had matched mean and variance, but the Laplace distribution has higher kurtosis and thus greater stimulus sparseness. A higher stimulus sparseness increased the efficiency of the ON–OFF system relative to the ON–ON system, as measured by the information rates (Fig. 7B, dashed) and the ratio of mean spike counts required to achieve a given information (Fig. 7C, dashed). Similar to the effects of adding input noise, increasing stimulus sparseness also pushed the thresholds closer to the two ends of the stimulus range (Fig. 7D, E).

These results demonstrate that, in the presence of input noise, the ON–OFF system again outperforms the ON–ON system. In fact, the benefits of ON–OFF coding can be substantially greater than without input noise. Bigger improvements in ON–OFF performance are observed for sparser stimulus distributions where one of the ON cells of the ON–ON system fires for a larger fraction of stimuli near the stimulus median.

Linear decoding as a measure of coding efficiency

Thus far, we have measured the total information transmitted by model neurons, without regard to how it can be decoded. An alternative criterion for coding efficiency is the ease with which these messages can be decoded by downstream neurons. In particular, it has been suggested that the important variables represented by a population of neurons should be obtainable by linear decoding of their spike trains (Seung and Sompolinsky, 1993; Marder and Abbott, 1995; Eliasmith and Anderson, 2002; Gollisch and Meister, 2010). Hence, we now ask how well an optimal linear decoder can extract the stimulus value from the firing of the two model neurons. In each coding window, the decoder sums the two spike counts with appropriate decoding weights to estimate the stimulus value. We measure the quality of the decoding by the SNR, where signal is the power of the stimulus and noise is the power of the difference between the true stimulus and the estimate of the decoder (see Materials and Methods). For any model cell pair, the decoder is optimized by choosing the two decoding weights that maximize the SNR. As before, we compare an ON–OFF pair and an ON–ON pair, each with optimized thresholds.

Because we wanted to compare the coding benefit realized by two measures, mutual information and the SNR of the optimal linear decoder, we returned to the original noise model in which the sole source of noise was at the level of the spike output with Poisson statistics. Unlike the mutual information (Fig. 3), the results of linear decoding depend on the shape of the stimulus distribution (see Materials and Methods); hence, we examine the role of stimulus sparseness on the accuracy of stimulus decoding by studying Laplace and Gaussian stimulus distributions. As before (Fig. 3), we also explored the effects of two different constraints on encoding: first on the maximal firing rate of each neuron and then on the total mean firing rate while allowing the maximal firing rates of the cells to vary.

As expected, a higher maximal or mean firing rate constraint produced higher SNR after decoding (Fig. 8*A,B*). However, the

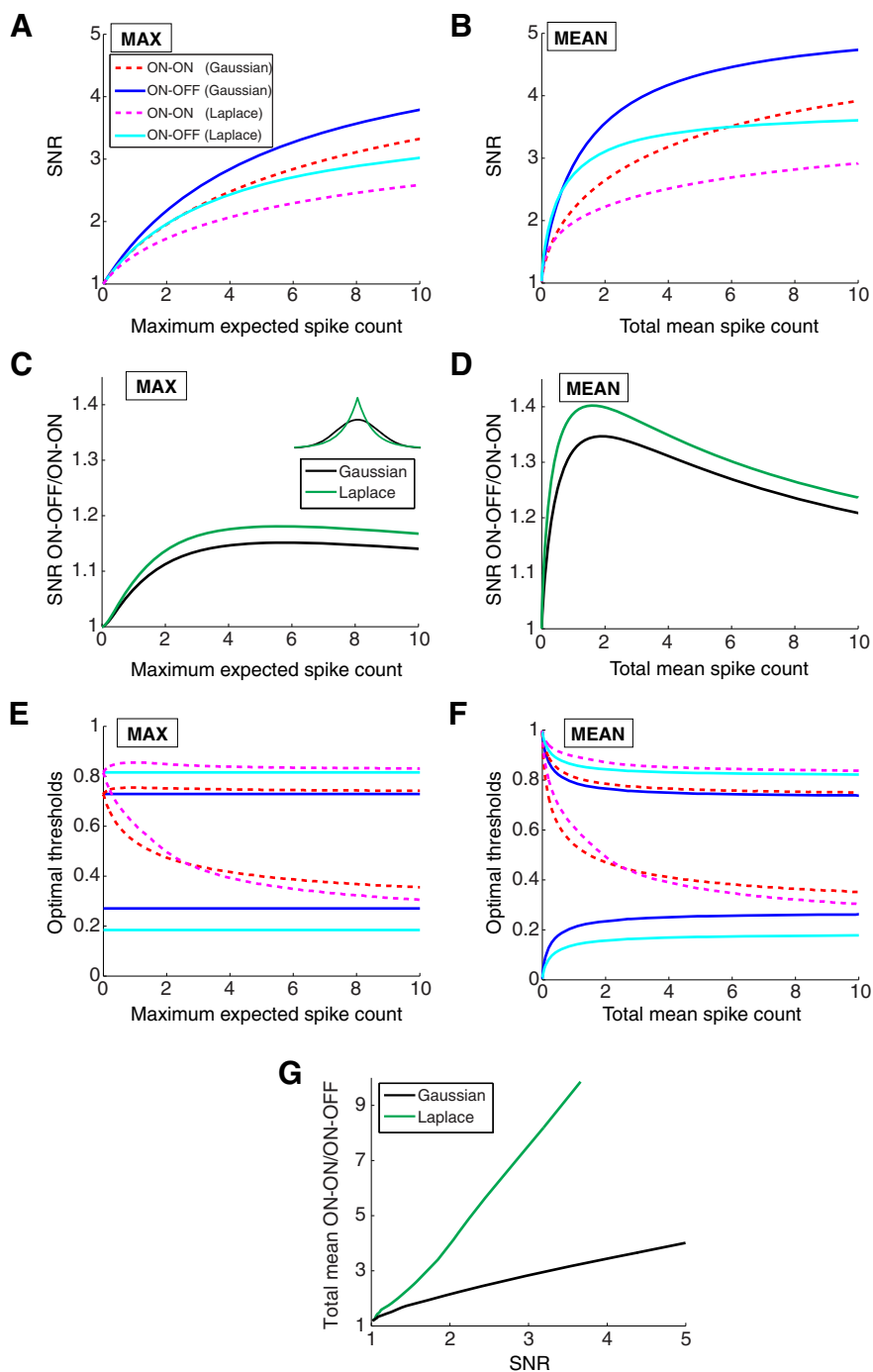


Figure 8. Linear decoding of stimuli from cell responses increases the benefits of ON–OFF splitting. *A, B*, The SNR of the stimulus reconstruction for two stimulus distributions (Gaussian and Laplace) and subject to two different constraints (maximum expected spike count and total mean spike count). *C, D*, The ratio of SNR curves shown in *A* and *B* under the two different constraints. *E, F*, The optimal thresholds (shown as the fraction of stimuli below threshold) for the response functions under the two different constraints (compare with Fig. 3*B, D*). *G*, The ratio of total mean spike count needed by the ON–ON versus the ON–OFF system to achieve the same SNR. Larger benefits for ON–OFF are observed for the Laplace distribution.

ON–OFF system achieved higher SNR than the ON–ON system, regardless of how the firing rates are constrained (Fig. 8*C,D*). With a constraint on the mean firing rate, the most pronounced benefit of ON–OFF coding was 40% (Fig. 8*D*). Alternatively, one can ask how much firing would be required to accomplish a certain quality of decoding, and again the ON–OFF system was much more efficient (Fig. 8*G*). For example, achieving the highest possible SNR requires 4-fold more spikes with the ON–ON sys-

tem than with the ON–OFF system if the stimulus distribution is Gaussian (Fig. 8G). With the more kurtotic Laplacian stimulus distribution, this increases to a 10-fold difference (Fig. 8G).

The optimal thresholds of the encoding neurons behave similarly to what was seen in previous instances of the model: at high spike count, the two neurons partition the stimulus range (Fig. 8E,F), although not quite into thirds as seen under the mutual information criterion (Fig. 3B,D). The improved performance of the ON–OFF system relative to ON–ON in the case of the Laplace stimulus distribution occurs with placing the optimal thresholds closer to the two ends of the stimulus range as in the case of the mutual information with input noise (Fig. 7D,E). At low spike counts, the ON–ON system again opts for redundant coding at the positive tail of the stimulus distribution.

In summary, this section shows that, if one adopts the alternative criterion of linear decodability, the ON–OFF system outperforms the ON–ON system by an even larger margin than seen before, both in terms of the SNR at a fixed maximum spike count, and the total mean spike count required to achieve a certain SNR.

Decoding a time-dependent stimulus

We next examined a more realistic encoding model in which the task is to reconstruct a stimulus with a temporal structure. Dong and Atick (1995) measured the spatiotemporal correlation structure of a large ensemble of natural movies and found that, in the regime of low spatial frequencies, the temporal power spectrum is well fit by the universal power law with exponent -2 . As a limiting case of this result, we used a Gaussian low-pass filtered stimulus with a correlation time longer than all other time constants in the system (see Materials and Methods). To mimic how the circuits of the retina encode such a stimulus, we first passed it through a linear filter with biphasic impulse response (Fig. 9), as seen in measurements from primate ganglion cells (Chichilnisky and Kalmar, 2002). This filtered stimulus was passed to the two neurons in the encoding model (with binary response functions) and spikes were derived as an inhomogeneous Poisson process. The task of the linear decoder is to reconstruct the time-dependent stimulus by summing the output of two temporal filters applied to the spike trains (Fig. 9). For a given set of thresholds, the decoding filters were derived analytically by maximizing the SNR of the decoder (see Materials and Methods).

This generalized encoding model led to qualitatively identical conclusions: the ON–OFF system outperformed the ON–ON system under all conditions (Fig. 10). Quantitatively, the benefits of the ON–OFF system were further enhanced over previous instantiations of the model. For example, the ON–OFF system produced a maximal SNR improvement of 24% compared with ON–ON at the same maximal firing rate of each cell (Fig. 10C) and 51% at the same total mean firing rate (Fig. 10D). At the same SNR, the ON–OFF system used 4.5 times fewer spikes (Fig. 10G). Interestingly, as in the case of the mutual information with input noise and the linear decoder of static stimuli, the more favorable performance of the ON–OFF system compared with ON–ON goes hand-in-hand with a shifting of the optimal thresholds away from the stimulus median toward the tails of the distribution (compare Figs. 8E,F, 10E,F).

The coding benefit of ON–OFF splitting is comparable with that from decorrelation

Under the assumptions considered so far, the ON–OFF system generally outperformed the ON–ON system by an average of 25% in information rate (15% in Fig. 3 and 34% in Fig. 7). How significant is a 25% gain for the purpose of efficient coding in sensory systems? We compare the present effects to another pop-

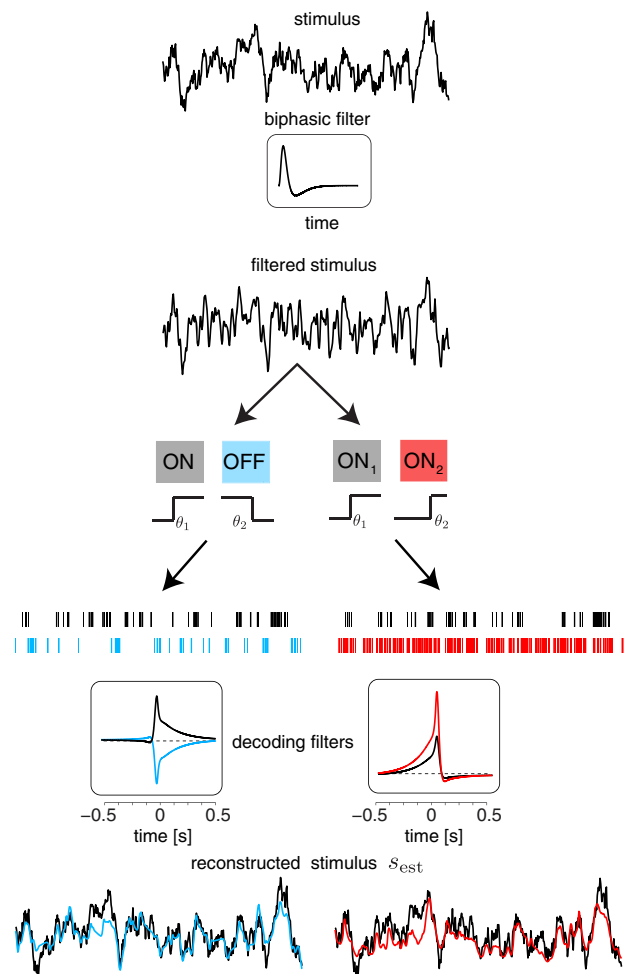


Figure 9. Schematic of linear decoding of time-varying stimuli from cell responses. A stimulus with temporal correlations was first passed through a biphasic filter and then encoded by an ON–OFF or ON–ON system. The resulting time series of spike counts were convolved with decoding filters to produce the reconstructed stimulus estimate. Examples of the decoding filters for each system (ON–OFF and ON–ON) and the resulting reconstructed stimuli are shown for $N_{\max} = 10$. Stimulus (black), stimulus estimate from ON–OFF (blue), and stimulus estimate from ON–ON (red).

ular application of efficient coding theory: decorrelation. According to the theory, early sensory circuits are designed to remove redundancies that arise in natural stimuli attributable to the strong correlations they induce within the population of sensory receptors, for example, among nearby pixels of a visual image (Ruderman, 1994). The resulting neural signals are more independent, thus allowing greater information transmission given the biophysical constraints on neural spike trains (Atick and Redlich, 1990, 1992; van Hateren, 1992b). How much greater?

Consider the simplest such system of just two neurons, each encoding a scalar stimulus variable using a step-shaped response function (Fig. 11; Pitkow and Meister, 2012). Suppose the two stimulus variables have positive correlation. To achieve optimal encoding, they first need to be decorrelated by lateral inhibition so that the inputs to the two neurons become statistically independent before passing through the nonlinear response function (Pitkow and Meister, 2012). We ask how much information is gained by the decorrelation step under the most favorable conditions. First, decorrelation is most effective in the regime of low noise, because the procedure tends to amplify the unshared noise

over the shared signal (Atick and Redlich, 1992). Thus, we will assume zero noise. Furthermore, the greatest gain accrues if the two signals are strongly correlated, say almost identical. Without lateral inhibition, the two neurons are forced to represent essentially the same stimulus variable (Fig. 11A). They can adjust their thresholds so as to encode three different regions of the stimulus (as in Fig. 3) and thus transmit up to $\log_2(3)$ bits of information (Fig. 11B). In contrast, if the two variables are first decorrelated and then each encoded with a binary neuron, the information is 2 bits (Fig. 11C). Thus, the benefit for a two-neuron system from implementing stimulus decorrelation is $[2 - \log_2(3)]/\log_2(3) = 26\%$ in the best case. We conclude that, for the goal of coding efficiency, ON–OFF pathway splitting is just as significant as decorrelation, a circuit function whose interpretation stands as a celebrated success of efficient coding theory (Atick and Redlich, 1990; Atick, 1992).

A sparse stimulus distribution emphasizes the benefits of ON–OFF coding

Whereas previous sections assumed that all stimulus values were in a sense equally important for transmission, we now suppose that the downstream brain areas are particularly concerned about large deviations in the stimulus of either sign. This is ecologically plausible in several situations. For example, consider a visual observer of a scene in which a potential prey animal moves against a static background (Dong and Atick, 1995). Almost every pixel in the image stays constant across time, except a few pixels at the leading and trailing edges of the moving object. These show a sudden increase or decrease in intensity depending on the relative brightness of target and background. Here all the ecological relevance lies in the rare stimulus values with large positive or negative changes. They are embedded in a sea of values with zero or very small change, which do not need to be discriminated downstream. Similar scenarios can be drawn for other senses, such as olfaction and electrosensation.

To implement this subjective criterion of ecological relevance, we can compress the stimulus distribution into just three discrete values: 0, -1 , and $+1$. Most of the time, the stimulus is 0 and only rarely $+1$ or -1 . We suppose that $+1$ and -1 occur with equal frequency $p_{-1} = p_1$. Thus, the stimulus distribution has only one parameter, and we will use the “sparseness,” $1/p_1$. Again, the stimulus drives two model neurons with binary rate functions and Poisson firing statistics, and we compute the mutual information between stimuli and responses, with a constraint on the

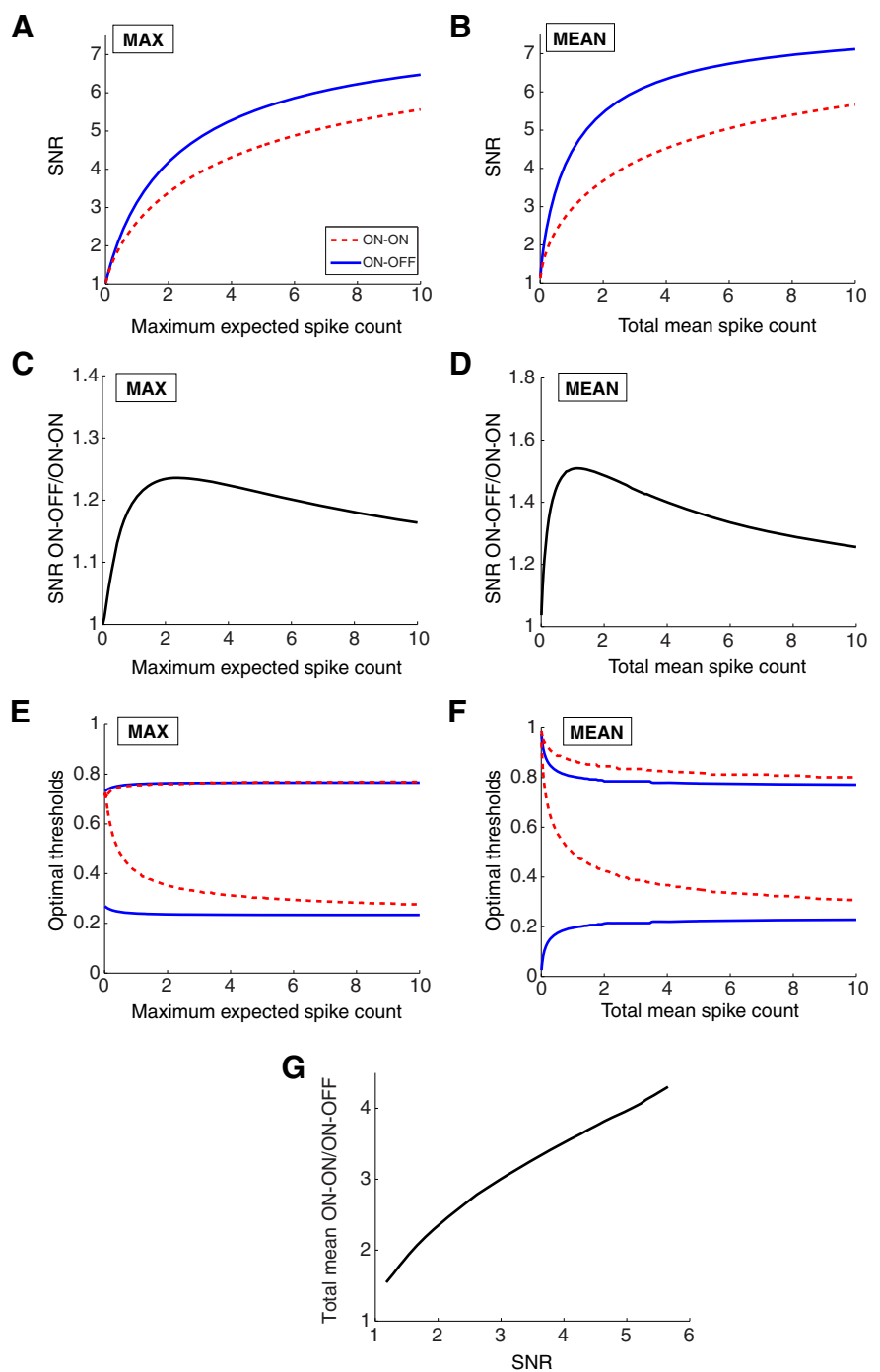


Figure 10. Decoding of temporally correlated stimuli benefits from ON–OFF splitting. **A, B**, The SNR from linear reconstruction of a dynamic stimulus while constraining the maximum expected spike count (**A**) or the total mean spike count (**B**). **C, D**, The SNR achieved by the ON–OFF versus the ON–ON system under the conditions of **A** and **B**. **E, F**, The optimal thresholds (shown as the fraction of stimuli below threshold) for the response functions of the ON–OFF and ON–ON systems of **A** and **B**. **G**, The ratio of total spike counts used by the ON–ON versus the ON–OFF system to achieve the same SNR.

total mean firing rate while we allow for the cells to adjust their maximal firing rates (Fig. 12A).

As expected, the information rate increases with the mean firing rate (Fig. 12B). There is an extended region of firing rates in which the ON–OFF system transmits twice as much information as ON–ON. This region increases with the sparseness of the stimulus. For example, for a sparseness of 100, the region extends over a ~ 100 -fold range of firing rates. As before, one can also phrase the comparison in terms of how many spikes are needed to

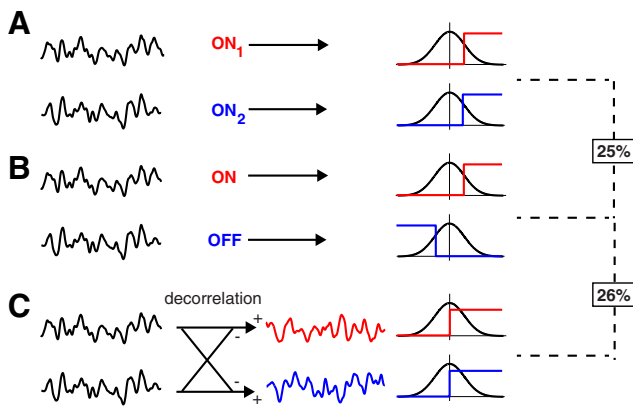


Figure 11. Comparing the effects of ON–OFF splitting and decorrelation. **A, B**, Two neurons encode almost identical stimulus variables. As shown in the previous figures, ON–OFF coding enhances the transmitted information by $\sim 25\%$. **C**, Here the two stimulus variables are first decorrelated by lateral inhibition to statistical independence and then encoded by two neurons. Under best conditions, this transmits 2 bits, a 26% increase over the $\log_2(2)$ bits in **B**.

achieve a certain information rate. If the target rate is above half-maximum, then the ON–OFF system requires far lower firing rate than the ON–ON system (Fig. 12C). The factors here are enormous, on the order of the sparseness of the stimulus distribution.

These large effects can be understood by considering the dilemma of the ON–ON system (Fig. 12A). It can signal the $s = -1$ events only by cessation of firing in one of the neurons, but that requires maintaining a constant firing rate during the much more frequent $s = 0$ events. If a low mean firing rate is specified, then that is simply not an option. Thus, the optimal ON–ON model forgoes signaling the $s = -1$ events entirely and therefore has to discard half the available stimulus information (Fig. 12B). Only when the available mean firing rate increases enough does the encoding through silence become an option. In contrast, the ON–OFF system can allocate spikes exclusively to the rare events $s = \pm 1$ and thus can encode both of them throughout the range of mean firing rates.

This section illustrates that the coding advantages of an ON–OFF coding scheme become truly dramatic when the important stimuli consist of sparse events on either side of the stimulus median.

Discussion

The goal of this study was to illuminate why sensory systems create multiple parallel pathways to encode the same signal. We focused on one aspect of pathway splitting: the use of ON-type and OFF-type neurons to signal the same stimulus variable, which is a prominent feature in several sensory modalities across the animal kingdom. Using a range of models for the encoding of a stimulus, we asked what the functional benefits are of using two neurons of opposite response polarity (ON–OFF) compared with two neurons of the same polarity (ON–ON). The general approach was to set a plausible constraint on the nature of the response models, allow each of the two systems to optimize the parameters of the response, and then evaluate their relative performance by criteria related to the efficiency of neural coding.

Our analysis shows that the benefit of the ON–OFF architecture is not as universal and straightforward as it would initially seem. First, we find that the two systems have identical capacity for information transmission when only the maximal firing rate is constrained (Fig. 3), despite the fact that they tile stimulus

space differently and possess different overall noise statistics (Fig. 2). However, the ON–OFF system is more efficient, because it conveys higher information per spike (Fig. 3). In the relevant regime for the retina, with an average of one or a few spikes per coding window (Uzzell and Chichilnisky, 2004; Pillow et al., 2008; Pitkow and Meister, 2012), the ON–OFF system offers an increase in information about the stimulus of up to 15% or a reduction in spikes required for the same information rate of up to 50% under the most conservative conditions of binary response models and Poisson output noise (Fig. 3). This benefit increases for more realistic versions informed by neural coding in the retina to 20–40% gain in information and 80% in average firing rate in Figures 5 and 6.

If the two neurons share substantial input noise on the common sensory variable, then the benefits of ON–OFF coding increase, to 34% enhancement of the maximal information rate, or a 3.4-fold reduction in spikes required for a given information rate (Fig. 7). In the presence of input noise, the thresholds are pushed to the tails of the stimulus distribution, resulting in the ON–ON system being active most of the time. The input noise forces the system to discriminate between noise and signal; this effect is accentuated when the stimulus distribution is sparse, making the tails most informative (Fig. 7). For certain stimulus environments, the benefits of ON–OFF coding become dramatic, offering virtually unlimited savings of spikes for a given performance. In particular, this occurs when the behaviorally relevant stimuli consist of rare but large fluctuations of both polarities embedded in a sea of frequent small fluctuations (Fig. 12).

The benefit of ON–OFF splitting is apparent also when optimality criteria other than information are used. When we demand that the stimulus value be read out by a simple linear decoder, the ON–OFF system again offers substantial improvements, with a maximum gain of 40% in the SNR and a 4-fold to 10-fold reduction in the number of spikes required for a given performance (Fig. 8). Here again, the advantage of the ON–OFF system further increases with the sparseness of the stimulus distribution. The ON–OFF system also achieves better linear decoding of dynamic stimuli that include temporal correlations, of the type encountered by the retina in natural vision (Figs. 9, 10). Finally, we demonstrated that the efficiency gains achieved by ON–OFF splitting compare favorably with those that arise from lateral inhibition, a circuit function widely held to implement decorrelation and efficient coding of signals in the retina.

Representing the distribution of natural signals

Although ON–OFF pathway splitting reaps large benefits for coding large stimulus deviations of either sign, sensory systems clearly encode small fluctuations in the stimuli as well. For example, humans can detect visual stimuli with a contrast of $< 1\%$ (De Valois et al., 1974). One approach to cover the entire range of the natural stimulus distribution is to use different circuits for encoding large and small contrasts. For instance, the primate retina includes ganglion cell types with very different contrast sensitivity: highly sensitive “phasic” cells, and less sensitive “tonic” cells (Lee et al., 1989). Each of these cell classes includes ON and OFF varieties, consistent with the present modeling framework. Indeed, one can extend this formalism to cover multiple cell types with very different thresholds that span a broad stimulus distribution (Gjorgjieva et al., 2014). A second solution encountered in biological systems is adaptation to contrast, whereby the system adjusts its thresholds dynamically within the prevailing stimulus distribution (Smirnakis et al., 1997; Chander and Chichilnisky,

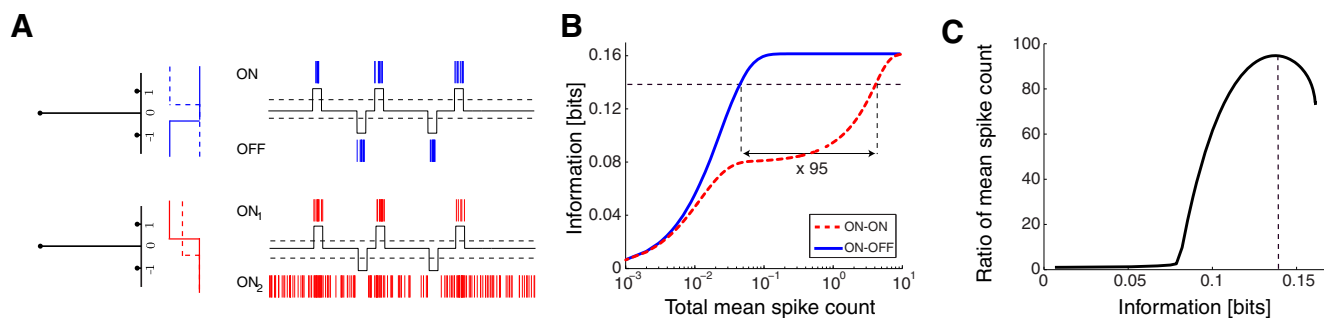


Figure 12. Sparse stimulus distributions enhance the advantage of ON–OFF splitting. **A**, Left, A stimulus distribution consisting of three events, $s = -1, 0, 1$, that occur with different probabilities $p_{-1}, p_0, p_1 = p_{-1}$, indicated by the bars. Right, Schematic responses of an ON–OFF system (top) and an ON–ON system (bottom) to such a stimulus. **B**, The mutual information for the ON–OFF and ON–ON systems as a function of the total mean spike count. Here the sparseness of the distribution is $1/p_1 = 100$. Given a desired level of information (horizontal dashed line), the ON–ON system needs 95 times higher firing rate than the ON–OFF system for the sparse stimulus in **A** (compare with Fig. 3). **C**, The ratio of the overall mean spike count used by the ON–ON versus the ON–OFF system as a function of the mutual information. The vertical dashed line corresponds to the dashed line in **B**.

2001). This contrast adaptation allows a system to maintain the proper choice of thresholds within the stimulus distribution.

In all our modeling, we assumed that large positive and negative stimulus values are equally represented and equally relevant. Is this justified? If one analyzes raw stimulus values—such as the light intensity in a natural scene or the intensity of natural sounds—the resulting distributions can be very asymmetric (Ruderman and Bialek, 1994; Dong and Atick, 1995; van Hateren, 1997; Thomson, 1999; Singh and Theunissen, 2003; Geisler, 2008; Ratliff et al., 2010). However, there are several aspects of early neural processing, universal across nervous systems, that effectively symmetrize the distribution of the relevant sensory signals.

Fundamentally, living systems are interested in the time derivative of natural signals, because a change in the environment requires action, whereas static features do not. A diverse set of adaptation mechanisms, both cellular and network based, accomplishes this sensitivity to change. For example, all retinal ganglion cells studied to date have a biphasic impulse response, characteristic of a differentiator, and the same applies to other animal senses (Reyes et al., 1994; Geffen et al., 2009; Kim et al., 2011; Nagel and Wilson, 2011) and even the sensory systems of bacteria (Block et al., 1982; Segall et al., 1986; Lazova et al., 2011) and fungi (Dennison and Foster, 1977; Foster and Smyth, 1980). After differentiation, the resulting signal is much more balanced in terms of positive and negative changes than the raw stimulus distribution because of the basic principle of “what goes up must come down.” Indeed, if one takes the retinal scenes that human subjects encounter during natural locomotion or saccadic eye movements and applies to them the biphasic filter characteristic of retinal processing, one obtains an almost perfectly symmetrical distribution of signals (Ratliff et al., 2010). Therefore, on a basic level, the cross-species universality of time differentiation in sensory systems leads to a symmetry between increments and decrements of the relevant stimulus variables, which in turn drives the universal appearance of ON–OFF coding in early sensory pathways.

At a more refined level, one often finds asymmetries in the distribution of natural stimuli that depend on the sensory modality in question. For example, the statistics of photon counts imply that bright and dim lights differ in SNR, and this has been invoked to explain certain asymmetries between ON and OFF pathways in the retina (Pandarinath et al., 2010; Nichols et al., 2013). These ON–OFF asymmetries propagate to downstream visual areas, including the thalamus and cortex (Jin et al., 2008; Yeh et

al., 2009; Kremkow et al., 2014), although it remains unknown whether ON and OFF ganglion cell signals are decoded together. Similarly, natural scenes tend to contain more regions of negative than of positive spatial contrast (Tadmor and Tolhurst, 2000), and indeed one finds that OFF ganglion cells tend to have smaller receptive fields and outnumber ON cells (Ratliff et al., 2010). Under an asymmetric stimulus distribution, our efficient coding models will also favor asymmetries in the response functions of ON and OFF cells, as observed in the experiment (Chichilnisky and Kalmar, 2002). Of course, pressures other than efficient coding play an important role. For instance, amphibians have a strong preference for hiding in dark spots (Himstedt, 1967; Roth, 1987), and this has been invoked to explain the numerical dominance of the OFF pathway in their retina. Given that the ON and OFF signals in the retina split already at the first synapse, evolution has been free to sculpt the downstream circuitry in each pathway to accomplish such fine adjustments to the two halves of the stimulus distribution.

Other drivers for ON–OFF coding

Here we treated ON–OFF coding as a specific instance of splitting into multiple sensory pathways and found that it improves the efficiency of neural transmission as measured by mutual information or by mean squared error of the optimal linear estimator. These two measures are entirely agnostic about the content of the transmission. They do not take into account that certain stimuli carry greater relevance than others or that downstream areas of the nervous system may need the information formatted in a certain way. It seems plausible that an agnostic criterion of this sort provides the evolutionary driving force toward ON–OFF coding, because this organization emerges in such a disparate collection of sensory systems that process anything from electric fields to odors. Nevertheless, one can envision other benefits that result from an ON–OFF pathway split. In particular, the split may greatly simplify downstream computations.

Consider the visual processing of moving objects. An object passing over a background produces an intensity change of one polarity at the leading edge and the opposite polarity at the trailing edge. Both edges are informative for analyzing the motion of the object. To process both edges by a standard motion-detecting circuit, it would be useful if both polarities were represented in the same way, namely by excitation of an interneuron with a rectified response. In fact, the retina itself already reaps these benefits: the ON–OFF direction-selective ganglion cells respond to motion of both bright and dark edges (Barlow et al., 1964).

This results from computations in both the outer and the inner plexiform layer using essentially identical circuits of ON- or OFF-bipolar cell terminals, starburst amacrine cells, and retinal ganglion cell dendrites (Borst and Euler, 2011). Motion computations in later stages, like the superior colliculus or visual cortex, can benefit similarly from the ON–OFF diversification. If, instead, the retina used only sensors of one polarity, then motion detection for bright edges and dark edges would require the construction of two very different circuits.

Another benefit from ON–OFF splitting is that it enables downstream computations based on the timing of action potentials. Rapid saccades of the eyes break vision up into a sequence of snap shots. After a saccade, many ganglion cells fire a brief burst of spikes, and the timing of the first action potential conveys a great deal of information about the image (Gollisch and Meister, 2008). It has been suggested that downstream visual centers can compute on the image content by combining just one spike per ganglion cell, with millisecond sensitivity to spike timing (Delorme and Thorpe, 2001; Gütig et al., 2013). This can work only if the entire range of image contrasts, including both increases and decreases, is signaled by a burst of spikes. Thus, any such spike-time computation requires retinal coding by ON and OFF cells.

As one seeks to understand further what motivated the split into 20 visual pathways emerging from the retina, the concerns of efficient coding will likely need to be supplemented by an understanding of the specific image functions computed in these circuits and what sensorimotor behaviors they support.

References

- Ala-Laurila P, Greschner M, Chichilnisky EJ, Rieke F (2011) Cone photoreceptor contributions to noise and correlations in the retinal output. *Nat Neurosci* 14:1309–1316. [CrossRef Medline](#)
- Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? *Network* 3:213–251. [CrossRef](#)
- Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. *Neural Comput* 2:308–320. [CrossRef](#)
- Atick JJ, Redlich AN (1992) What does the retina know about natural scenes? *Neural Comput* 4:196–210. [CrossRef](#)
- Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* 61:183–193.
- Barlow HB (1961) Possible principles underlying the transformations of sensory messages. In: *Sensory communication*, Chap 13, pp 217–234. Cambridge, MA: Massachusetts Institute of Technology.
- Barlow HB, Hill RM, Levick WR (1964) Retinal ganglion cells responding to selectively to direction and speed of image motion in the rabbit. *J Physiol* 173:377–407. [Medline](#)
- Bell AJ, Sejnowski TJ (1997) The “independent components” of natural scenes are edge filters. *Vision Res* 37:3327–3338. [CrossRef Medline](#)
- Bennett MVL (1971) Electrophysiology. In: *Fish physiology*, Vol V (Hoar WS, Randall DJ, eds), pp 347–491. New York: Academic.
- Berry MJ, Warland DK, Meister M (1997) The structure and precision of retinal spike trains. *Proc Natl Acad Sci U S A* 94:5411–5416. [CrossRef Medline](#)
- Bethge M, Rotermund D, Pawelzik K (2003) Optimal neural rate coding leads to bimodal firing rate distributions. *Network* 14:303–319. [CrossRef Medline](#)
- Block SM, Segall JE, Berg HC (1982) Impulse responses in bacterial chemotaxis. *Cell* 31:215–226. [CrossRef Medline](#)
- Borst A, Euler T (2011) Seeing things in motion: models, circuits, and mechanisms. *Neuron* 71:974–994. [CrossRef Medline](#)
- Chalasan SH, Chronis N, Tsunozaki M, Gray JM, Ramot D, Goodman MB, Bargmann CI (2007) Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature* 450:63–70. [CrossRef Medline](#)
- Chander D, Chichilnisky EJ (2001) Adaptation to temporal contrast in primate and salamander retina. *J Neurosci* 21:9904–9916. [Medline](#)
- Chichilnisky EJ, Kalmar RS (2002) Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neurosci* 22:2737–2747. [Medline](#)
- Delorme A, Thorpe SJ (2001) Face identification using one spike per neuron: resistance to image degradations. *Neural Netw* 14:795–803. [CrossRef Medline](#)
- Dennison DS, Foster KW (1977) Intracellular rotation and the phototropic response of *Physcomyces*. *Biophys J* 18:103–123. [CrossRef Medline](#)
- De Valois RL, Morgan H, Snodderly DM (1974) Psychophysical studies of monkey vision—III. Spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Res* 14:75–81. [CrossRef Medline](#)
- Dong DW, Atick JJ (1995) Statistics of natural time-varying images. *Network* 6:345–358. [CrossRef](#)
- Eliasmith C, Anderson CH (2002) *Neural engineering: computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: Massachusetts Institute of Technology.
- Field DJ (1994) What is the goal of sensory coding? *Neural Comput* 6:559–601. [CrossRef](#)
- Foster KW, Smyth RD (1980) Light antennas in phototactic algae. *Microbiol Rev* 44:572–630. [Medline](#)
- Gallio M, Ofstad TA, Macpherson LJ, Wang JW, Zuker CS (2011) The coding of temperature in the *Drosophila* brain. *Cell* 144:614–624. [CrossRef Medline](#)
- Geffen MN, Broome BM, Laurent G, Meister M (2009) Neural encoding of rapidly fluctuating odors. *Neuron* 61:570–586. [CrossRef Medline](#)
- Geisler WS (2008) Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol* 59:167–192. [CrossRef Medline](#)
- Gjorgjieva J, Meister M, Sompolinsky H (2014) Optimal sensory coding by populations of ON and OFF neurons. Presented at the Meeting of Computational and Systems Neuroscience, Salt Lake City, UT, February 27 to March 2.
- Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. *Science* 319:1108–1111. [CrossRef Medline](#)
- Gollisch T, Meister M (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* 65:150–164. [CrossRef Medline](#)
- Gütig R, Gollisch T, Sompolinsky H, Meister M (2013) Computing complex visual features with retinal spike times. *PLoS One* 8:e53063. [CrossRef Medline](#)
- Himstedt W (1967) Experimentelle analyse der optischen sinnesleistungen im beutefangverhalten der einheimischen urodelen. *Zool Jahrb Physiol* 73:281–320.
- Jin JZ, Weng C, Yeh CI, Gordon JA, Ruthazer ES, Stryker MP, Swadlow HA, Alonso JM (2008) On and off domains of geniculate afferents in cat primary visual cortex. *Nat Neurosci* 11:88–94. [CrossRef Medline](#)
- Joesch M, Schnell B, Raghu SV, Reiff DF, Borst A (2010) ON and OFF pathways in *Drosophila* motion vision. *Nature* 468:300–304. [CrossRef Medline](#)
- Kastner D, Baccus SA, Sharpee TO (2014) Second order phase transition describes maximally informative encoding in the retina. Presented at the Meeting of Computational and Systems Neuroscience, Salt Lake City, UT, February 27 to March 2.
- Kim AJ, Lazar AA, Slutskiy YB (2011) System identification of *Drosophila* olfactory sensory neurons. *J Comput Neurosci* 30:143–161. [CrossRef Medline](#)
- Kremkow J, Jin J, Komban SJ, Wang Y, Lashgari R, Li X, Jansen M, Zaidi Q, Alonso JM (2014) Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proc Natl Acad Sci U S A* 111:3170–3175. [CrossRef Medline](#)
- Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. *J Neurophysiol* 16:37–68. [Medline](#)
- Lazova MD, Ahmed T, Bellomo D, Stocker R, Shimizu TS (2011) Response rescaling in bacterial chemotaxis. *Proc Natl Acad Sci U S A* 108:13870–13875. [CrossRef Medline](#)
- Lee BB, Martin PR, Valberg A (1989) Sensitivity of macaque retinal ganglion cells to chromatic and luminance flicker. *J Physiol* 414:223–243. [Medline](#)
- Marder E, Abbott LF (1995) Theory in motion. *Curr Opin Neurobiol* 5:832–840. [CrossRef Medline](#)
- Masland RH (2012) The neuronal organization of the retina. *Neuron* 76:266–280. [CrossRef Medline](#)
- Nagel KI, Wilson RI (2011) Biophysical mechanisms underlying olfactory receptor neuron dynamics. *Nat Neurosci* 14:208–216. [CrossRef Medline](#)
- Nichols Z, Nirenberg S, Victor J (2013) Interacting linear and nonlinear characteristics produce population coding asymmetries between ON and OFF cells in the retina. *J Neurosci* 33:14958–14973. [CrossRef Medline](#)
- Nikitin AP, Stocks NG, Morse RP, McDonnell MD (2009) Neural popula-

- tion coding is optimized by discrete tuning curves. *Phys Rev Lett* 103:138101. [CrossRef Medline](#)
- Pandarinath C, Victor JD, Nirenberg S (2010) Symmetry breakdown in the ON and OFF pathways of the retina at night: functional implications. *J Neurosci* 30:10006–10014. [CrossRef Medline](#)
- Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454:995–999. [CrossRef Medline](#)
- Pitkow X, Meister M (2012) Decorrelation and efficient coding by retinal ganglion cells. *Nat Neurosci* 15:628–635. [CrossRef Medline](#)
- Pitkow X, Meister M (2014) Neural computation in sensory systems. In: *The cognitive neurosciences V* (Gazzaniga MS, Mangun GR, eds), pp 305–318. Cambridge, MA: Massachusetts Institute of Technology.
- Pitkow X, Sompolinsky H, Meister M (2007) A neural computation for visual acuity in the presence of eye movements. *PLoS Biol* 5:2898–2911. [CrossRef Medline](#)
- Ratliff CP, Borghuis BG, Kao YH, Sterling P, Balasubramanian V (2010) Retina is structured to process an excess of darkness in natural scenes. *Proc Natl Sci USA* 107:17368–17373. [CrossRef](#)
- Reyes AD, Rubel EW, Spain WJ (1994) Membrane properties underlying the firing of neurons in the avian cochlear nucleus. *J Neurosci* 14:5352–5364. [Medline](#)
- Roth G (1987) *Visual behavior in salamanders*. Berlin: Springer.
- Ruderman DL (1994) The statistics of natural images. *Network* 5:517–548. [CrossRef](#)
- Ruderman DL, Bialek W (1994) Statistics of natural images: Scaling in the woods. *Phys Rev Lett* 73:814–817. [CrossRef Medline](#)
- Schiller PH (1992) The on and off channels of the visual system. *Trends Neurosci* 15:86–92. [CrossRef Medline](#)
- Schiller PH, Sandell HH, Maunsell JHR (1986) Functions of the on and off channels of the visual system. *Nature* 322:824–825. [CrossRef Medline](#)
- Scholl B, Gao X, Wehr M (2010) Nonoverlapping sets of synapses drive On responses and Off responses in auditory cortex. *Neuron* 65:412–421. [CrossRef Medline](#)
- Segall JE, Block SM, Berg HC (1986) Temporal comparisons in bacterial chemotaxis. *Proc Natl Acad Sci U S A* 83:8987–8991. [CrossRef Medline](#)
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753. [CrossRef Medline](#)
- Shamai S (1990) Capacity of a pulse amplitude modulated direct detection photon channel. *IEE Proc Commun Speech Vis* 137:424–430.
- Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114:3394–3411. [CrossRef Medline](#)
- Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386:69–73. [CrossRef Medline](#)
- Stein RB (1967) The information capacity of nerve cells using a frequency code. *Biophys J* 7:797–826. [CrossRef Medline](#)
- Tadmor Y, Tolhurst DJ (2000) Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Res* 40:3145–3157. [CrossRef Medline](#)
- Thomson MG (1999) Visual coding and the phase structure of natural scenes. *Network* 10:123–132. [CrossRef Medline](#)
- Uzzell VJ, Chichilnisky EJ (2004) Precision of spike trains in primate retinal ganglion cells. *J Neurophysiol* 92:780–789. [CrossRef Medline](#)
- van Hateren JH (1992a) Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 171:157–170. [CrossRef](#)
- van Hateren JH (1992b) A theory of maximizing sensory information. *Biol Cybern* 68:23–29. [CrossRef Medline](#)
- van Hateren JH (1997) Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res* 37:3407–3416. [CrossRef Medline](#)
- Wässle H (2004) Parallel processing in the mammalian retina. *Nat Rev Neurosci* 5:747–757. [CrossRef Medline](#)
- Wässle H, Boycott BB, Illing RB (1981a) Morphology and mosaic of on- and off-beta cells in the cat retina and some functional considerations. *Proc R Soc Lond B* 212:177–195. [CrossRef Medline](#)
- Wässle H, Peichl L, Boycott BB (1981b) Morphology and topography of on- and off-alpha cells in the cat retina. *Proc R Soc Lond B* 212:157–175. [CrossRef Medline](#)
- Westheimer G (2007) The ON–OFF dichotomy in visual processing: from receptors to perception. *Prog Retin Eye Res* 26:636–648. [Medline](#)
- Yeh CI, Xing D, Shapley RM (2009) “black” responses dominate macaque primary visual cortex V1. *J Neurosci* 29:11753–11760. [CrossRef Medline](#)