bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Optimal sensory coding by populations of ON and OFF neurons

Julijana Gjorgjieva[1,2], Markus Meister[3], Haim Sompolinsky[2,4]
(1) Max Planck Institute for Brain Research, Frankfurt, Germany
(2) Center for Brain Science, Harvard University, Cambridge, MA, USA
(3) Division of Biology, California Institute of Technology, Pasadena, CA, USA,
(4) The Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem, Israel

Abstract

In many sensory systems the neural signal is coded by multiple parallel pathways, suggesting an evolutionary fitness benefit of general nature. A common pathway splitting is that into ON and OFF cells, responding to stimulus increments and decrements, respectively. According to efficient coding theory, sensory neurons have evolved to an optimal configuration for maximizing information transfer given the structure of natural stimuli and circuit constraints. Using the efficient coding framework, we describe two aspects of neural coding: how to optimally split a population into ON and OFF pathways, and how to allocate the firing thresholds of individual neurons given realistic noise levels, stimulus distributions and optimality measures. We find that populations of ON and OFF neurons convey equal information about the stimulus regardless of the ON/OFF mixture, once the thresholds are chosen optimally, independent of stimulus statistics and noise. However, an equal ON/OFF mixture is the most efficient as it uses the fewest spikes to convey this information. The optimal thresholds and coding efficiency, however, depend on noise and stimulus statistics if information is decoded by an optimal linear readout. With non-negligible noise, mixed ON/OFF populations reap significant advantages compared to a homogeneous population. The best coding performance is achieved by a unique mixture of ON/OFF neurons tuned to stimulus asymmetries and noise. We provide a theory for how different cell types work together to encode the full stimulus range using a diversity of response thresholds. The optimal ON/OFF mixtures derived from the theory accord with certain biases observed experimentally.

Introduction

The efficient coding hypothesis states that sensory systems have evolved to optimally transmit information about the natural world given limitations on their biophysical components and constraints on energy use [4]. This theory has been successfully applied to explain the structure of neuronal receptive fields in the mammalian retina [2, 3] and fly lamina [23, 43] based on the statistics of natural scenes. Similar arguments have been made to explain why early sensory pathways often split into parallel channels that represent different stimulus variables, for example different auditory waveforms [37], or local visual patterns [35]. However, even neurons that encode the same sensory variable often split further into distinct types. A commonly encountered diversification is into ON and OFF types: ON cells fire when the stimulus increases and OFF cells when it decreases. This basic ON-OFF dichotomy is found in many modalities, including vertebrate vision [22], invertebrate vision [18], thermosensation [14], and chemosensation [9]. Furthermore, even among neurons that encode the same sensory variable with the same sign, one often encounters distinct types that have different response thresholds, for example among touch receptors [41] and electroreceptors [6]. The same principle seems to apply several synapses downstream from the receptors [20], and even in the organization of the motor periphery, where motor neurons that activate the same muscle have a broad range of response thresholds [17]. In the present article we consider this pathway splitting among neurons that represent the same variable and explore whether it can be understood based on efficient coding theory.

One reason why the ON and OFF pathways have evolved may be to optimize information about both increments and decrements in stimulus intensity by providing excitatory signals for both [32]. For instance, if there were only ON neurons, such a cell would need high baseline firing rate to encode stimulus decrements, which can be very costly. A population of many neurons, however, could resolve this issue by tuning their thresholds so that they jointly code for the stimulus. We have previously addressed the benefits for having ON and OFF cells in a small population of just two cells [16]. Since ON and OFF neurons often exhibit a broad distribution of firing thresholds [6, 41, 17, 20], an important question is thus, what distribution of thresholds yields the

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

most efficient coding. Here we study optimal information transmission in sensory populations comprised of different mixtures of ON and OFF neurons that code for a common stimulus variable. We develop the problem parametrically in the neuronal noise and the distribution of stimuli that the cells encode.

The efficient coding hypothesis does not specify what quantity the neural population should optimize. Therefore, we consider two alternative measures of optimal coding that are in common use: first we maximize the mutual information between stimulus and response, and second we optimize the estimate of the stimulus obtained by a linear decoder of the response. When constraining the maximal firing rate of each cell, we find that counter to our expectations the mutual information is identical for any mixture of ON and OFF cells once the thresholds of all cells are optimized. This result is independent of the shape of the stimulus distribution or the level of neuronal noise. However, the total mean spike count is lowest for the population with equal numbers of ON and OFF cells, making this arrangement optimal in terms of bits per spike. Optimizing the linear decoder requires determining not only the cells' thresholds, but also the decoding weights in order to minimize the mean square error between the stimulus and its estimate. Under this criterion, the optimal ON/OFF mixture and cells' thresholds depend on the asymmetries in the stimulus distribution and the noise level. Our theory yields surprising results regarding the optimal organization of sensory populations comprised of ON and OFF cells. We make distinct predictions for the optimal distribution of thresholds under the two optimality measures, providing insight into the diverse coding strategies of these populations across different sensory modalities and species.

Results

Population coding model. We study a population of ON and OFF neurons that respond to a common stimulus. Model neurons are assumed to transmit information about a common scalar stimulus through the spike count observed during a short coding window. The duration of this coding window, T, is chosen based on the observed dynamics of neuronal responses; for instance, for retinal ganglion cells, T is typically in the range of 10-50 ms [42, 29]. Neuronal spike counts are stochastic and their mean is modulated by the stimulus through a binary response function: ON (OFF) neurons fire Poisson spikes with an average mean count $\nu_{\max}$ whenever the stimulus intensity s is above (below) their threshold θ_i, and zero otherwise, i.e. $\nu_i(s) = \nu_{\max}\Theta(s - \theta_i)$ for ON neurons and $\nu_i(s) = \nu_{\max}\Theta(\theta_i - s)$ for OFF neurons (Figure 1A), where Θ is the Heaviside function. The binary response function is provably optimal under the constraint of Poisson spiking and short coding window encountered in many sensory areas [38, 34, 7, 27] and offers a reasonable approximation of neural behavior in several systems [24, 29].

Since we use a discrete rate function, we can replace θ_i by the corresponding cumulative threshold which essentially maps any stimulus distribution into a uniform distribution from 0 to 1 (Figure 1B). Since 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.

Maximal mutual information for mixtures of ON and OFF neurons. Population responses are determined by the ratio of ON vs. OFF cells, as well as the distribution of firing thresholds of these cells. We determine the values of these parameters that maximize the Shannon mutual information between stimulus and population response, while constraining the maximum expected spike count $R = \nu_{\max}T$ for each cell. Biophysically, such a constraint on the maximal firing rate arises naturally from refractoriness of the spike-generating membrane. We have analytically proven the following theorem (SI Appendix):

Equal Coding Theorem: For a population of N ON and OFF Poisson neurons coding a one-dimensional stimulus in a fixed time window T by binary rate functions with maximal firing rate $\nu_{\max}$, the mutual information is identical for all ON/OFF mixtures when the thresholds are optimized, for all N, $\nu_{\max}$ and stimulus distributions.

Optimized mutual information equals the negative log probability of the quiescent state. The value of this maximal information depends on the maximal expected spike count, $R = \nu_{\max}T$, which controls the Poisson noise level, since in the active state (where the cell's firing rate is nonzero) the variance of spike count is R. It is therefore useful to introduce the noise parameter $q = e^{-R}$, ranging from $q = 0$ in the noiseless limit and $q = 1$ in the high noise limit. We find that for all ON/OFF mixtures, the information with optimized

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

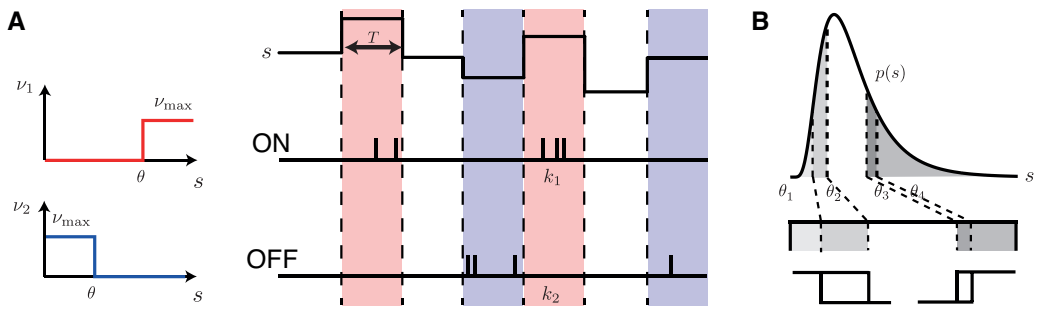

Figure 1. Neuron model and population coding framework. A. A stimulus s is encoded by a population of ON (red) and OFF (blue) cells. Each cell is described by a binary response nonlinearity ν with a threshold θ and maximal firing rate $\nu_{\max}$. During a coding window of fixed duration T the stimulus is constant and the spike count k is drawn from a Poisson distribution with a mean rate ν. **B**. When measuring coding efficiency using the mutual information between stimulus and spike count response, the neurons' thresholds can be interpreted as quantiles of the original stimulus distribution, thus mapping an arbitrary stimulus distribution $p(s)$ into a uniform distribution (four thresholds shown).

thresholds is given by (Figure 2A)

$$I = \log\left(N(1-q) + q^{-q/(1-q)}\right) + \frac{q}{1-q}\log q \tag{1}$$

Interestingly, we can write $I = -\log P(0)$ where $P(0)$ denotes the probability that no cells fire, i.e. the quiescent state. There are two contributions to $P(0)$: (1) the probability that the stimulus is in an interval in stimulus space which is not coded by any ON or OFF cells (white region in Figure 1B), and (2) the probability that the stimulus is in an interval in stimulus space that is coded by a subset of ON or OFF cells (grey regions in Figure 1B), but no cell fires due to the Poisson variability. In the noiseless limit, $R \to \infty$ (i.e. $q = 0$), I reaches its upper bound $I = \log(N+1)$. The effect of noise is most prominent when R is of order $1/N$, so that the total spike count RN is of order 1, implying that the signal-to-noise of the entire population is of order 1. We call this the **high noise regime**, and here we obtain $I \to \log(RN/e + 1)$.

Optimal distribution of thresholds. In the case of a discrete rate function, the information does not depend directly on the stimulus distribution $p(s)$, but only on the areas of $p(s)$ between consecutive thresholds. It is therefore useful to define the optimal threshold intervals $p_i = \int_{\theta_i}^{\theta_{i+1}} \mathrm{d}s\, p(s)$ where the neurons' thresholds are ordered $\theta_1 \leq \ldots \leq \theta_N$ (and we define the special $\theta_0 = -\infty$ and $\theta_{N+1} = \infty$). We find a surprisingly simple structure for the optimal p_i. The optimal thresholds divide stimulus space into intervals of equal area, p (Eq. 11), except for the *'edge'* intervals, p_{edge} (Eq. 12), and the *'silent'* interval, p_0, which separates the ON and OFF thresholds (Figure 2B). This p_0 is the only non-noisy response state where the firing rate of each cell is zero. We call this optimal threshold structure the **infomax** solution. We consider several limiting cases: first, a large population $N \gg 1$ and maximal firing rate per neuron R, which is much larger than $1/N$, i.e. $1-q = \mathcal{O}(1)$. We call this the **large population regime**. In this regime, $p_{\text{edge}} = p = p_0 = 1/(N+1)$, so the N thresholds divide stimulus space into $N+1$ equal intervals.

In this large population regime, we can rewrite the optimal thresholds as a continuous function of cumulative stimulus space; we replace θ_i with $\theta(x)$, where $x = i/N$ is between 0 and 1. Then the optimal thresholds equalize the area under the stimulus density,

$$x(\theta) = \frac{1}{Z}\int_{-\infty}^{\theta} p(\theta')\,\mathrm{d}\theta', \tag{2}$$

where Z is a normalization factor. Therefore, the population of cells achieves 'histogram equalization', a strategy that has been proven optimal for a single cell, which codes a stimulus using a continuous function (the cell's graded response) [23, 26].

In contrast, in the high noise regime, the system performs redundant coding so that $p \ll 1$ and the only two substantial threshold intervals are the edge intervals $p_{\text{edge}} = 1/e$, and the silent interval, $p_0 = 1 - 2/e$, where e

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

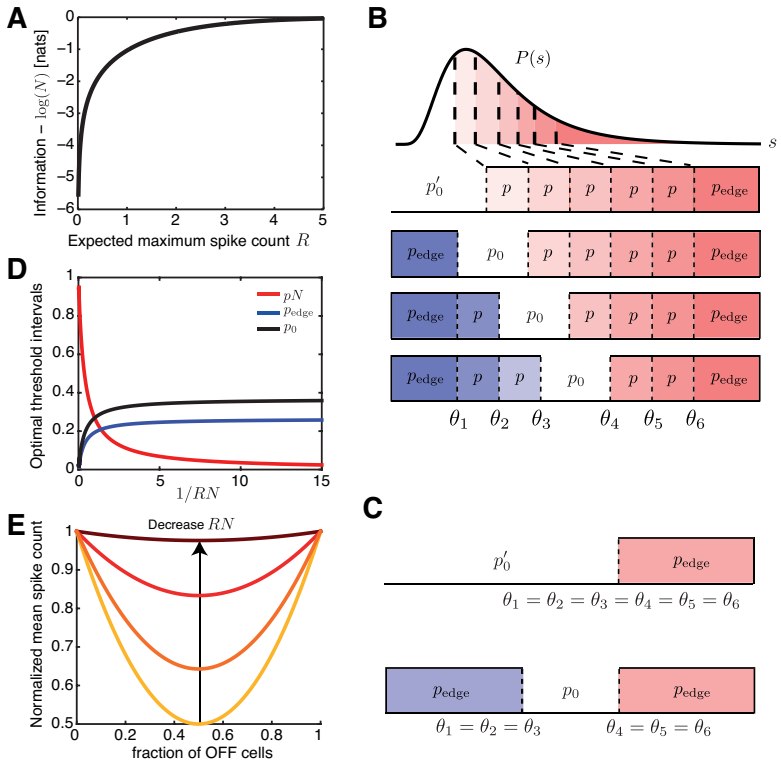

Figure 2. Mutual information when constraining the maximal spike count. A. The mutual information between stimulus and response for any combination of N ON and OFF cells is identical when constraining the maximum expected spike count, R. **B**. The optimal threshold intervals for all possible mixtures of ON (red) and OFF (blue) cells in a population of six cells that achieve the same mutual information about a stimulus s from an arbitrary distribution $p(s)$. **C**. The optimal threshold intervals in the high noise regime, $RN \to 0$, for the homogeneous mixture of all ON cells (top) and for the equal ON/OFF mixture (bottom). **D**. The optimal threshold intervals as a function of $1/RN$. **E**. The mean spike count required to transmit the same information (see A) by populations with a different fraction of OFF neurons (α), normalized by the mean spike count of the homogeneous population with $\alpha = 0$. The different curves denote $RN = \{0.1, 1, 5, 100\}$.

denotes exp(1) (Figure 2C,D). This implies that the optimal solution is to place all ON thresholds, and similarly all OFF thresholds, at roughly the same value, maximizing redundancy and noise reduction.

Mean population firing rate depends on the ON/OFF mixture. Despite equality in information for all ON/OFF mixtures (Figure 2A), each optimized ON/OFF population uses a different mean spike count to achieve this information. In the large population regime, the mean spike count per neuron is $r(\alpha) = R(\alpha^2 + (1-\alpha)^2)/2$ where α is the fraction of OFF cells. This mean spike count per neuron is minimized at $\alpha = 1/2$, where it is half of the mean spike count for the homogeneous population, $r(0) = R/2$ (Figure 2E, yellow). As the noise increases, the relative benefits of the equally mixed relative to the homogeneous population decrease (Figure 2E). In the high noise regime, all mixtures produce roughly the same mean spike count per neuron of R/e (Figure 2E, brown).

Minimizing mean square error of the optimal linear readout. The mutual information tells us how well the population represents the stimulus without regard for how it can be decoded. An alternative criterion for coding efficiency is the ability of downstream neurons to decode this information. A simple biologically plausible decoding mechanism is linear decoding [11, 33, 46]. Here we study the accuracy of a downstream neuron that estimates the stimulus value s using a weighted sum of spike counts n_i of the upstream population

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

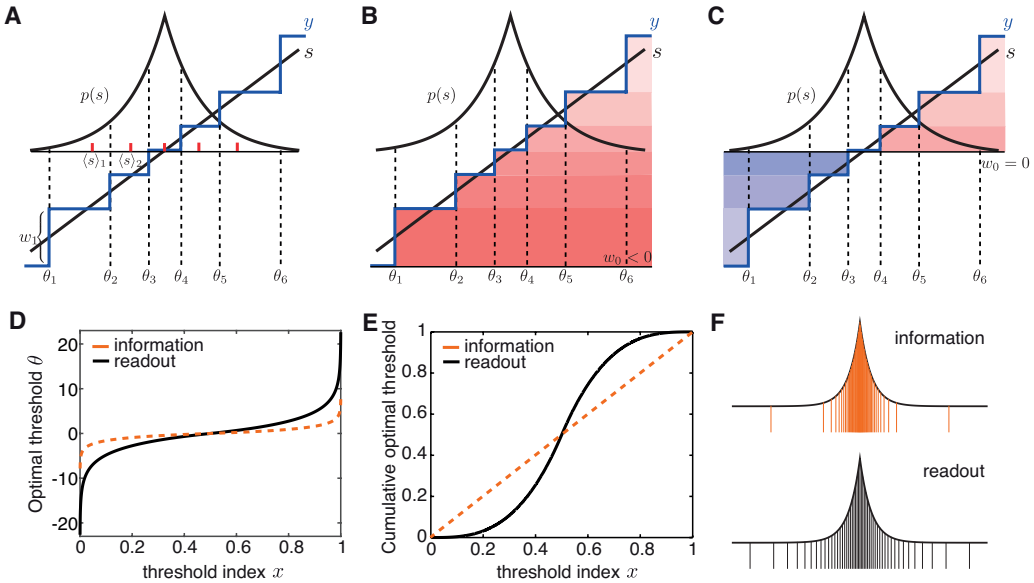

Figure 3. Optimal linear decoding of stimuli without noise is independent of ON/OFF mixture. A. Minimizing the MSE between a stimulus s drawn from distribution $p(s)$ (black) and its linear estimate y (blue) by a population of (6) ON and OFF cells, in the absence of noise, i.e. $R \to \infty$. The optimal weights w_i are the difference between, while the optimal thresholds θ_i are the average of two centers of mass $\langle s \rangle_i$ of each interval between neighboring thresholds (red dashes). **B,C**. Any ON/OFF population can achieve the same error with the same set of optimal thresholds and weights but a different constant, w_0. B: 6 ON cells ($w_0 < 0$), and 3 OFF and 3 ON cells ($w_0 = 0$). **D**. The optimal thresholds equalize not the area under the stimulus density (as in the case of the mutual information, Eq. 2), but the area under its one-third power (Eq. 6). The optimal thresholds are shown for the large population regime and a Laplace distribution. **E**. The cumulative optimal thresholds $\int_{-\infty}^{\theta(x)} p(\theta')\mathrm{d}\theta'$. **F**. The infomax thresholds partition the Laplace distribution into intervals that code for stimuli with higher likelihood of occurrence, while minimizing the MSE favors rarer stimuli near the tails of the distribution.

of binary Poisson neurons with thresholds θ_i

$$y = \sum_{i=1}^{N} w_i \, n_i + w_0. \tag{3}$$

The weights w_i, constant w_0 and thresholds θ_i are optimized to minimize the mean square estimation error (MSE).

Accuracy of the optimal linear readout without noise. We first consider the scenario of low noise ($R \to \infty$), in which case the limitation on the accuracy of the stimulus reconstruction comes solely from the discreteness of the rate functions of each cell in the population. Unlike maximizing the information, when minimizing the MSE both weights and thresholds depend on the stimulus distribution $p(s)$ (Figure 3). Interestingly, we find that in this low noise limit, the optimal MSE is proportional to $1/N^2$ and is the same for all ON/OFF mixtures, including the homogeneous population. The optimal weights are given by

$$w_i = \langle s \rangle_i - \langle s \rangle_{i-1} \tag{4}$$

where $\langle s \rangle_i$ are the centers of mass of intervals of $p(s)$ intersected by neighboring thresholds (Methods Eq. 15, Figure 3A). The optimal thresholds are the average of two neighboring centers of mass

$$\theta_i = \frac{1}{2}(\langle s \rangle_i + \langle s \rangle_{i-1}). \tag{5}$$

The constant term and the stimulus interval not coded by any cell depend on the ON/OFF mixture (Figure 3B,C). In the large population regime, we can rewrite the thresholds θ_i as a continuous function $\theta(x)$ of the

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cumulative stimulus space $x = i/N$ between 0 and 1. Interestingly, the optimal thresholds equalize not the area under the stimulus density, as in the case of the mutual information (Eq. 2), but the area under its one-third power

$$x(\theta) = \frac{1}{Z}\int_{-\infty}^{\theta} p(\theta')^{1/3}\,\mathrm{d}\theta' \tag{6}$$

where Z is a normalization factor. We invert this relationship to derive the optimal thresholds $\theta(x)$. Since the optimal MSE depends on the stimulus distribution, from now on we consider the Laplace distribution $p(s) = 1/2\, e^{-|s|}$, which arises from fits produced by filtering natural images with difference-of-Gaussian linear filters corresponding to center-surround receptive fields [13, 5]. In this case, the optimal thresholds can be derived from Eq. 6 (Figure 3D):

$$\theta(x) = \begin{cases} 3\log(2x), & x \leq \frac{1}{2} \\ -3\log(2(1-x)), & x > \frac{1}{2}. \end{cases} \tag{7}$$

The infomax thresholds are the same except that the pre-factor is 1 instead of 3, making them less spread out in the tails (Figure 3D). In particular, the largest thresholds (in magnitude) are $\pm 3\log(2N)$ when optimizing the MSE, three times as large as in the infomax case, $\pm\log(2N)$. To further compare the optimal thresholds from minimizing the MSE and the infomax solution, we also plot the cumulative optimal thresholds $\int_{\infty}^{\theta(x)} p(\theta')\mathrm{d}\theta'$ (Figure 3E). While the optimal strategy when maximizing the information is to emphasize stimuli with higher likelihood of occurring, minimizing the MSE of the optimal linear readout pushes thresholds towards relatively rare stimuli near the tails of the stimulus distribution (Figure 3F).

Mixed ON/OFF populations in the presence of noise. When we introduce noise into the system so that the maximum expected spike count R is order 1, mixed ON/OFF populations show a dramatic improvement of the MSE over predominantly homogeneous populations (Figure 4A). For the Laplace distribution we have considered so far, and different noise values, the optimal fraction of OFF cells in the population is $\alpha = 1/2$. Although there is a unique best ON/OFF mixture, the performance of populations with similar proportions of ON and OFF neurons is similar (i.e. the MSE around $\alpha = 1/2$ is flat), while the homogeneous population has the highest MSE. As the noise decreases (R increases), this difference in performance between the mixed and homogeneous populations becomes even more dramatic, see for example $R = 1$ (Figure 4A). For a large population, where we assume small difference between two neighboring thresholds, we can can derive the asymptotic limit (SI Appendix) by expanding $p(s)$ around each threshold to determine how the MSE scales with the population size N and the noise level R. For mixed ON/OFF populations, the variance term dominates the MSE scaling as $1/N$, while the constant term (determined solely by the discreteness of the rate functions) remains proportional to $1/N^2$ [33]. For homogeneous populations, the variance term still dominates the MSE scaling as $\log^2(RN)/(RN)$, while the constant term as $\log(RN)/(RN)$.

In addition to the big difference in coding performance between mixed and homogeneous populations, the presence of noise also qualitatively changes the distribution of optimal thresholds (Figure 4B). In the asymptotic limit, the thresholds for a mixed population with an equal number of ON and OFF cells follow logarithmic profiles as a function of cumulative stimulus space ($x = i/N$) similar to the noiseless case (Eq. 7), except that the pre-factor is 2 instead of 3. This is also the case for any population where neither cell type predominates over the other, see for instance the population with 2/3 OFF cells and 1/3 ON cells (Figure 4C, red, blue). Thus, the noise has the effect of concentrating the thresholds near more likely stimuli, increasing the redundancy of the code.

When there is a pronounced over-representation of one cell type in the mixture (as in the extreme case of the homogeneous population), the optimal thresholds $\theta(x)$ exhibit a distinct asymmetry. The thresholds corresponding to the more abundant population are distributed linearly with x, while the thresholds corresponding to the less abundant population are distributed logarithmically with x, as before (SI Appendix). We demonstrate this asymmetry in the extreme case of the homogeneous population, e.g. all ON neurons (Figure 4B,C, black):

$$\theta(x) = \begin{cases} -(1-x)\log(RN), & 0 < x \leq \left(1 - \frac{\sqrt{2}}{\log(RN)}\right)^{-1} \\ -2\log(1-x), & \left(1 - \frac{\sqrt{2}}{\log(RN)}\right)^{-1} < x \leq 1. \end{cases} \tag{8}$$

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

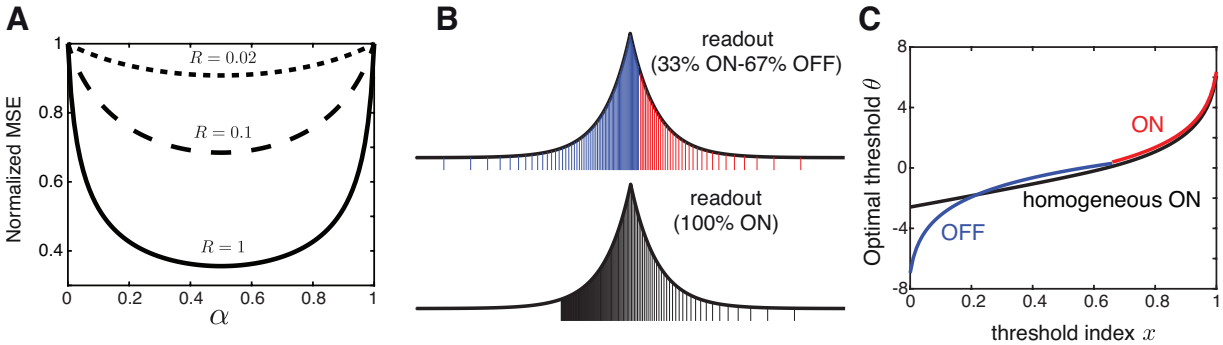

Figure 4. Optimal linear decoding of stimuli with noise favors different ON/OFF mixtures. A. The MSE as a function of the fraction of OFF cells, α, for a different maximum spike count, R. The MSE was normalized to the MSE for the homogeneous population of all ON cells. The MSE is shown for $N = 100$ cells and for the Laplace distribution. **B**. The optimal thresholds for the homogeneous population partition the Laplace stimulus distribution starting with a much larger threshold than the mixed population with 2/3 OFF cells (blue) and 1/3 ON cells (red). **C**. The optimal thresholds for the Laplace distribution for a homogeneous population (black) and a mixed population with 2/3 OFF cells (blue) and 1/3 ON cells (red). In B and C, $R = 1$. Note the difference in the optimal threshold distribution between the mixed ON/OFF and the homogeneous ON population, especially for small $x = i/N$ (logarithmic vs. linear).

Moreover, the smallest threshold for the homogeneous population is much larger than the smallest threshold for any mixed population, suggesting that there is a large region of stimuli that is not coded by any cell in the homogeneous case.

In summary, we conclude that introducing noise has a dramatic effect on the coding efficiency of different ON/OFF mixtures when the MSE of the optimal linear readout is minimized: coding by mixed ON/OFF populations is much better compared to coding by populations dominated by one cell type, such as the homogeneous population. When considering the optimal thresholds, we find that introducing noise does not qualitatively affect the shape of the optimal threshold profile for mixed ON-OFF populations compared to the noiseless case (except for a constant pre-factor). However, for the homogeneous population, the addition of noise qualitatively changes the shape of the optimal thresholds, leaving a large stimulus region that is not coded by any cell.

The optimal ON-OFF mixture of the linear readout depends on the asymmetry in the stimulus distribution. Depending on the sensory modality, the distribution of natural stimuli may be asymmetric around the most likely stimulus. In the case of vision, it has been shown that the distribution of contrasts in natural images is indeed skewed towards more negative values [30, 39]. Therefore, we instead consider an asymmetric Laplace distribution $p(s) \propto e^{s/\tau_-}$ for $s < 0$ and $p(s) \propto e^{-s/\tau_+}$ for $s \geq 0$. In this case, by minimizing the MSE we predict that the optimal ON/OFF mixture will be tuned to these stimulus asymmetries. At a fixed noise (or maximum spike count R), increasing the negative stimulus bias favors more OFF cells of the optimally configured population (Figure 5A,B). Similarly, increasing the positive bias in the stimulus distribution would favor an increase of the ON cells in the population. At a fixed level of stimulus bias, increasing the value of the noise accentuates the asymmetry in the optimal ON/OFF mixture (Figure 5C).

In summary, our theory predicts different optimal ON-OFF mixtures at which the lowest MSE is achieved depending on an asymmetry in the stimulus distribution and changing the noise level. Even in nature, the relative predominance of ON and OFF cells in populations can be different, for instance in the retinas of different species [10, 30]. Therefore, if we know the natural stimulus distribution being encoded by a population and the bounds on cells' firing rates, we can predict the optimal ON/OFF ratio, as well as the tuning properties of the cells and compare them to experimental observations. We anticipate that our theory would be a good match in early sensory populations of primary sensory neurons, which have mostly likely been adapted to the statistics of the stimulus distribution.

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

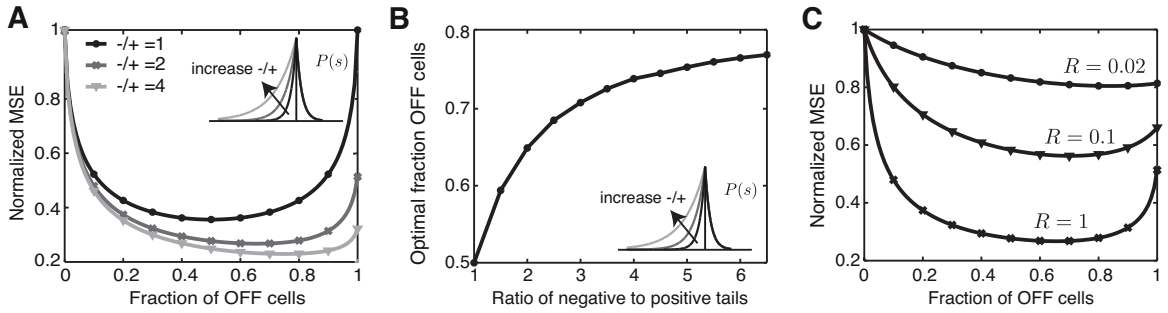

Figure 5. The optimal ON/OFF mixture derived from the linear readout is tuned to asymmetries in the stimulus distribution. A. The MSE as a function of the fraction of OFF cells in a population when constraining the maximum spike count, R. The MSE was normalized to the MSE for the homogeneous population of all ON cells. The MSE is shown for an asymmetric Laplace distribution with varying negative to positive bias $-/+$ and noise level $R = 1$ and $N = 100$. **B**. The optimal fraction OFF cells as a function of stimulus bias for the asymmetric Laplace distribution and noise level $R = 1$. **C**. Same as A but for an asymmetric Laplace distribution with a negative bias $-/+ = 2$ and varying noise.

Discussion

Information in neural circuits is processed by many different cell types, but it remains a challenge to understand how these distinct cell types work together. Here we treat a puzzling aspect of neural coding, how do discrete cell types conspire to collectively encode a single relevant variable? We use information theoretic measures to offer a plausible explanation: that neurons diversify their responses to maximize the coding efficiency of the range of occurring stimuli given biophysical constraints. The efficient coding framework we developed covers two aspects of the population code: (1) how to optimally split a population into ON and OFF pathways, and (2) how to allocate the thresholds of the individual neurons in the population as a function of the noise level, the stimulus distribution and the optimality measure.

We first examined optimal coding by maximizing mutual information: when constraining the maximum firing rate, all ON/OFF mixtures yield the same maximal information about the stimulus when thresholds are optimized (Figure 2). The optimal thresholds divide stimulus space into equal intervals with size dependent on the noise, except for the edge intervals coded by the largest ON and OFF thresholds. However, the system with an equal number of ON and OFF cells is most efficient, because it conveys the highest information per spike (Figure 2). The invariance in coding performance by different ON/OFF mixtures is also present when we demand the stimulus to be read out by an optimal linear readout, but only in the absence of noise (Figure 3). In the biologically relevant regimes of non-negligible noise, noise has a dramatic influence on the optimal performance realized by different ON/OFF mixtures (Figure 4). Populations with a similar number of ON and OFF cells have a much smaller decoding error than populations dominated by one cell type. The extreme case of the homogeneous population performs dramatically worse than any mixed population. Our theory predicts the optimal ON/OFF mixture depending on the asymmetry in the stimulus distribution and the amount of noise – thus making it relevant for sensory systems beyond the retina where ON and OFF pathways are encountered (Figure 5). However, the distribution of optimal thresholds is not much affected by the presence of noise, except for a scaling factor. The huge difference in performance reflected in the MSE between the homogeneous and any mixed population seems to be due to the threshold distribution – the optimal thresholds acquire qualitatively a very different structure in the homogeneous case, leaving a large stimulus region not coded by any cell.

How comparable is coding by different ON/OFF mixtures for the two optimality measures? Generally, the infomax criterion implements an optimal strategy which emphasizes stimuli that occur with higher probability. In contrast, minimizing the mean square error of the linear readout implements a more conservative strategy that utilizes more cells in the encoding of rarer stimuli due to a larger error penalty (Figure 3F, 4B). What ON-OFF thresholds one finds in a biological system could be indicative of the optimality measure utilized by that system.

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our work predicts the firing thresholds of neuronal populations in the high noise regime, which corresponds to short coding windows commonly encountered in biology; for instance, in the mammalian retina computations can be performed with an average of a few spikes per coding window [28, 42, 29]. In the low noise regime when the coding window is sufficiently long, or there is a large number of neurons, our results agree with previous studies on infomax and the optimal linear readout [23, 25, 45]. Efficient coding in the high noise regime has previously been examined, but only in terms of the transfer function of a single neuron [7, 27]. Using numerical simulations it was shown that the optimal transfer function is binary, as commonly encountered in biological systems [7, 27]. We go beyond this work and provide analytical solutions for how a population of binary neurons should coordinate their response ranges to optimally represent a given stimulus in the realistic regimes of short encoding times.

In our theory we assumed that variability in the spiking output of each cell is the dominant source of noise; this is in agreement with the small shared variability observed among retinal ganglion cells [29]. However, nearby ganglion cells can also exhibit strong noise correlations under certain conditions, presumably arising from noise in shared presynaptic neurons as early as the photoreceptors [1], or produced by circuit interactions [40, 47]. We expect that our results will continue to hold at low levels of input noise. At high levels of input noise, a different coding strategy may be optimal as previous work has shown that neurons perform redundant coding whereby the thresholds converge to the same value [40, 16, 21, 8].

Our predictions for the optimal ON/OFF mixture and the population thresholds could be directly compared to experiment. For instance, when deriving the optimal ON/OFF mixture based on the optimal linear readout, the imbalance in ON vs. OFF cell arises as a result of asymmetries in the stimulus distribution and the amount of noise. Indeed, 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 skewed towards negative stimuli [31, 12, 44, 39, 36, 15, 30]. Our theory then predicts that more resources should be spent on OFF, which is consistent with the predominance of OFF retinal ganglion cells in the retinas of a variety of species [10, 30]. Other related theories of ON/OFF splitting which have examined efficient coding including the spatial dimension, have also shown that OFF cells predominate [19, 30]. Besides predicting the optimal ON/OFF ratio, in our encoding model of one stimulus variable, we derive how the optimal ON and OFF subpopulations should coordinate their thresholds to achieve optimal encoding.

Given the ubiquity of ON/OFF pathway splitting in different sensory modalities and species, it may be possible to test our predictions in other sensory systems where the neuronal responses properties may be different, depending on differences in the natural stimulus distribution and the bounds on cells' firing rates. The challenge in each case might be to determine the natural stimulus distributions that the populations might be optimized to encode. Comparing the predicted optimal and the measured threshold distributions would provide a test whether the efficient coding criteria proposed here are a likely constraint that shapes the evolution of sensory systems.

Materials and Methods

Mutual information. The Shannon mutual information between the stimulus s and the spiking response $\boldsymbol{n}$ of the population is the difference between response and noise entropy:

$$I(s,\boldsymbol{n}) = H(\boldsymbol{n}) - H(\boldsymbol{n}|s) = -\langle \log p(\boldsymbol{n})\rangle_{\boldsymbol{n}} + \sum_{i=1}^{N} \langle \log p(n_i|s)\rangle_{n_i,s} \tag{9}$$

where $\langle\cdot\rangle_x$ denote averages over the distribution $p(x)$ and $p(\boldsymbol{n}) = \langle p(\boldsymbol{n}|s)\rangle_s$. We assume that stimulus encoding by all neurons is statistically independent conditional on s. Given the noise model, knowing the stimulus s unambiguously determines the response firing rate ν. Therefore, we can replace $p(n_i|s)$ with $p(n_i|\nu)$ which is Poisson distributed: $p(n_i|\nu) = \frac{[\nu T]^{n_i}}{n_i!} e^{-\nu T}$.

For a binary response function with two firing rate levels, 0 and $\nu_{\max}$, we can lump together all states with

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nonzero spike counts into a single state which we denote as **1**. Correspondingly, the state with zero spikes is **0**:

$$\begin{aligned} p(\mathbf{0}|\nu=0)&=1, & p(\mathbf{1}|\nu=0)&=0,\\ p(\mathbf{0}|\nu=\nu_{\max})&=q, & p(\mathbf{1}|\nu=\nu_{\max})&=1-q, \end{aligned} \tag{10}$$

where $q=e^{-R}$ and $R=\nu_{\max}T$ denote the level of noise in the system.

The optimal cumulative thresholds $\theta_1 \leq \ldots \leq \theta_N$ divide the stimulus into $N+1$ intervals which have equal area

$$p_i = p = \int_{\theta_i}^{\theta_{i+1}} \mathrm{d}s\, p(s) = \frac{1-q}{q^{-q/(1-q)}+N(1-q)} \tag{11}$$

except for the two *'edge'* intervals,

$$p_{\text{edge}} = \int_{-\infty}^{\theta_1} \mathrm{d}s\, p(s) = \int_{\theta_N}^{\infty} \mathrm{d}s\, p(s) = \frac{p}{1-q}, \tag{12}$$

and the *'silent'* interval that separates the ON and OFF thresholds, $p_0 = 1-(N-2)p-2p_{\text{edge}}$ (SI Appendix). Note that for the homogeneous population, $p_0' = 1-(N-1)p-p_{\text{edge}}$.

Optimal linear readout. No noise ($R\to\infty$). Here we consider the ON and OFF thresholds separately. In a population of N neurons, we assume that there are N_- OFF neurons and N_+ ON neurons. The mean square error can be written as:

$$E = \langle (s-y)^2\rangle = \langle (s-w_0)^2\rangle + w^T C w - 2w^T U \tag{13}$$

where y is the stimulus estimate from Eq. 3, C denotes the matrix of pairwise correlations between the cells' responses, $C_{ij}=C_i$ for $i\geq j$,and U denotes the correlation between response and stimulus:

$$C_i = \langle \Theta(s-\theta_i)\rangle \;\text{ and }\; U_i = \langle s\Theta(s-\theta_j)\rangle \tag{14}$$

with $i=1,\ldots,N_+$ for the ON cells and similarly for the OFF cells with $i=1,\ldots,N_-$. In the optimal solution the ON and OFF responses do not overlap; thus, there is no correlation between them. Optimizing the error with respect to the weights gives Eq. 4 for the optimal weights, while optimizing with respect to the threshold gives Eq. 5 for the optimal thresholds, with the centers of mass $\langle s\rangle_i$ defined as:

$$\sum_{0\leq j\leq i} w_j = \frac{\int_{\theta_i}^{\theta_{i+1}} ds\, s\, p(s)}{\int_{\theta_i}^{\theta_{i+1}} ds\, p(s)} = \langle s\rangle_i,\; 1\leq i\leq N \tag{15}$$

where we have defined $\theta_{N+1}=\infty$. Optimizing with respect to the constant term yields for the homogeneous population:

$$w_0 = \frac{\int_{-\infty}^{\theta_1} ds\, s\, p(s)}{\int_{-\infty}^{\theta_1} ds\, p(s)} \tag{16}$$

and for a mixed population:

$$w_0^m = \frac{\int_{\theta_1{}^{\text{OFF}}}^{\theta_1^{\text{ON}}} \mathrm{d}s\, s\, p(s)}{\int_{\theta_1{}^{\text{OFF}}}^{\theta_1^{\text{ON}}} \mathrm{d}s\, p(s)} \tag{17}$$

where $\theta_1{}^{\text{OFF}}$ denotes the largest OFF threshold and $\theta_1{}^{\text{ON}}$ denotes the smallest ON threshold in the population.

Non-negligible noise (finite R). We normalize readout, for convenience

$$y = \frac{1}{R}\sum_i w_i n_i + w_0. \tag{18}$$

The error can be written as before (Eq. 13) with correlations

$$C_{ij} = \frac{1}{R}\langle\langle n_i\rangle_n\langle n_j\rangle_n\rangle + \frac{1}{R^2}\delta_{ij}\langle\langle n_i\rangle_n\rangle = \langle\Theta(s-\theta_i)\Theta(s-\theta_j)\rangle + \frac{1}{R}\delta_{ij}\langle\Theta(s-\theta_i)\rangle \tag{19}$$

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $\delta_{ij} = 1$ if $i = j$ and otherwise 0. If we define $C_i = \langle \Theta(s - \theta_i) \rangle$ as before, then for $i \geq j$:

$$C_{ij} = C_i + \frac{1}{R}\delta_{ij} C_i \quad \text{and now} \quad U_i = \langle s\Theta(s - \theta_j) \rangle - w_0 C_i \tag{20}$$

with $i = 1, \ldots, N_+$ for the ON cells and similarly for the OFF cells with $i = 1, \ldots, N_-$. Optimizing with respect to the ON (OFF) weights:

$$w = C^{-1} U \tag{21}$$

and optimizing with respect to the ON (OFF) thresholds:

$$\theta_i = w_0 + \sum_{j \leq i} w_j - \frac{w_i}{2}\left(1 - R^{-1}\right), \tag{22}$$

with $i = 1, \ldots, N_+$ for the ON cells and similarly for the OFF cells with $i = 1, \ldots, N_-$. To solve these equations numerically, we implement an iterative procedure that rapidly converges to the optimal solution: starting from an ansatz for the thresholds, we compute C and U and obtain w from Eq. 21, which is used to derive the new set of thresholds.

Acknowledgments

JG was supported by the Max Planck Society and a Burroughs-Wellcome Career Award at the Scientific Interface. All authors were supported by the NIH, the Gatsby Charitable Foundation and the Swartz Foundation.

References

1. P. Ala-Laurila, M. Greschner, E. J. Chichilnisky, and F. Rieke. Cone photoreceptor contributions to noise and correlations in the retinal output. *Nat Neurosci*, 14:1309–1316, 2011.

2. J. J. Atick and A. N. Redlich. Towards a theory of early visual processing. *Neural Comput*, 2:308–320, 1990.

3. J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural Comput*, 4:196–210, 1992.

4. H. B. Barlow. Possible principles underlying the transformations of sensory messages. In *Sensory Communication*, chapter 13, pages 217–234. MIT Press, 1961.

5. A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Res*, 37:3327–3338, 1997.

6. C. C. Bell. Mormyromast electroreceptor organs and their afferent fibers in mormyrid fish. III. Physiological differences between two morphological types of fibers. *J Neurophysiol*, 63:319–332, 1990.

7. M. Bethge, D. Rotermund, and K. Pawelzik. Optimal neural rate coding leads to bimodal firing rate distributions. *Network: Comput Neur Syst*, 14:303–319, 2003.

8. B. A. W. Brinkman, A. I. Weber, F. Rieke, and E. Shea-Brown. How do efficient coding strategies depend on origins of noise in neural circuits. *PLoS Comp Biol*, 12:e1005150, 2016.

9. S. H. Chalasani, N. Chronis, M. Tsunozaki, J. M. Gray, D. Ramot, M. B. Goodman, and C. I. Bargmann. Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature*, 450:63–70, 2007.

10. E. J. Chichilnisky and R. S. Kalmar. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neurosci*, 22:2737–2747, 2002.

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11. P. Dayan and L. F. Abbott. *Theoretical neuroscience: computational and mathematical modelling of neural systems*. The MIT Press, Cambridge, Massachusetts, London, England, 2001.

12. D. W. Dong and J. J. Atick. Statistics of natural time-varying images. *Network: Comput Neural Syst*, 6:345–358, 1995.

13. D. J. Field. What is the goal of sensory coding? *Neural Comput*, 6:559–601, 1994.

14. M. Gallio, T. A. Ofstad, L. J. Macpherson, J. W. Wang, and C. S. Zuker. The coding of temperature in the *Drosophila* brain. *Cell*, 144:614–624, 2011.

15. W. S. Geisler. Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol*, 59:167–192, 2008.

16. J. Gjorgjieva, H. Sompolinsky, and M. Meister. Benefits of pathway splitting in sensory coding. *J Neurosci*, 34:12127–12144, 2014.

17. E. F. Hodson-Tole and J. M. Wakeling. Motor unit recruitment for dynamic tasks: current understanding and future directions. *J Comp Physiol B*, 179:57–66, 2009.

18. M. Joesch, B. Schnell, S. V. Raghu, D. F. Reiff, and A. Borst. ON and OFF pathways in *Drosophila* motion vision. *Nature*, 468:300–304, 2010.

19. Y. Karklin and E. P. Simoncelli. Efficient coding of natural images with a populations of noisy linear-nonlinear neurons. In *Adv Neural Inf Proc Syst 24*. MIT Press, Cambridge, MA, 2011.

20. D. B. Kastner and S. A. Baccus. Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nat Neurosci*, 14:1317–1322, 2011.

21. D. B. Kastner, S. A. Baccus, and T. O. Sharpee. Critical and maximally informative encoding between neural populations in the retina. *Proc Natl Acad Sci USA*, 112:2533–2538, 2015.

22. S. W. Kuffler. Discharge patterns and functional organization of mammalian retina. *J Neurophysiol*, 16:37–68, 1953.

23. S. A. Laughlin. Simple coding procedure enhances a neuron's information capacity. *Z Naturforsch C*, 36:910–912, 1981.

24. G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck. Neural coding of naturalistic motion stimuli. *Network: Comput Neural Syst*, 12:317–329, 2001.

25. J.-P. Nadal and N. Brunel. Nonlinear feedforward networks with stochastic output: infomax implies redundancy reduction. *Network Comput Neural Syst*, pages 207–217, 1998.

26. J.-P. Nadal and N. Parga. Nonlinear neurons in the low noise limit: a factorial code maximizes information transfer. *Network Comput Neural Syst*, 5:565–581, 1994.

27. A. P. Nikitin, N. G. Stocks, and M. D. McDonnell. Neural population coding is optimized by discrete tuning curves. *Phys Rev Lett*, 103:138101, 2009.

28. J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454:995–999, 2008.

29. X. Pitkow and M. Meister. Decorrelation and efficient coding by retinal ganglion cells. *Nat Neurosci*, 15:628–635, 2012.

30. C. P. Ratliff, B. G. Borghuis, Y.-H. Kao, P. Sterling, and V. Balasubramanian. Retina is structured to process an excess of darkness in natural scenes. *Proc Natl Sci USA*, 107:17368–17373, 2010.

bioRxiv preprint doi: https://doi.org/10.1101/131946; this version posted April 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31. D. L. Ruderman. The statistics of natural images. *Network: Comput Neural Syst*, 5:517–548, 1994.

32. P. H. Schiller. The ON and OFF channels of the visual system. *Trends Neurosci*, 15:86–92, 1992.

33. H. S. Seung and H. Sompolinsky. Simple models for reading neuronal population codes. *Proc Natl Acad Sci USA*, 90:10749–10753, 1993.

34. S. Shamai. Capacity of a pulse amplitude modulated direct detection photon channel. *IEE Proc Commun. Speech Vis.*, 137:424–430, 1990.

35. E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annu Rev Neurosci*, 24:1193–1216, 2001.

36. N. C. Singh and F. E. Theunissen. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am*, 114:3394–3411, 2003.

37. E. C. Smith and M. S. Lewicki. Efficient auditory coding. *Nature*, 439:978–982, 2006.

38. R. B. Stein. The information capacity of nerve cells using a frequency code. *Biophys J*, 7:797–826, 1967.

39. Y. Tadmor and D. J. Tolhurst. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Res*, 40:3145–3157, 2000.

40. G. Tkacik, J. S. Prentice, V. Balasubramanian, and E. Schneidman. Optimal population coding by noisy spiking neurons. *Proc Natl Acad Sci USA*, 107:14419–14424, 2010.

41. M. Tsunozaki and D. M. Bautista. Mammalian somatosensory mechanotransduction. *Curr Opin Neurobiol*, 19:362–369, 2009.

42. V. J. Uzzell and E. J. Chichilnisky. Precision of spike trains in primate retinal ganglion cells. *J Neurophysiol*, 92:780–789, 2004.

43. J. H. van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J Comp Physiol A*, 171:157–170, 1992.

44. J. H. van Hateren. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res*, 37:3407–3416, 1997.

45. Z. Wang, A. A. Stocker, and D. D. Lee. Efficient neural codes that minimize L_p reconstruction error. *Neural Comput*, 28:2656–2686, 2016.

46. D. K. Warland, P. Reinagel, and M. Meister. Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol*, 78:2336–2350, 1997.

47. J. Zylberberg, J. Cafaro, M. H. Turner, E. Shea-Brown, and F. Rieke. Direction-selective circuits shape noise to ensure a precise population code. *Neuron*, 89:369–383, 2016.