Efficient Neural Codes under Metabolic Constraints

Zhuo Wang *
Department of Mathematics
University of Pennsylvania
Philadelphia, PA 19104
wangzhuo@sas.upenn.edu

Xue-Xin Wei *
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104
weixxpku@gmail.com

Alan A. Stocker
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104
astocker@sas.upenn.edu

Daniel D. Lee
Department of Electrical and System Engineering
University of Pennsylvania
Philadelphia, PA 19104
ddlee@seas.upenn.edu

Abstract

Neural codes are inevitably shaped by various kinds of biological constraints, *e.g.* noise and metabolic cost. Here we formulate a coding framework which explicitly deals with noise and the metabolic costs associated with the neural representation of information, and analytically derive the optimal neural code for monotonic response functions and arbitrary stimulus distributions. Our framework can be applied to a neuronal pool of various sizes. For a single neuron, the theory predicts a family of optimal response functions depending on the metabolic budget and noise characteristics. Interestingly, the well-known histogram equalization solution can be viewed as a special case when metabolic resources are unlimited. For a pair of neurons, our theory suggests that under more substantial metabolic constraints, ON-OFF coding is an increasingly more efficient coding scheme compared to ON-ON or OFF-OFF. For a larger neural population, the theory predicts that the optimal code should divide the neurons into an ON-pool and an OFF-pool; neurons in the same pool have similar yet non-identical response functions. These analytical results may provide a theoretical basis for the predominant segregation into ON- and OFF-cells in early visual processing areas. Overall, the theory provides a unified framework for optimal neural codes with monotonic tuning curves in the brain, and makes predictions that can be directly tested with physiological experiments.

1 Introduction

The efficient coding hypothesis [1, 2] plays a fundamental role in understanding neural codes, particularly in early sensory processing. Going beyond the original idea of redundancy reduction by Horace Barlow [2], efficient coding has become a general conceptual framework for studying optimal neural coding [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Efficient coding hypothesizes that the neural code is such that it maximizes the information conveyed about the stimulus variable. Any formulation of efficient coding necessarily relies on a set of constraints. These constraints can come in various ways as reflected by the many real world limitations neural systems are facing. For examples, noise, limited metabolic energy budgets, constraints on the shape of tuning curves, the number of neurons in the system *etc.* all limit the dynamic range and accuracy of the neural code.

---

*equal contribution

Previous studies mainly considered only a small subset of these constraints. For example, the original proposal of redundancy reduction by Barlow focused on utilizing the dynamical range of the neurons efficiently [2, 15], but did not address the problem of noise and energy consumption. Some studies explicitly dealt with the metabolic costs of the system but did not consider the constraints imposed by the limited firing rates of neurons as well as their detailed tuning properties [16, 7, 17, 18]. Histogram equalization has been proposed as the mechanism in determining the optimal tuning curve of a single neuron with monotonic response characteristics [19]. However, this result relies on restrictive assumptions of the neural noise and does not take metabolic costs into consideration. In terms of neural population coding, most previous studies have focused on bell-shaped tuning curves. Optimal neural coding for monotonic tuning curves have received only little attention [20, 21].

We developed a formulation of efficient coding that explicitly deals with multiple biologically relevant constraints, including neural noise, limited range of the neural output, and metabolic constraints. We use our formulation to study neural codes based on monotonic response characteristics that have been frequently observed in biological neural systems. We were able to derive analytical solutions for a wide range of conditions in the small noise limit. We present results for neural pools of different sizes, including the cases of a single neuron, pairs of neurons, and larger neural populations. The results are in general agreements with observed coding schemes for monotonic tuning curves. The results also provides various quantitative predictions which are readily testable with targeted physiology experiments.

2 Optimal Code for a Single Neuron

2.1 Models and Methods

We start with the simple case where a scaler stimulus s with prior $p(s)$ is encoded by a single neuron. To model the neural response for stimulus s, we denote the mean output level as $h(s)$. Such value $h(s)$ is a deterministic mapping from s and could be the mean firing rate in the context of rate coding or just the mean membrane potential. In either case, the actual response r is noisy and has a probabilistic distribution $P(r|h(s))$. Throughout the paper, we constrain ourself to neural codes with monotonic response functions. The mutual information between the neural response r and the input stimulus is denoted as $\mathrm{MI}(s, r)$.

We formulate the efficient coding problem as the neural code seeks to maximize the mutual information between the stimulus and the response, *e.g.*, $\mathrm{MI}(s, r)$ [3]. To complete the formulation of this problem, it is crucial to choose a set of constraints which characterizes the limited resource available to the neural system. One constraint is the finite range of the neural output [19]. Another plausible constraint is on the mean metabolic cost [16, 7, 17, 18], which limits the mean activity level of neural output. Under these constraints, the efficient coding problem is mathematically formulated as following:

$$
\begin{aligned}
\text{maximize} \quad & \mathrm{MI}(s, r) \\
\text{subject to} \quad & 0 \le h(s) \le r_{\max}, \quad h'(s) \ge 0 && \text{(range constraint)} \\
& \mathbf{E}_s[K(h(s))] \le K_{\text{total}} && \text{(metabolic constraint)}
\end{aligned}
$$

We seek the optimal response function $h(s)$ under various choices of the neural noise model $P(r|h(s))$ and certain metabolic cost function $K(h(s))$, as discussed below.

Neural Noise Models: Neural noise in early sensory area can often be well characterized by a Poisson distribution [22, 23]. Under the Poisson noise model, the number of spikes N_T over a duration of T is a Poisson random variable with mean $h(s)T$ and variance $h(s)T$. In the long T limit, the mean response $r = N_T/T$ approximately follows a Gaussian distribution

$$
r \sim \mathcal{N}(h(s), h(s)/T) \tag{1}
$$

Non-Poisson noise have also been observed where the variance of response N_T can be greater or smaller than the mean firing rate [22, 23, 24]. Therefore we consider a more generalized family of noise models parametrized by α

$$
r \sim \mathcal{N}(h(s), h(s)^\alpha/T) \tag{2}
$$

This generalized family of noise model naturally includes the additive Gaussian noise case (when $\alpha = 0$), which is useful to describe the stochasticity of the membrane potential.

Metabolic Cost: The metabolic cost K is a function of the neural output

$$K(h(s)) = h(s)^\beta \tag{3}$$

where $\beta > 0$ is a parameter to model how does the energy cost scale up as the neural output is increasing. For a single neuron we will demonstrate with the general energy cost function but when we generalize to the case of multiple neurons, we will use a linear model suggested by [25] for clarity

$$K(h(s)) = K_0 + K_1 h(s) \tag{4}$$

In the context of rate coding, $K_0 = K(0)$ can be understood as the energy cost per unit time to maintain a resting neuron and K_1 is the energy cost for each extra spike per unit time. Because the metabolic constraint is also linear in $K(h(s))$, this is equivalent to the above cost function with $\beta = 1$ and properly adjusted K_{total}.

2.2 Derivation of the Optimal $h(s)$

This efficient coding problem can be greatly simplified thanks to the fact that it is invariant under any re-parameterization of the stimulus variable s (see supplementary materials for details). We take the advantage by mapping s to a uniform random variable $u = F(s) \in [0, 1]$ via the cumulative distribution function $F(s)$ [26]. If we define $g(u) = g(F(s)) = h(s)$, it suffices to solve the following new problem which optimizes $g(u)$ for a re-parameterized input u with uniform prior.

$$
\begin{aligned}
\text{maximize} \quad & \text{MI}(u, r) \\
\text{subject to} \quad & 0 \leq g(u) \leq r_{\max}, \quad g'(u) \geq 0 \\
& \mathbf{E}_u[K(g(u))] \leq K_{\text{total}}
\end{aligned}
$$

Once the optimal form of $g(u)$ is obtained, the optimal $h(s)$ is naturally given by $g(F(s))$. To solve this simplified problem, first we express the objective function in terms of $g(u)$. In the small noise limit (large integration time T), the *Fisher information* $I_F(u)$ of the neuron with noise model in Eq. (2) is calculated and the mutual information can be approximated as ([27, 14]; see supplementary notes for details)

$$I_F(u) = T\frac{g'(u)^2}{g(u)^\alpha} + O(1) \tag{5}$$

$$\text{MI}(u, r) = H(U) + \frac{1}{2}\int p(u) \log I_F(u) \, du = \frac{1}{2}\int_0^1 \log \frac{g'(u)^2}{g(u)^\alpha} \, du + \frac{1}{2}\log T + O(1/T) \tag{6}$$

where $H(U) = 0$ is the entropy and $p(u) = 1_{\{0 \leq u \leq 1\}}$ is the density of the uniform distribution. Furthermore, each constraints can be rewritten as integrals of $g'(u)$ and $g(u)$ respectively:

$$g(1) - g(0) = \int_0^1 g'(u) \, du \leq r_{\max} \tag{7}$$

$$\mathbf{E}_u[K(g(u))] = \int_0^1 g(u)^\beta \, du \leq K_{\text{total}} \tag{8}$$

This problem can be analytically solved by using the Lagrangian multiplier approach (details see the supplementary proof) and the optimal response function must take the form

$$g(u) = \left[\frac{1}{a}\gamma_q^{-1}\left(u\gamma_q(b)\right)\right]^{1/\beta}, \quad h(s) = g(F(s)) \tag{9}$$

$$\text{where } q = (1 - \alpha/2)/\beta, \quad \gamma_q(x) = \int_0^x z^{q-1}\exp(-z) \, dz. \tag{10}$$

The function $\gamma(q)$ is called the incomplete gamma function and γ_q^{-1} is its inverse function. The constants a, b can be determined by equalizing both the range and metabolic constraints. Here we give the more intuitive conclusion while leaving the detailed proof in the supplementary materials. Depending on the relative magnitude of $r_{\max}$ and K_{total}:

- **Range constraint dominates:** This is the case when there is more than sufficient energy to achieve the optimal solution $K_{\text{total}} \geq K_{\text{thre}}$. There is a threshold value K_{thre} beyond which

the metabolic constraint will become non-binding. The exact value of K_{thre} depends on the model parameters $r_{\max}$, α and β. As a loose estimation, if $K_{\text{total}} \geq r_{\max}^{\beta}$, the metabolic constraint is automatically satisfied for any α. In this case:

$$b \to 0^{+}, \quad a = b/r_{\max}^{\beta}, \quad g(u) = r_{\max} \cdot u^{1/q} \qquad (11)$$

- **Both constraints:** This is the general case when K_{total} is about the same magnitude as $r_{\max}$. We choose $a = b/r_{\max}^{\beta}$ to satisfy the range constraint and b is set to the minimum value for which the metabolic constraint is satisfied.

- **Metabolic constraint dominates:** This happens when $K_{\text{total}} \ll r_{\max}^{\beta}$. In this case we choose $a = b/r_{\max}^{\beta}$ and b is often very large.
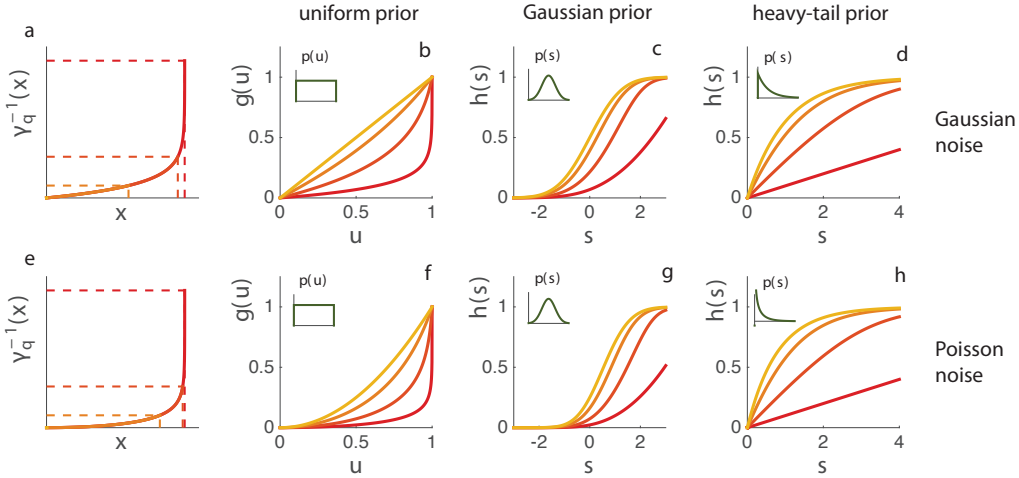

Figure 1: The process of determining optimal tuning curves $g(u)$ and corresponding $h(s)$ for different prior distributions and different noise models (top row: constant Gaussian noise $\alpha = 0$; bottom row: Poisson noise $\alpha = 1$). (a) A segment of the inverse incomplete gamma function is taken depending on the constraints. The higher the horizontal dash lines, the more substantial the metabolic constraint is. (b) The optimal $g(u)$ is determined for uniform variable u. (c) The corresponding optimal $h(s)$ for Gaussian prior. (d) The corresponding optimal $h(s)$ for Gamma distribution $p(s) \propto s^{q-1} \exp(-s)$. Specifically, in the absence of maximum response constraint and assuming the input follows this heavy tail distribution, the optimal tuning curve is exactly linear. (e-h) Similar to (a-d), but for Poisson noise.

2.3 Properties of the Optimal $h(s)$

We have predicted the optimal response function for arbitrary values of α (which constrain the noise) and β (which quantifies the cost). Here we specifically focus on a few biologically most relevant situations.

Additive gaussian noise We begin with the simple additive Gaussian noise model, i.e. $\alpha = 0$. This model could provide a good characterization of the response mapping from the input stimulus to the membrane potential of a neuron [19]. With more than sufficient metabolic supply, the optimal solution falls back to the principle of histogram equalization (see Figure 1b, yellow straight line). With less and less available metabolic budget, the optimal tuning curve bends downwards to satisfy this constraint. In general, the optimal solution strikes a balance between the the constraints, resulting in a family of optimal response functions in between of the two extrema mentioned above.

Poisson noise For neural spiking activity, it is observed that the variability often varies systematically with the mean firing rate [22, 23]. In the case of Poisson spiking, the theory predicts the optimal response function should bend more downwards compared to the case of Gaussian noise (see Figure 1). In the extreme case when the main resource constraint comes from the limit firing rate range, the model predicts a square tuning curve for uniform input (Figure 1f, yellow curve), which is consistent with early studies [28, 29].

2.4 Distribution of the response magnitude

In the case of Gaussian noise and K_{total} is large, the response magnitude is equally distributed in the response range. This is consistent with the histogram equalization solution which uses the response range equally well. However, as the metabolic constraint plays an increasingly important role when K_{total} is diminishing, the large response will be penalized more severely, resulting in more density at small response magnitude. We also found that Poisson noise leads to more penalization on large response magnitude compared to Gaussian noise, suggesting an interplay between noise and metabolic cost in shaping the optimal neural response distribution. Furthermore, in the case that K_{total} goes to 0, the response distribution converges to a gamma distribution, with heavy tail. This phenomenon represents the sparse codes [7]. It also gives a simple yet quantitative characterization of how the energy budget may push the neural responses toward a sparse coding regime.
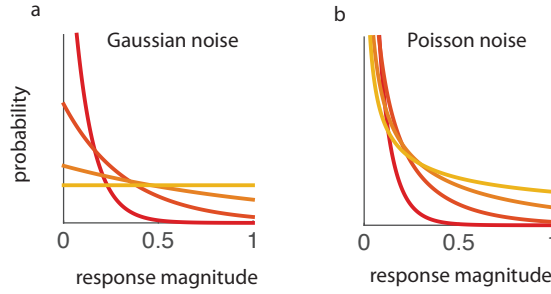

Figure 2: Distribution of the response based on the optimal response function of a single neuron. (a) Gaussian noise. (b) Poisson noise. In the extreme case of Gaussian noise with effectively no metabolic constraint, the distribution is uniformly distributed on the whole range.

3 Optimal Code for a Pair of Neurons

We next study the optimal coding in the case of two neurons with monotonic response functions. We denote the neural responses as $\mathbf{r} = (r_1, r_2)$. Therefore the efficient coding problem becomes:

$$
\begin{aligned}
\text{maximize} \quad & L(\mathbf{h}) = \text{MI}(s, \mathbf{r}) \\
\text{subject to} \quad & 0 \le h_i(s) \le r_{\max}, \quad i = 1, 2. \qquad \text{(range constraint)} \\
& \mathbf{E}_s \left[K(h_1(s)) + K(h_2(s)) \right] \le 2K_{\text{total}} \qquad \text{(metabolic constraint)}
\end{aligned}
$$

Assuming the neural noise is independent across neurons, the system of two neurons has total Fisher information just as the linear sum of Fisher information contributed from each neuron $I_F(s) = I_1(s) + I_2(s)$.

3.1 Optimal response functions

Previous studies on neural coding with monotonic response functions have typically assumed that $h_i(s)$ has sigmoidal shape. It is important to emphasize that we do not make any a priori assumptions on the detailed shape of the tuning curve other than being monotonic and smooth. We define each neuron's active region $A_i = A_i^+ \cup A_i^-$ where $A_i^\pm = \{s | \pm h_i'(s) > 0\}$. Without going into detailed proof (see supplementary materials, section B), we list the main conclusions

1. Neurons should have non-overlapping active regions $A_i \cap A_j = \emptyset$ if $i \ne j$.

2. If the metabolic constraint is binding, ON-OFF coding (A_1^+, A_2^- are non-empty or vice versa) is better than ON-ON coding (A_i^+'s are non-empty) or OFF-OFF coding (A_i^-'s are non-empty). Otherwise all three coding schemes can achieve the same mutual information.

3. For ON-OFF coding, it is better to have ON regions on the right side: $\sup A_i^- \le \inf A_j^+$.

4. For ON-ON coding (or OFF-OFF), each neuron should have roughly the same tuning curve $h_i(s) \approx h_j(s)$ while still have disjoint active regions. Within the ON-pool or OFF-pool, the optimal tuning curve is same as the optimal solution from the single neuron case.

In Figure 3 we illustrate how these conclusions can be used to determine the optimal pair of neurons, assuming additive Gaussian noise $\alpha = 0$ and linear metabolic cost $\beta = 1$ (for other α, β, the process is similar). Crucially, our theory allows us to predict the precise shape of the optimal response functions. This represents a significant advantage over the previous studies which also investigated ON-OFF coding scheme [13, 30].

3.2 Comparison between ON-OFF and ON-ON codes

From Figure 3e we can see that, the maximum possible mutual information is monotonically increasing as a function of available energy K_{total} until they both saturate the limit at $K_{\text{ON-ON}} = 0.5r_{\text{max}}$ and $K_{\text{ON-OFF}} = 0.25r_{\text{max}}$ respectively (see the yellow tuning curves in Figure 3a-d). Note that this specific saturation limit is only valid for $\alpha = 0$ and $\beta = 1$. In order to encode exactly the same amount information, the most energy efficient ON-ON pair (or OFF-OFF) always requires twice as much compared to the most energy efficient ON-OFF pair.

On the other hand, we can compare the ON-ON and ON-OFF by fixing a value of $K_{\text{total}} < 0.5r_{\text{max}}$ (i.e. when metabolic constraint is binding for ON-ON pairs). The optimal mutual information achieved by ON-ON neurons is always smaller than that achieved by ON-OFF neurons and the difference is plotted. If in the mutual information we use logarithm of base 2, this difference will saturate at -1 when the available energy is very limited $K_{\text{total}} \ll r_{\text{max}}$. In this extreme case, the ON-ON code is only half as efficient as the ON-OFF code. In other words, it takes as much as twice amount of time T for the ON-ON code to achieve same amount of mutual information as the ON-OFF code.
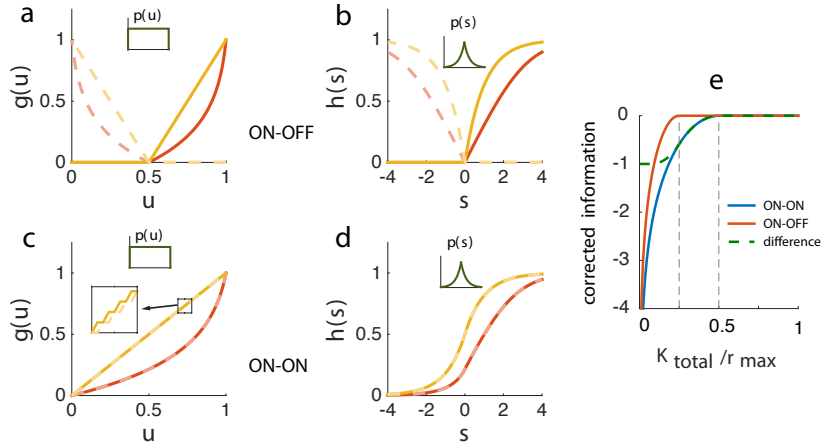

Figure 3: The optimal response functions for a pair of neurons assuming Gaussian noise. (a) The optimal response functions for a uniform input distribution assuming ON-OFF coding scheme. Solid yellow curve and dash yellow curve represent the optimal solution with weak metabolic constraint. Solid red and dash red curves are the optimal solution with substantial metabolic constraint. (b) Similar to panel a, but for input stimuli with heavy tail distribution. (c) The optimal response functions for a uniform input distribution assuming ON-ON coding scheme. Solid and dash yellow curves are for little metabolic constraint. Notice that two curves appear to be identical but are actually different at finer scales (see the inserted panel). Solid and dash red are for substantial metabolic constraint. (d) Similar to panel c, but for input stimuli with heavy tail distribution. (e) A comparison between the ON-ON scheme and ON-OFF scheme. The x-axis represents the relative importance of metabolic constraint. The y-axis represents the corrected information, defined as the amount of information actually transmitted minus the maximal information that can possibly be transmitted. The green dash line represent the difference between the information transmitted by the two schemes. Negative difference indicates an advantage of ON-OFF over ON-ON.

Our analysis provides a quantitative characterization of the advantage of ON-OFF over ON-ON and shows how it depends on the relative importance of the metabolic constraint. The encoding efficiency of ON-OFF ranges from double (with very limited metabolic budget) to equal amount of the ON-ON efficiency (with unlimited metabolic budget). This wide range includes the previous

conclusions drawn by Gjorgjieva and colleagues, who only found a mild increase ($\sim 15\%$ at most) in the efficiency when comparing ON-OFF to ON-ON under short integration time limit [30]. It is well known that in the retina of many animal species, there is a split of ON and OFF pathways [31, 32]. The substantial increase of efficiency in the regime of strong metabolic constraint supports the idea that metabolic constraint may be one of the main drives for such pathway splitting in evolution.

In a recent study by Karklin and Simoncelli [13], it is observed numerically that training a simple linear-nonlinear network on natural images by maximizing mutual information subject to metabolic constraint would lead to ON-OFF coding scheme in certain noise regime. Our result may provide a theoretical bases for this observation, although we do not directly model the natural image, rather the neurons can been seen as encoding the local contrast in this context. Intriguingly, we found that in the case that the input distribution is a heavy tail distribution (see Figure 3b), the optimal response functions are two rectified non-linear functions who split the encoding range, which is the kind of non-linearity that has been observed physiologically in retina.

4   Optimal coding of large neural population

The framework could be generalized to study a large population of neurons ($N = 2k$, k is large). In this case, we consider the following problem:

$$
\begin{aligned}
\text{maximize} \quad & \text{MI}(s, \mathbf{r}) \\
\text{subject to} \quad & r_{\min} \leq h_i(s) \leq r_{\max} && \text{(range constraint)} \\
& \mathbf{E}_s[\sum_i K(h_i(s))] \leq N K_{\text{total}} && \text{(metabolic constraint)}
\end{aligned}
$$

We can again solve this problem analytically by exploiting the Fisher information approximation of mutual information [27, 14]. Interestingly, we found the optimal codes should be divided into two pools of neurons of equal size k. One pool of neuron with monotonic increasing response function (ON-pool), and the other with monotonic decreasing response function (OFF-pool). For neurons within the same pool, the optimal response functions appear to be identical on the macro-scale but are quite different when zoomed in. We have shown that the optimal code must have disjoint active regions for each neuron. This is illustrated in the inset panel of Figure 3c, in which we show the case for two ON seemingly identical tuning curves.

We ask how the energy should be allocated across different neurons. Assume that metabolic cost is linear in terms of the response level and Poisson noise, each neuron across two different pools should share the same maximum firing rate. This generalizes to other noise type with considered ($\alpha > 0$) and other metabolic cost function ($\beta > 0$).

We quantify the amount of the information increase by using optimal coding schemes compared to using all ON neurons or all OFF neurons. Interestingly, the results we found in the Fig 3e for the a pair of neurons generalize to the current case. Specifically, in the case of strong metabolic constraint (*i.e.*, K_{total} is small), the optimal $2k$-ON neuron scheme is close to half of the efficiency of the optimal k-ON/k-OFF scheme.

The optimal coding scheme is reminiscent of the opponent coding observed in some neural systems, for example, the sound location system [33]. In our results the support of the response function of an ON-neuron does not overlap with that of an OFF-neuron. We notice that in the physiological data [33], there appears to be some overlap between two neuron which belong to different pools. However, in the case that there is noise in the input, it is possible that some amount of the overlap might be beneficial.

5   Discussion

We presented a theoretical framework for studying optimal neural codes under biologically relevant constraints. We emphasized the importance of two constraints – the noise characteristics of the neural responses and the metabolic cost. Throughout the paper, we have focused on neural codes with smooth monotonic response functions. We demonstrated that, maybe surprisingly, analytical solutions exist for a wide family of noise characteristics and metabolic cost functions.

An interesting venue for future research is to see whether the framework and techniques developed here could be used to define the optimal neural codes based on bell-shape tuning curves. Another interesting question is the optimal code in case of an odd number of neurons. Presumably, the solution for the case of $N = 2k + 1$ is close to $N = 2k$ for a large pool of nuerons. However, when k is small, the difference due to symmetry breaking may substantially change the result. We have not addressed these results due to the lack of biological relevance for this case. Also, we have only considered the case of maximizing mutual information as the objective function; it will be interesting to see whether the results generalized to other objective functions such as, *e.g.*, minimizing decoding error[34, 35].

Due to the limited scope of the paper, we have ignored several important other factors when formulating the efficient coding problem. First, we have not modeled the spontaneous activity (baseline firing rate) of neurons. Second, we have not considered the noise correlations between the responses of neurons. Third, we have ignored the noise in the input to the neurons. We think that the first two factors are unlikely to significantly change our main results. However, incorporating the input noise may significantly change the results. In particular, for the cases of multiple neurons, our current results suggest that the response functions for ON and OFF neurons should not overlap. However, it is possible that this prediction does not hold in the presence of the input noise. Intuitively, introducing some redundancy by making the response functions partially overlap might be beneficial in this case. Including these factors into the framework should allow us to make a detailed and quantitative comparison to physiologically measured data in the future.

References

[1] Fred Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183, 1954.

[2] Horace B Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234, 1961.

[3] Ralph Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, 1988.

[4] Joseph J Atick and A Norman Redlich. Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320, 1990.

[5] Joseph J Atick. Could information theory provide an ecological theory of sensory processing? *Network: Computation in neural systems*, 3(2):213–251, 1992.

[6] F Rieke, DA Bodnar, and W Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 262(1365):259–265, 1995.

[7] Bruno Olshausen and David Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[8] Anthony J Bell and Terrence J Sejnowski. The "independent components" of natural scenes are edge filters. *Vision research*, 37(23):3327–3338, 1997.

[9] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.

[10] Allan Gottschalk. Derivation of the visual contrast response function by maximizing information rate. *Neural computation*, 14(3):527–542, 2002.

[11] Nicol S Harper and David McAlpine. Optimal neural population coding of an auditory spatial cue. *Nature*, 430(7000):682–686, 2004.

[12] Mark D McDonnell and Nigel G Stocks. Maximally informative stimuli and tuning curves for sigmoidal rate-coding neurons and populations. *Physical review letters*, 101(5):058103, 2008.

[13] Yan Karklin and Eero P Simoncelli. Efficient coding of natural images with a population of noisy linear-nonlinear neurons. *Advances in neural information processing systems*, 24:999, 2011.

[14] Xue-Xin Wei and Alan A Stocker. Mutual information, fisher information, and efficient coding. *Neural computation*, 2016.

[15] Horace Barlow. Redundancy reduction revisited. *Network: computation in neural systems*, 12(3):241–253, 2001.

[16] William B Levy and Robert A Baxter. Energy efficient neural codes. *Neural computation*, 8(3):531–543, 1996.

[17] Simon B Laughlin, Rob R de Ruyter van Steveninck, and John C Anderson. The metabolic cost of neural information. *Nature neuroscience*, 1(1):36–41, 1998.

[18] Vijay Balasubramanian, Don Kimber, and Michael J Berry II. Metabolically efficient information processing. *Neural Computation*, 13(4):799–815, 2001.

[19] Simon B Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch*, 36(910-912):51, 1981.

[20] Deep Ganguli and Eero P Simoncelli. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Computation*, 26(10):2103–2134, 2014.

[21] David B Kastner, Stephen A Baccus, and Tatyana O Sharpee. Critical and maximally informative encoding between neural populations in the retina. *Proceedings of the National Academy of Sciences*, 112(8):2533–2538, 2015.

[22] George J Tomko and Donald R Crapper. Neuronal variability: non-stationary responses to identical visual stimuli. *Brain research*, 79(3):405–418, 1974.

[23] DJ Tolhurst, JA Movshon, and ID Thompson. The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental brain research*, 41(3-4):414–419, 1981.

[24] Mark M Churchland et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378, 2010.

[25] David Attwell and Simon B Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145, 2001.

[26] Xue-Xin Wei and Alan A Stocker. A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts. *Nature Neuroscience*, 2015.

[27] Nicolas Brunel and Jean-Pierre Nadal. Mutual information, Fisher information, and population coding. *Neural Computation*, 10(7):1731–1757, 1998.

[28] Matthias Bethge, David Rotermund, and Klaus Pawelzik. Optimal short-term population coding: when fisher information fails. *Neural Computation*, 14(10):2317–2351, 2002.

[29] Don H Johnson and Will Ray. Optimal stimulus coding by neural populations using rate codes. *Journal of computational neuroscience*, 16(2):129–138, 2004.

[30] Julijana Gjorgjieva, Haim Sompolinsky, and Markus Meister. Benefits of pathway splitting in sensory coding. *The Journal of Neuroscience*, 34(36):12127–12144, 2014.

[31] Peter H Schiller. The on and off channels of the visual system. *Trends in neurosciences*, 15(3):86–92, 1992.

[32] Heinz Wässle. Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, 5(10):747–757, 2004.

[33] G Christopher Stecker, Ian A Harrington, and John C Middlebrooks. Location coding by opponent neural populations in the auditory cortex. *PLoS Biology*, 3(3):e78, 2005.

[34] Tvd Twer and Donald IA MacLeod. Optimal nonlinear codes for the perception of natural colours. *Network: Computation in Neural Systems*, 12(3):395–407, 2001.

[35] Zhuo Wang, Alan A Stocker, and Daniel D Lee. Optimal neural tuning curves for arbitrary stimulus distributions: Discrimax, infomax and minimum L_p loss. In *Advances in Neural Information Processing Systems NIPS*, pages 2177–2185, 2012.