PHYSICAL REVIEW RESEARCH **5**, 043205 (2023)

Efficient population coding of sensory stimuli

Shuai Shao,[1,2] Markus Meister,[3] and Julijana Gjorgjieva[4,1,*]
[1]*Computation in Neural Circuits Group, Max Planck Institute for Brain Research, 60438 Frankfurt, Germany*
[2]*Donders Institute and Faculty of Science, Radboud University, 6525 GD, Nijmegen, Netherlands*
[3]*Division of Biology and Biological Engineering, California Institute of Technology, 91125 Pasadena, California, USA*
[4]*School of Life Sciences, Technical University of Munich, 85354 Freising, Germany*

(Received 24 July 2022; accepted 19 October 2023; published 5 December 2023)

This article is part of the Physical Review Research *collection titled* Physics of Neuroscience.

The efficient coding theory postulates that single cells in a neuronal population should be optimally configured to efficiently encode information about a stimulus subject to biophysical constraints. This poses the question of how multiple neurons that together represent a common stimulus should optimize their activation functions to provide the optimal stimulus encoding. Previous theoretical approaches have solved this problem with binary neurons that have a step activation function, and have assumed that spike generation is noisy and follows a Poisson process. Here we derive a general theory of optimal population coding with neuronal activation functions of any shape, different types of noise and heterogeneous firing rates of the neurons by maximizing the Shannon mutual information between a stimulus and the neuronal spiking output subject to a constraint on the maximal firing rate. We find that the optimal activation functions are discrete in the biological case of non-negligible noise and demonstrate that the information does not depend on how the population is divided into ON and OFF cells described by monotonically increasing vs decreasing activation functions, respectively. However, the population with an equal number of ON and OFF cells has the lowest mean firing rate, and hence encodes the highest information per spike. These results are independent of the shape of the activation functions and the nature of the spiking noise. Finally, we derive a relationship for how these activation functions should be distributed in stimulus space as a function of the neurons' firing rates.

DOI: 10.1103/PhysRevResearch.5.043205

I. INTRODUCTION

In many neuronal systems, sensory information is processed by multiple neurons in parallel, forming a population code. However, how a population of neurons works together to efficiently encode a sensory stimulus in the presence of different biological constraints is still an open question. Many experimental and theoretical studies have proposed that neuronal coding is optimal [1–5]. Determining optimality is typically considered in the context of various constraints provided by the biological system in question. These include various assumptions made about the structure of the neuronal population, the relationship between stimulus and neuronal firing, the source and magnitude of sensory noise, and different measures used to quantify coding efficiency. For example, a common way to describe the firing rate of a neuron as a function of the stimulus is through an activation function, which usually describes a nonlinear dependence determined by the various ion channels embedded in the neuron's membrane or elaborate dendrites morphologies [6,7]. The activation functions of sensory neurons can be monotonically increasing or decreasing as a function of the stimulus, referred to as ON or OFF, respectively [Fig. 1(a)], although in some sensory systems ON-OFF cells with nonmonotonic activation functions also exist [8,9]. ON and OFF cells are found in many sensory systems, including the retina where ON (OFF) ganglion cells code for increases (decreases) in visual stimulus intensity or contrast [10,11] and the insect mechanosensory system where they code for increases and decreases in leg angle [12]. In line with most optimal coding theories of neuronal populations, here we assume that multiple cells together encode a sensory stimulus more efficiently than single cells in the presence of sensory noise and biophysical constraints.

Populations of sensory neurons are typically affected by noise, which can come from different sources including from the sensory environment and biophysical constraints. Assuming a description of neuronal firing by activation functions, noise can enter before or after the activation function, called input vs output noise, respectively, and can have a different influence of stimulus coding [13,14]. Since neurons communicate via action potentials, theoretical studies of optimal coding have commonly assumed that individual neurons generate spike counts in fixed coding windows following Poisson statistics [15–18]. Under conditions of low spike count

---

[*] gjorgjieva@tum.de

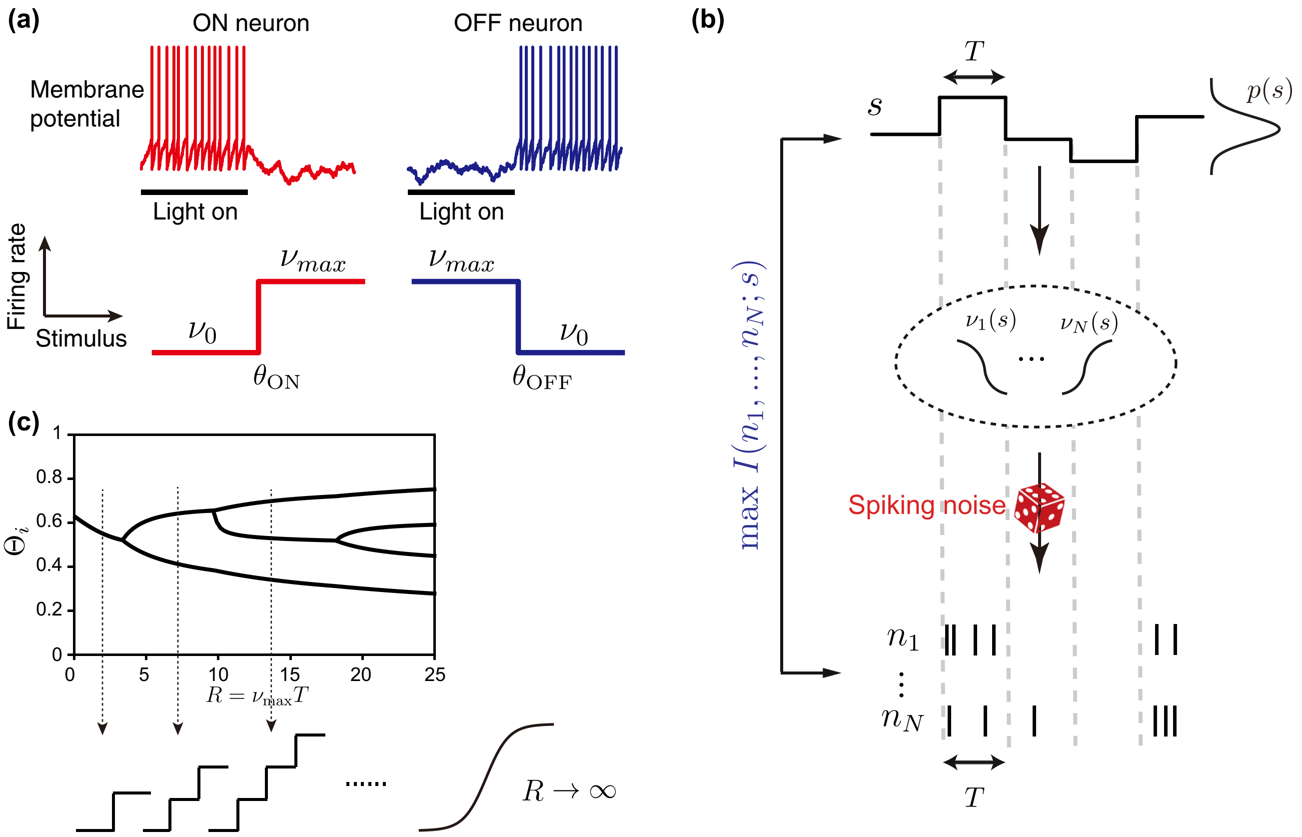

FIG. 1. Efficient coding framework of a population of ON and OFF neurons. (a) A schematic of ON and OFF neurons. An ON neuron fires more frequently when the stimulus (which is light in this example) is high and fires at the spontaneous rate (here 0) when the stimulus is absent. The opposite is true for an OFF neuron. (b) The population coding model. Sensory stimuli s, which are constant in the coding windows of size T, are drawn from a distribution $p(s)$. The stimuli are encoded by a population of neurons with firing rates $\nu_i(s)$, which fire noisy spike trains, n_i. The distribution of n_i is given by the conditional probability $p(n_i|\nu_i(s))$, which denotes the spiking noise. The efficiency of the neuronal coding is quantified by the Shannon mutual information between the stimuli s and the spike trains n_i, i.e., $I(n_1, \ldots, n_N; s)$. (c) The optimal activation function, which maximizes the mutual information for a single ON neuron is discrete. (Upper) The optimal thresholds of a single neuron that maximizes the Shannon mutual information. Θ_i denotes the cumulative probability of s above a threshold. (Lower) Schematics depicting that the number of steps of the optimal activation function increases with the product of the maximal firing rate and the coding window, i.e., $R = \nu_{\max} T$. For low R, the optimal activation function is binary and has one threshold ($i = 1$). As R increases, the optimal activation function become ternary ($i = 2$), etc. The activation function becomes continuous in the limit of $R \to \infty$.

intensity of the Poisson process, the optimal activation functions of single neurons can be proven to be discrete with a single step, i.e., binary [15,16]. However, when the spike count intensity increases, binary neurons are no longer optimal, but rather the number of steps in the activation function increases as a function of spike count intensity [16]. Especially in biological systems, many of these assumptions need to be relaxed. First, activation functions in different sensory systems usually do not manifest as binary and may appear continuous due to the presence of noise [14,19,20]. Neuronal spike counts can also be non-Poisson, for instance, in the retina [21,22]. Therefore, it is an interesting question what optimal configuration of activation functions can be achieved in theoretical frameworks of efficient coding where spike counts follow statistics other than Poisson.

What quantity might neural populations optimize? Two measures have been commonly used [17,23–27]. The Shannon mutual information between the stimulus and neuronal responses does not assume how the information should be decoded downstream. Alternatively, the stimulus can be estimated using a decoder and the difference between the stimulus and the estimate can be minimized. These two measures can generate very different predictions about the optimal population coding strategy [18,28].

Here, we develop a general efficient coding theory based on a population coding model with multiple ON and OFF neurons that code for a scalar stimulus from a given distribution assuming any (monotonic) nonlinear activation function and any noise statistics. We use the Shannon mutual information between the stimulus and the neuronal spikes as a measure of coding efficiency, and discover that this measure is independent of how the population is divided into ON and OFF neurons. We also investigate how the optimal firing thresholds of ON and OFF neurons partition the stimulus space as a function of the maximal neuronal firing rates. When these firing rates are equal for all neurons, we find that the thresholds divide the stimulus distribution into surprisingly regular stimulus regions.

II. THEORETICAL FRAMEWORK

We propose a theoretical framework of population coding with the following assumptions [Fig. 1(a)]:

(i) A population of ON and OFF neurons code for a one-dimensional stimulus, with monotonically increasing and decreasing firing rates as a function of the stimulus (respectively), called activation functions;

(ii) Each neuron i in the population has a minimum (spontaneous) firing rate ν_0 usually assumed to be 0, and a maximal firing rate constraint $\nu_{\max,i}$;

(iii) The dynamic range of each neuron i, defined as the stimulus that leads to nonzero and nonmaximal firing rate ν_i (with $\nu_0 < \nu_i < \nu_{\max,i}$), does not overlap with the dynamic range of other neurons;

(iv) The dynamic ranges of OFF neurons are lower than those of ON neurons.

For the second assumption, we start with a simple case in which the maximal firing rates in a population are identical across the cells, i.e., $\nu_{\max,i} = \nu_{\max}$. Later in this paper (Secs. III D and III H) we also consider neuron populations with heterogeneous $\nu_{\max,i}$. The assumption of zero spontaneous firing rate ensures analytical tractability. Our conclusions hold, at least in the case of binary activation functions for all cells with Poisson noise, even if this assumption is relaxed [18].

We denote the sensory stimulus to be encoded by a population of N cells as the scalar s, which is drawn from a distribution $p(s)$. We denote the activation function of each neuron as $\nu_i(s)$, where the subscript i is the index of neurons in the population. We define "the coding window" T as the time period when the stimulus s is constant [Fig. 1(b)]. The coding window depends on the neuronal dynamics in the specific sensory population. For instance, in the mammalian retina, retinal ganglion cells have a coding window of 10 to 50 ms [17,29,30]. In the mouse auditory system, auditory nerve fibers, have a coding window of 50 ms [14,19]. Defining a coding window allows us to define the spike count n_i for neuron i within a coding window T, which has an expected value of $\nu_i(s)T$. Therefore, the stimulus s is encoded by a vector of noisy spike counts $\vec{n} = \{n_1, \dots, n_N\}$, which represents the population code.

We consider a general noise model where the spike counts follow a probability distribution $p(\vec{n}|s)$, which only directly depends on the expected value $\vec{\nu}(s)T$. Since the firing rate vector $\vec{\nu}$ is a deterministic function of the stimulus s, and assuming the noise of different neurons is independent of each other, the probability distribution $p(\vec{n}|s)$ can also be written as a product of the spike count probability distribution of every neuron, i.e., $p(\vec{n}|\vec{\nu}(s)) = \prod_i p(n_i|\nu_i(s))$. Because ν_i is the firing rate and $\nu_i T$ is the expected value of the spike count n_i of neuron i, by definition, we have

$$\sum_{n_i=0}^{+\infty} p(n_i|\nu_i) = 1, \tag{1}$$

$$\sum_{n_i=0}^{+\infty} p(n_i|\nu_i) n_i = \nu_i T. \tag{2}$$

While the noise can follow any distribution, a special case commonly used in previous study is the Poisson noise where $p(\vec{n}|s) = p(\vec{n}|\vec{\nu}(s)) = \prod_i \frac{(\nu_i(s)T)^{n_i}}{n_i!} e^{-\nu_i(s)T}$. We quantify the coding efficiency of this population code using the Shannon mutual information between the population spike count $\vec{n}$ and stimulus s,

$$I(s, \vec{n}) = \sum_{\vec{n}} \int ds\, p(s) p(\vec{n}|s) \log \frac{p(\vec{n}|s)}{P(\vec{n})} \tag{3}$$

where

$$P(\vec{n}) = \int ds\, p(s) \prod_i p(n_i|\nu_i), \tag{4}$$

and $\sum_{\vec{n}} = \sum_{n_1=0}^{+\infty} \dots \sum_{n_N=0}^{+\infty}$ denotes the sum over all possible spike counts of all the neurons.

Because the firing rates $\vec{\nu}$ depend deterministically on the stimulus s, the mutual information between s and $\vec{n}$ is the same as the mutual information between $\vec{\nu}$ and $\vec{n}$ (see Supplemental Material, SM [31]),

$$I(s, \vec{n}) = I(\vec{\nu}, \vec{n}) = \sum_{\vec{n}} \int_{\vec{\nu}} d^N\vec{\nu}\, p(\vec{\nu}) p(\vec{n}|\vec{\nu}) \log \frac{p(\vec{n}|\vec{\nu})}{P(\vec{n})}. \tag{5}$$

III. RESULTS

We seek to derive the optimal activation functions $\{\nu_i(\cdot)\}_i$ of an entire population of ON and OFF neurons, which maximize the mutual information $I(s, \vec{n})$ [Eq. (5)], when the conditional probability $p(n_i|\nu_i)$ is given. We also aim to determine how this maximal mutual information depends on the ON-OFF composition of the neuronal population.

A. The optimal activation function for a single neuron is discrete

We first investigate a population with only a single neuron subject to the constraints from Sec. II. Previous studies have found that under these conditions and with Poisson-distributed spike counts, the optimal activation function for a single neuron should be discrete, with an increasing number of steps as a function of the product $R = \nu_{\max} T$, i.e., the maximum expected spike count [15,16] [Fig. 1(c)]. In two steps, we generalize this result to any analytic conditional probability $p(n|\nu)$ (analytic in terms of ν) using the fact that mutual information is convex in the input space [32].

In step 1, we prove that the mutual information $I(\nu, n)$ is distributed proportionally to the probability density $p(\nu)$ in the optimal configuration. Defining the "density of mutual information" as

$$i(\nu) = \sum_n p(n|\nu) \log \frac{p(n|\nu)}{P(n)} \tag{6}$$

we can write

$$I(\nu, n) = \int_\nu d\nu\, p(\nu)\, i(\nu). \tag{7}$$

We can then prove that in the optimal case,

$$i(\nu) = I^{\max} \text{ for all possible } \nu \tag{8}$$

where $I^{\max}$ is the maximal mutual information (see the SM [31]).

Then in step 2, we show that Eq. (8) cannot be true if the activation function $\nu(s)$ is continuous, therefore concluding that it must be discrete. To do this, we first redefine the activation function using a function F_ν. For an ON neuron (the case for an OFF neuron follows similarly), we can write for any arbitrary firing rate $\tilde{\nu}$,

$$F_\nu(\tilde{\nu}) = \int_{-\infty}^{s_{\max}(\tilde{\nu})} ds\, p(s), \tag{9}$$

where $s_{\max}(\tilde{\nu})$ is defined as the highest s that makes $\nu(s) \leqslant \tilde{\nu}$, i.e., $s_{\max} = \max\{s|\nu(s) \leqslant \tilde{\nu}\}$. Because $\nu(s)$ is a monotonically increasing function of s, $s_{\max}(\tilde{\nu})$ is also monotonically increasing, making $F_\nu(\tilde{\nu})$ a monotonically increasing function of $\tilde{\nu}$. We can replace the variable in the integral of Eq. (9), leading to

$$F_\nu(\tilde{\nu}) = \int_{\nu(s\to-\infty)}^{\nu(s=s_{\max}(\tilde{\nu}))} d\nu\, p(\nu). \tag{10}$$

Therefore, F_ν becomes the cumulative distribution function of the firing rate ν,

$$F_\nu(\tilde{\nu}) = \int_0^{\tilde{\nu}} d\nu\, p(\nu). \tag{11}$$

Let F_ν^* denote the optimal activation function, which maximizes the mutual information $I(\nu, n)$ [Eqs. (5) and (7)]. We explicitly include the dependence of the density of mutual information $i(\nu)$ [Eq. (6)] on the activation function F_ν by writing $i(\nu, F_\nu)$ because $P(n)$ depends on F_ν. Then, Eq. (8) can be rewritten as

$$i(\nu, F_\nu^*) = I(F_\nu^*) \quad \text{for all } \nu \text{ in } E_\nu^* \tag{12}$$

where E_ν^* is the set of points at which F_ν^* increases.

From now on, we denote the conditional probability $p(n_i|\nu_i)$ by $L(n_i, \nu_i T)$, and call it the "noise generation function". If we assume $L(n, \nu T)$ is analytic with respect to νT, then we can show that the optimal activation function has a finite number of steps, i.e., E_ν^* is a finite set of points. Note that because of Eq. (11), E_ν is also the set of all possible firing rates, i.e., $E_\nu = \{\nu|p(\nu) > 0\}$. If E_ν^* has a finite number of points, then the optimal $\nu(s)$ will have a finite number of steps.

Let us first consider the case that E_ν^* is infinite. In the simplest case, if F_ν^* is continuous over the interval $[0, \nu_{\max}]$, then $E_\nu^* = [0, \nu_{\max}]$. As a result, $i(\nu, F_\nu^*) = \text{const}$ for any $\nu \in [0, \nu_{\max}]$.

If F_ν^* is not continuous but E_ν^* has an infinite number of points (e.g., F_ν^* is only continuous on a subinterval of $[0, \nu_{\max}]$), similar to previous study [15,32], one can use the Bolzano Weierstrass theorem [33] to prove that E_ν^* has a limit point in $[0, \nu_{\max}]$. Then by the identity theorem for analytic functions [34], if two analytic functions, in our case $i(\nu, F_\nu^*)$ and $I(F_\nu^*)$, have the same value on an infinite number of points and the limit of these points, then they are equal, i.e., $i(\nu, F_\nu^*) = \text{const}$ for any $\nu \in [0, \nu_{\max}]$. In short, assuming E_ν^* has an infinite number of points also implies that $i(\nu, F_\nu^*)$ is a constant over the interval $[0, \nu_{\max}]$.

If E_ν^* is infinite, assuming optimal coding, based on Eq. (8), we have

$$i(\nu) = \sum_{n=0}^{+\infty} L(n, \nu T) \log \frac{L(n, \nu T)}{P(n)} = I^{\max} = \text{const}. \tag{13}$$

Then the derivative with respect to νT,

$$i'(\nu) = \sum_{n=0}^{+\infty} L'(n, \nu T) \log \frac{L(n, \nu T)}{P(n)} = 0 \tag{14}$$

where $L'(n, \nu T)$ denotes $\frac{\partial L(n,\nu T)}{\partial(\nu T)}$. Similarly, the second derivative

$$i''(\nu) = \sum_{n=0}^{+\infty} \left[L''(n, \nu T) \log \frac{L(n, \nu T)}{P(n)} + \frac{L'(n, \nu T)^2}{L(n, \nu T)} \right] = 0. \tag{15}$$

Using mathematical induction, one can prove that for any $m \in \mathbb{N}^+$, the mth derivative of $i(\nu)$ with respect to νT, $i^{(m)}(\nu)$, contains the term

$$\sum_{n=0}^{+\infty} L^{(m)}(n, \nu T) \log \frac{L(n, \nu T)}{P(n)}. \tag{16}$$

According to Eq. (2), $\sum_n L(n, \nu T) n = \nu T$, we have

$$L(0, 0) = 1, \quad L(n \geqslant 1, 0) = 0. \tag{17}$$

Based on these two boundary conditions, $L(n, \nu T)$ can be written as a Maclaurin series

$$L(0, \nu T) = 1 + \sum_{k=1}^{+\infty} a_{0k} (\nu T)^k, \tag{18}$$

$$L(n, \nu T) = \sum_{k=1}^{+\infty} a_{nk} (\nu T)^k. \tag{19}$$

Substituting these two series into the fractional or polynomial terms of the derivatives of the noise generation function $L(n, \nu T), L'(n, \nu T), \ldots, L^{(m-1)}(n, \nu T)$, and also in the terms $\sum_n L^{(m)}(n, \nu T) \log P(n)$ in the derivatives $i^{(m)}(\nu)$ [Eq. (16)], we find that they all become fractional or polynomial terms of νT after doing the Maclaurin expansion with respect to νT around 0. For example, in $i''(\nu)$ [Eq. (15)],

$$\begin{aligned} &\frac{L'(n, \nu T)^2}{L(n, \nu T)} \\ &= \frac{\left(\sum_{k=1}^{+\infty} a_{nk} k (\nu T)^{k-1}\right)^2}{\sum_{k=1}^{+\infty} a_{nk} (\nu T)^k} \\ &= \frac{\left(\sum_{k=1}^{+\infty} a_{nk} k (\nu T)^{k-1}\right)^2}{a_{n1} \nu T} \left(1 + \sum_{k=2}^{+\infty} \frac{a_{nk}}{a_{n1}} (\nu T)^{k-1}\right)^{-1} \\ &= \frac{a_{n1}}{\nu T} + 3a_{n2} + \left(5a_{n3} + \frac{a_{n2}^2}{a_{n1}}\right)\nu T + \cdots \quad (n \geqslant 1). \end{aligned} \tag{20}$$

The only exception is the term containing $\log(\nu T)$ apart from the polynomial terms,

$$\begin{aligned}\sum_{n=0}^{+\infty} L^{(m)}(n,\nu T)\log L(n,\nu T) &= \sum_{k=m}^{+\infty} a_{0k}\frac{k!}{(k-m)!}(\nu T)^{k-m}\log\left[1+\sum_{l=1}^{+\infty}a_{0l}(\nu T)^{l}\right] + \sum_{n=1}^{+\infty}\left[\sum_{k=m}^{+\infty}a_{nk}\frac{k!}{(k-m)!}(\nu T)^{k-m}\right]\\ &\quad\times\left[\log a_{n,j(n)} + j(n)\log(\nu T) + \log\left(1+\sum_{l>j(n)}\frac{a_{nl}}{a_{nj}}(\nu T)^{l-j(n)}\right)\right]\end{aligned} \tag{21}$$

where $j(n)$ is the minimal index of k that makes $a_{nk} > 0$ when n is given. When $\nu T \to 0$, we can see that the first term in Eq. (21) is finite. The second term can be expanded as the sum of polynomial terms and other terms proportional to $(\nu T)^{k-m}\log(\nu T)$, which converge to 0 if $k > m$. The only diverging term is $(\nu T)^{k-m}\log(\nu T)$ when $k = m$, which becomes $\log(\nu T)$. Hence, the second term diverges as

$$\sum_{n=1}^{+\infty} a_{nm}\, j(n)\log(\nu T) \tag{22}$$

while other terms of $i^{(m)}(\nu)$ either converge to a finite value or diverge even faster than $\log(\nu T)$, because they are either polynomial or fractional terms of νT. The sum of the coefficients a_{nm} of all the fractional terms with the same order should then be 0. If we could not find a relationship among a_{nm} that make the sum 0, a paradox would arise completing the proof. In addition, the sum of the coefficients a_{nm} of $\log(\nu T)$ terms should also be 0, i.e.,

$$\sum_{n=1}^{+\infty} a_{nm}\, j(n) = 0 \quad \text{for all } m \geqslant 1. \tag{23}$$

According to Eq. (17), when $\nu T = 0$, $L(n \geqslant 1, \nu T)$ reaches its lower bound 0. Then the derivative $L'(n, 0)$, which equals to a_{n1} [see Eq. (19)], is positive or 0 for any $n \geqslant 1$, i.e.,

$$a_{n1} \geqslant 0. \tag{24}$$

Combining with Eq. (23), and noting that $j(n) > 0$, we have

$$a_{n1} = 0 \quad \text{for all } n \geqslant 1. \tag{25}$$

Similarly, based on $a_{n1} = 0$, we can derive $a_{n2} = 0$. This is because the second derivative $L''(n \geqslant 1, 0)$ also needs to be positive or 0, given that $L(n \geqslant 1, \nu T)$ is at its lower bound and its first derivative is 0. Continuing this process, we get

$$a_{nm} = 0 \tag{26}$$

for all $n \geqslant 1$ and $m \geqslant 1$. Substituting into Eq. (19), we have

$$L(n, \nu T) = 0 \quad \text{for any } n \geqslant 1 \text{ and any } \nu, \tag{27}$$

which leads to

$$L(0, \nu T) = 1 \quad \text{for any } \nu. \tag{28}$$

This is in contradiction to Eq. (2), $\sum_n L(n, \nu T)n = \nu T$, since $\nu > 0$ means that the neuron fires and $L(0, \nu T)$ cannot be 1. Therefore, Eq. (13) leads to a paradox, which indicates that the set of increasing points E_ν^* cannot be infinite.

In summary, this proves that a continuous activation function is inconsistent with Eq. (8). This means that the optimal activation function for a single neuron must be discrete for any noise generation function.

B. The optimal activation functions for a population of neurons are discrete

Next, we investigate a population of N neurons, made up of ON and OFF neurons that have monotonically increasing and decreasing activation functions as a function of the stimulus s, respectively. We continue to consider the same constraints of a maximal firing rate and zero spontaneous firing rate [Fig. 1(a)]. Under these conditions, the optimal activation functions for all neurons in the population continues to be discrete for any analytic noise generation function $L(n_i, \nu_i T)$.

We define the "dynamic range" of a neuron to be the interval of s that leads to unsaturated firing rates, i.e., $\{s|0 < \nu_i(s) < \nu_{\max}\}$ for neuron i (see Sec. II). For a discrete activation function, the dynamic range is the interval between the lowest and highest threshold. We assume that the dynamic ranges of any two neurons do not overlap and also assumed that any OFF neuron encodes smaller stimuli than any ON neuron (see Sec. II), which is consistent with experimental measurements [12] and previous theoretical study [18].

We consider a mixed population of m ON neurons and $N - m$ OFF neurons. To proceed, we label all ON neurons with decreasing indices (m to 1) from low to high dynamic ranges, where the ON neuron with the highest dynamic range has index 1. Similarly, we label all OFF neurons with increasing indices ($m + 1$ to N) from low to high dynamic ranges to ensure symmetry in our mathematical expressions (note this ordering is different from previous work [18] to ensure symmetry of the expressions).

If one of the ON neurons $1, 2, \ldots, m$ fires, assuming that spontaneous firing rates are 0, we know that the stimulus s is higher than, or is at least within the dynamic range of neuron m. Then we also know the firing rates of neurons $m + 1, m + 2, \ldots, N$, which means the spike counts of these neurons cannot give any new information about the stimulus s. Based on this, we can write the mutual information encoded by the mixture of m ON neurons and $N - m$ OFF neurons as

$$I_N(F_1, \ldots, F_N) = I_m(F_1, \ldots, F_m) + Q_m I_{N-m}\left(F_{m+1}^{(m)}, \ldots, F_N^{(m)}\right). \tag{29}$$

Here $F_i = F_{\nu_i}$ is defined in the same way as before [Eq. (11)], while Q_m denotes the probability that none of the ON neurons $1, 2, \ldots, m$ fires. We additionally define the terms $F^{(m)}$ to denote the "revised" distribution functions under the condition that none of the neurons $1, 2, \ldots, m$ fires, i.e., given an arbitrary firing rate $\tilde{\nu}$,

$$F_i^{(m)}(\tilde{\nu}) = \text{Prob}(\nu_i \leqslant \tilde{\nu} | n_1 = \cdots = n_m = 0). \tag{30}$$

From Bayes rule, we can write

$$F_i^{(m)}(\tilde{\nu}) = \frac{F_i(\tilde{\nu})\,\text{Prob}(n_1 = \cdots = n_m = 0 | \nu_i \leqslant \tilde{\nu})}{Q_m}. \tag{31}$$

Here, Q_m does not depend on $\tilde{\nu}$. Within the dynamic range of neuron i (where $i > m$), the firing rate of neurons $1, \ldots, m$ are all 0, which means $\text{Prob}(n_1 = \cdots = n_m = 0 | \nu_i \leqslant \tilde{\nu})$ also does not depend on $\tilde{\nu}$ in the dynamic range of F_i. Therefore, if F_i is discrete, $F_i^{(m)}$ will also be discrete, and vice versa. This relationship also exists between F_i and $F_i^{(j)}$ where j is an arbitrary positive integer smaller than i.

Following a similar logic, we can also decompose the mutual information encoded by a population of N neurons in Eq. (29) into N single terms, each containing the mutual information encoded by one neuron, i.e.,

$$\begin{aligned} I_N = {} & I(F_1) + P_1(0)\big\{I\big(F_2^{(1)}\big) + P_2^{(1)}(0)\big[I\big(F_3^{(2)}\big) + \cdots \\ & + P_{N-1}^{(N-2)}(0)\, I\big(F_N^{(N-1)}\big)\big]\big\} \end{aligned} \tag{32}$$

where $P_i(0) = \int L(0, \nu_i T)\, dF_i$ denotes the probability that neuron i does not fire, i.e., $n_i = 0$. Furthermore, we have used $I(F_i^{(i-1)})$ to denote the mutual information of neuron i assuming that neurons $1, \ldots, i-1$ do not fire. Since m does not explicitly appear in this equation, Eq. (32) applies to any mixed ON-OFF population, including homogeneous ON populations (where $m = N$) or homogeneous OFF populations (where $m = 0$).

We use mathematical induction to demonstrate that the optimal activation functions in a population are all discrete. Having already shown this for a single neuron, we assume it is true for a population of $N - 1$ cells. Then we add an additional neuron and show the optimal activation functions of all N neurons are discrete. Without loss of generality, we assume that the newly added neuron is an ON neuron with a highest dynamic range, labeled with 1, and the remaining $N - 1$ neurons $2, \ldots, N$. The sum of all the terms multiplying $P_1(0)$ in Eq. (32) has the same mathematical form as I_{N-1}. As a result, the sum equals $I_{N-1}^{\max}$ when optimizing $F_2^{(1)}, \ldots, F_N^{(N-1)}$, allowing us to write

$$I_N = I(F_1) + P_1(0) I_{N-1}^{\max}. \tag{33}$$

Meanwhile, because we assumed that optimal activation functions are discrete in a population of $N - 1$ neurons, the optimal $F_2^{(1)}, \ldots, F_N^{(N-1)}$ are all discrete. As we argued before, since F_i and $F_i^{(j)}$ are either both discrete or both continuous, this means that $F_2, \ldots, F_N$ are all discrete. As before [Eq. (6)], we can also define the density of mutual information here as

$$\begin{aligned} \widetilde{i}(\nu_1) &= \sum_{n_1} p(n_1|\nu_1) \log \frac{p(n_1|\nu_1)}{P(n)} + p(n_1 = 0|\nu_1) I_{N-1}^{\max} \\ &= \sum_{n_1} L(n_1, \nu_1 T) \log \frac{L(n_1, \nu_1 T)}{P(n)} + L(0, \nu_1 T) I_{N-1}^{\max}. \end{aligned} \tag{34}$$

Therefore, maximizing I_N is equivalent to optimizing F_1 assuming optimal $F_2^{(1)}, \ldots, F_N^{(N-1)}$ as in Eq. (33). If the optimal F_1 is continuous, when I_N is maximized we have

$$\widetilde{i}(\nu_1) = I_N^{\max}, \quad \nu_1 \in [0, \nu_{\max}] \tag{35}$$

and this leads to (see the SM [31])

$$L(n_1 = 0, \nu_1 T) = 1 \quad \text{for any } \nu_1. \tag{36}$$

Similar to Eq. (28), here Eq. (36) is also in contradiction to Eq. (2), $\sum_{n_1} L(n_1, \nu_1 T) n_1 = \nu_1 T$. Therefore, the optimal F_1 must be discrete and we have proved that all the N optimal activation functions need to be discrete.

Hence, using mathematical induction, we have proved that all the neurons' optimal activation functions in a population of any number of neurons are discrete.

C. The optimal thresholds and the maximal mutual information for a population of binary neurons

Having shown that the optimal activation functions in population are discrete for any noise generation function, we first consider the simplest discrete activation function, which is binary, to derive the optimal thresholds and the maximal mutual information. As before, we study a combination of a total of N neurons, m ON and $N - m$ OFF neurons. Assuming that the spontaneous firing rate (the firing rate when the stimulus s is subthreshold) is 0 [Fig. 2(a)], only two parameters characterize the activation function of neuron i, $\nu_i(s)$: the threshold (denoted as θ_i) and the maximal firing rate, which as before we assume is the same for all neurons ($\nu_{\max,i} = \nu_{\max}$ for all i).

Because there is only one threshold for every neuron, the dynamic range of every neuron is compressed to a single point, the neuron's threshold. Labeling all the neurons as before

$$\theta_{m+1} < \cdots < \theta_N < \theta_m < \cdots < \theta_1, \tag{37}$$

we note that there is only one noisy firing level at the maximum firing rate. The absence of noise in the zero firing state enables us to lump all firing states with nonzero spike count into one [18] (see also SM [31]).

Because the optimal activation functions are discrete, following [18], we can replace the firing thresholds with the intervals of stimulus space partitioned by those thresholds [Fig. 2(a)] and optimize them instead of directly optimizing thresholds, i.e., we define

$$u_i = \text{Prob}(\nu_i = \nu_{\max}) = \begin{cases} \int_{\theta_i}^{+\infty} ds\, p(s), & \text{for ON} \\ \int_{-\infty}^{\theta_i} ds\, p(s), & \text{for OFF.} \end{cases} \tag{38}$$

Denoting

$$\begin{aligned} R &= \nu_{\max} T, \\ q &= L(0, R) = 1 - \sum_{n=1}^{+\infty} L(n, R), \end{aligned} \tag{39}$$

we extend the finding of [18] to any noise generation function that the maximal mutual information is

$$I_N^{\max} = \log(1 + N(1-q) q^{q/(1-q)}) = -\log(P(\vec{0})) \tag{40}$$

where $P(\vec{0})$ is the probability that spike counts are all 0 (see the SM [31]). The maximal information $I_N^{\max}$ is independent of the composition of ON neurons and OFF neurons and

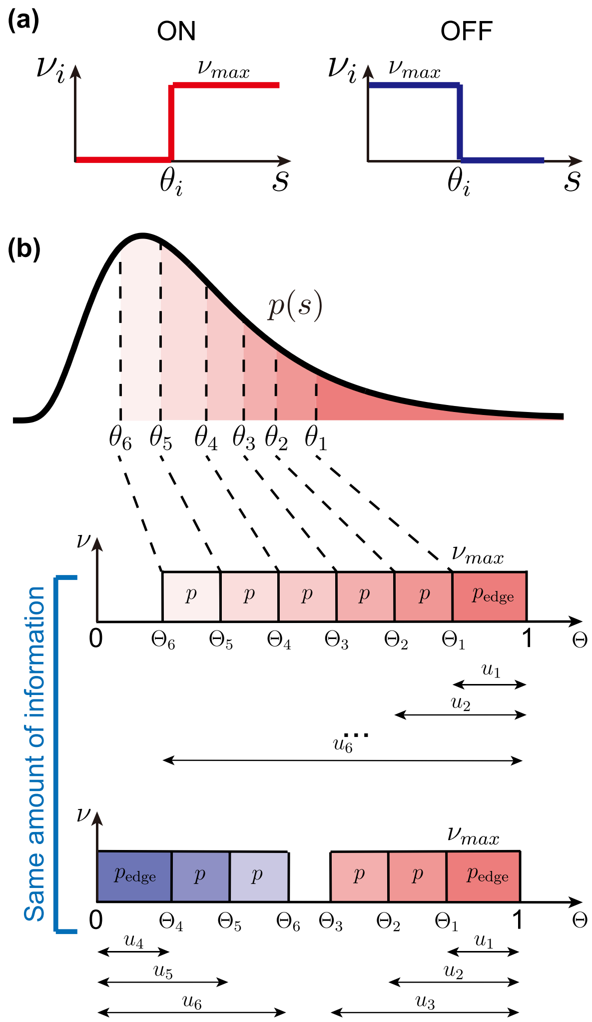

FIG. 2. Efficient population coding of binary neurons. (a) Activation functions of ON and OFF binary neurons. Each neuron has the same maximal firing rate ν_{max}, with an activation function described by a single threshold θ_i. (b) Optimal configurations of homogeneous populations with only ON neurons and mixed ON and OFF neurons. Θ_i denotes the cumulative probability of s above threshold θ_i. The optimal thresholds partition the cumulative stimulus space into regular intervals [Eqs. (46)–(48)]. The optimized mutual information is independent of the ON-OFF mixture for any noise generation function [Eq. (49)].

only depends on the total number of neurons N. Hence, we have generalized the previously termed "equal coding theorem" to other noise generation functions than Poisson [18]. In addition, comparing the maximal mutual information of a single neuron population I_1^{max}, and of an N-neuron population I_N^{max}, reveals that the maximum mutual information encoded by a population of neurons increases logarithmically with the number of neurons,

$$I_N^{\text{max}} = \log\left[N\left(\exp\left(I_1^{\text{max}}\right) - 1\right) + 1\right]. \tag{41}$$

Given this maximum mutual information, we next calculate the optimal threshold distribution of the population's binary activation functions. We can show (see the SM [31]) that the optimal $\{u_i\}$ for the ON neurons are

$$u_i = \frac{1 + (i-1)(1-q)}{N(1-q) + q^{-q/(1-q)}} \tag{42}$$

and for the OFF neurons

$$u_i = \frac{1 + (m-i+1)(1-q)}{N(1-q) + q^{-q/(1-q)}}. \tag{43}$$

The terms $\{u_i\}$ represent an arithmetic progression for any noise generation function L, whereby all the firing thresholds equally partition the probability space of stimuli similar to the case with Poisson noise [18]. If we define

$$\begin{aligned} &p_1 = u_1, \quad p_{m+1} = u_{m+1}, \\ &p_i = u_i - u_{i-1}, \quad i = 2, \dots, m, m+2, \dots, N, \end{aligned} \tag{44}$$

as the probabilities of the stimulus intervals, i.e., the intervals of stimuli s that lead to the same firing rates $\vec{\nu}$ [Fig. 2(b)], we have

$$\begin{aligned} &p_1 = p_{m+1} \overset{\text{def}}{=} p_{\text{edge}}, \\ &p_2 = \cdots = p_m = p_{m+2} = \cdots = p_N \overset{\text{def}}{=} p, \\ &p = (1-q)p_{\text{edge}}. \end{aligned} \tag{45}$$

This gives us the optimal thresholds in cumulative stimulus space [Fig. 2(b)],

$$\Theta_i = \int_{-\infty}^{\theta_i} ds\, p(s), \tag{46}$$

for the ON cells as:

$$\begin{aligned} \Theta_1 &= 1 - p_{\text{edge}}, \\ \Theta_2 &= 1 - p_{\text{edge}} - p, \\ &\cdots \\ \Theta_m &= 1 - p_{\text{edge}} - (m-1)p, \end{aligned} \tag{47}$$

and for the OFF cells as:

$$\begin{aligned} \Theta_{m+1} &= p_{\text{edge}}, \\ \Theta_{m+2} &= p_{\text{edge}} + p, \\ &\cdots \\ \Theta_N &= p_{\text{edge}} + (N-m-1)p. \end{aligned} \tag{48}$$

Given these optimal thresholds, we can combine this with Eq. (40) to find the expression for the optimal mutual information

$$I_N^{\text{max}} = -\log(1 - Np). \tag{49}$$

Hence, we conclude that, for a mixed ON-OFF population with binary activation functions, the optimal thresholds and mutual information look exactly the same for any noise generation function as for Poisson [18]. Homogeneous populations with only ON or OFF neurons, and mixed ON-OFF populations with any ON-OFF mixture can encode the same amount of information.

D. The optimal thresholds for a population of binary neurons with heterogeneous maximal firing rates

Different neurons might have different maximal firing rate constraints. For example, ON ganglion cells in the primate retina have higher firing rates than OFF ganglion cells [30]. To explore the effect of these maximal firing rate differences on efficient coding, next we assume that the different neurons in the population might have different maximal firing rates, and consider a heterogeneous population of ON and OFF neurons. In this case, we define $\nu_{\max,i}$ as the maximal firing rate of neuron i,

$$\begin{aligned} R_i &= \nu_{\max,i} T, \\ q_i &= L(0, R_i). \end{aligned} \tag{50}$$

Similar to Eq. (38), we define u_i as the probability that neuron i fires at its maximal firing rate, i.e., $u_i = \mathrm{Prob}(\nu_i = \nu_{\max,i})$. Then we can prove that the optimal thresholds are {see Eqs. (S6.9) and (S6.10) in the SM [31]}

$$u_i = \frac{q_i^{q_i/(1-q_i)} + \sum_{j=1}^{i-1}(1-q_j)\,q_j^{q_j/(1-q_j)}}{1+\sum_{j=1}^{N}(1-q_j)\,q_j^{q_j/(1-q_j)}} \tag{51}$$

for ON neurons and

$$u_i = \frac{q_i^{q_i/(1-q_i)} + \sum_{j=m+1}^{i-1}(1-q_j)\,q_j^{q_j/(1-q_j)}}{1+\sum_{j=1}^{N}(1-q_j)\,q_j^{q_j/(1-q_j)}} \tag{52}$$

for OFF neurons. The maximal mutual information now becomes

$$I_N = \log\left[1+\sum_{j=1}^{N}(1-q_j)\,q_j^{q_j/(1-q_j)}\right]. \tag{53}$$

This result tells us that as long as the distribution of the maximal firing rates is the same (i.e., the same set of $\{q_i\}$), shuffling the thresholds within the ON and OFF subpopulations, replacing ON neurons with OFF, or replacing OFF neurons with ON, does not change the maximal mutual information (Fig. 3). Similar to Eq. (44), defining

$$\begin{aligned} p_1 &= u_1, \\ p_{m+1} &= u_{m+1}, \\ p_i &= u_i - u_{i-1}, \quad i = 2,\dots,m,m+2,\dots,N, \end{aligned} \tag{54}$$

we can derive the stimulus intervals partitioned by the thresholds as

$$\begin{aligned} p_1 &= \frac{q_1^{q_1/(1-q_1)}}{1+\sum_{j=1}^{N}(1-q_j)\,q_j^{q_j/(1-q_j)}} \\ &= e^{-I_N}\, q_1^{q_1/(1-q_1)}, \\ p_{m+1} &= \frac{q_{m+1}^{q_{m+1}/(1-q_{m+1})}}{1+\sum_{j=1}^{N}(1-q_j)\,q_j^{q_j/(1-q_j)}} \\ &= e^{-I_N}\, q_{m+1}^{q_{m+1}/(1-q_{m+1})}, \\ p_i &= \frac{q_i^{q_i/(1-q_i)} - q_{i-1}^{1/(1-q_{i-1})}}{1+\sum_{j=1}^{N}(1-q_j)\,q_j^{q_j/(1-q_j)}} \\ &= e^{-I_N}\left[q_i^{q_i/(1-q_i)} - q_{i-1}^{1/(1-q_{i-1})}\right], \\ & i = 2,\dots,m,m+2,\dots,N. \end{aligned} \tag{55}$$

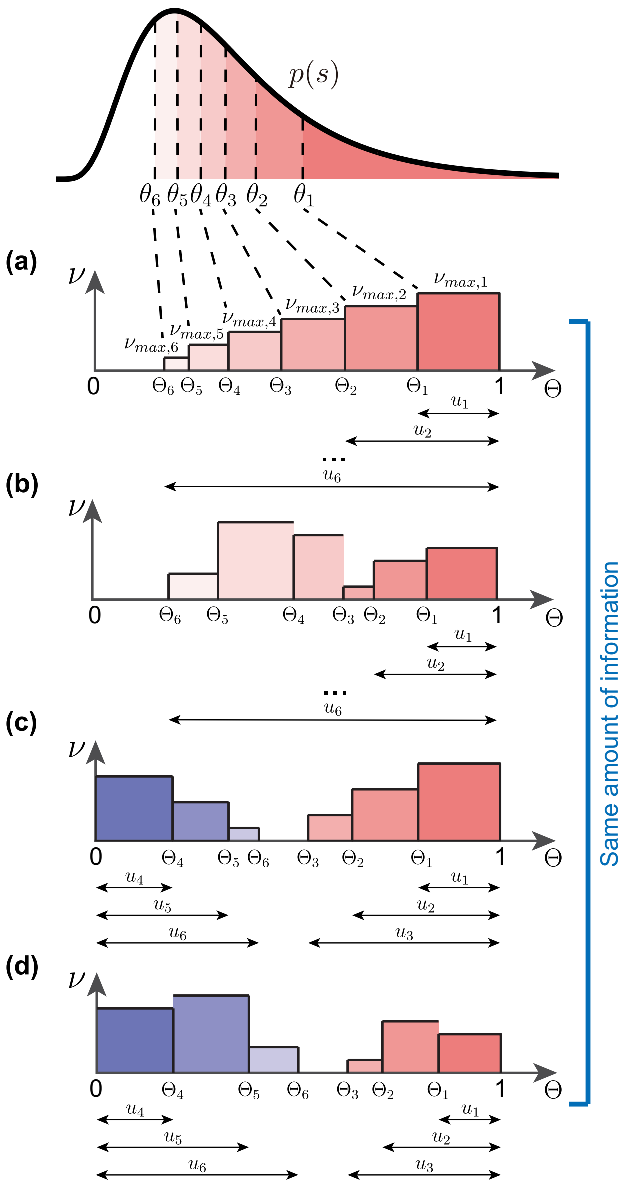

FIG. 3. Efficient population coding of binary neurons with heterogeneous maximal firing rates. (a) Optimal configurations of homogeneous populations with only ON neurons. Θ_i denotes the cumulative probability of s above threshold θ_i. The optimal thresholds partition the cumulative stimulus space into intervals, which increase with the maximal firing rate of the neurons within a population. (b) Same as (a) but with maximal firing rates shuffled. (c) Same as (a) but for mixed populations of ON and OFF neurons. (d) Same as (c) but with maximal firing rates shuffled. All populations code the same amount of mutual information assuming the same distribution of the maximal firing rates [Eq. (53)].

One can show that (see the SM [31])

$$\frac{d}{dq}\big(q^{q/(1-q)}\big) < 0, \qquad \frac{d}{dq}\big(q^{1/(1-q)}\big) > 0. \tag{56}$$

We note that recently [35] derived a similar solution for binary neurons.

In our solution [Eq. (56)], $q_i \to 0$ and $q_{i-1} \to 0$, p_i converges to its maximum e^{-I_N}. On the other hand, when $q_i \to 1$

and $q_{i-1} \to 1$, p_i converges to its minimum 0. Therefore, given a fixed amount of mutual information I_N, p_1 increases with $\nu_{\max,1}$, p_{m+1} increases with $\nu_{\max,m+1}$, and other p_i increases with both $\nu_{\max,i}$ and $\nu_{\max,i-1}$ (Fig. 3). This dependency indicates that when the maximal firing rates are heterogeneous in a population, stimulus intervals corresponding to high firing neurons are wider, which can be understood as a result of noise reduction.

E. The optimal activation functions with increasing maximal firing rate constraint

For a single neuron, Nikitin *et al.* [16] showed that the optimal activation function is discrete with an increasing number of steps as the maximal firing rate constraint increases [Fig. 1(c), assuming the coding window T is constant]. Given our result that even in a population of neurons the optimal activation functions remain discrete with increasing maximal firing rate constraint (Sec. III B), we next determine the number of steps in the activation function for each neuron.

Similar to a single neuron [16], the number of steps for all neurons in the population increases as a function of increasing maximal firing rate ($\nu_{\max,i}$), or equivalently decreasing level of noise ($R_i = \nu_{\max,i}T$). In a population of neurons with identical maximal firing rates ($\nu_{\max,i} = \nu_{\max}$ and hence $R_i = R$ for all i), all the activation functions have the same number of steps for a given noise level, such that, for example, the optimal activation function cannot be binary for one neuron and ternary for another neuron in the same population. To verify this, we perform extensive numerical calculations with multiple noise generation functions $L(n, \nu T)$ for small populations of neurons (Fig. 4, and Fig. S1 in the SM [31]), although we lack an analytical proof that this is universally true.

F. The optimal thresholds and the maximal mutual information for a population of ternary neurons

Knowing that at some critical level of maximal firing rate the optimal activation functions of all neurons in the population increase the number of steps, we next generalize our results beyond binary neurons to activation functions with any number of steps and any noise generation function. For the time being, we assume identical maximal firing rates across all neurons in the population. We first start with ternary activation functions with three possible firing rate levels [Fig. 5(a)]. In this case, one can no longer lump the firing states as we did previously for binary neurons because there is more than one noisy firing level in the activation function. We show that the "equal coding theorem" remains valid in that the same maximal information is encoded by any mixture of ON and OFF neurons, and the thresholds of the ternary activation functions of each neuron divide the cumulative stimulus space equally [Fig. 5(b)].

Two thresholds, θ_{i1} and θ_{i2}, describe a ternary neuron i, separating the spontaneous firing rate $\nu = 0$, an intermediate firing rate $\nu = f_i\nu_{\max}$, and the maximal firing rate $\nu = \nu_{\max}$ [Fig. 5(a)]. In a population of m ON neurons and $N - m$ OFF neurons, we assign all ON neuron indices, from the highest to the lowest threshold, and for the OFF neurons, from the lowest to the highest, as for the binary neurons [Eq. (37) and Fig. 2(b)].

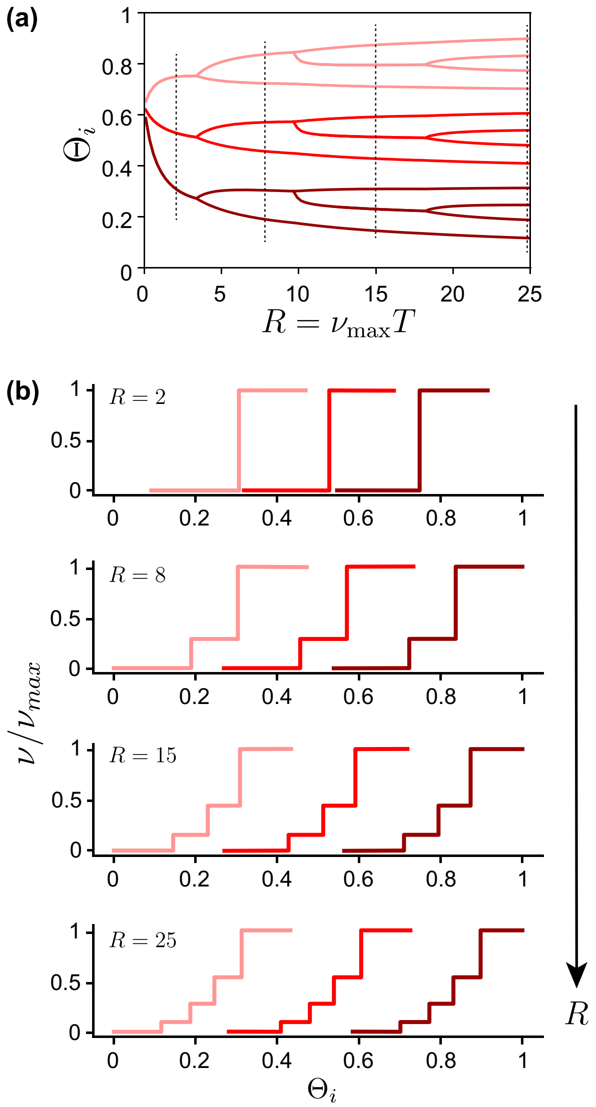

FIG. 4. The optimal activation functions in a neuronal population are discrete with the number of steps (i.e., thresholds) increasing with the maximal firing rate. (a) The optimal thresholds in a population of three ON neurons with Poisson noise generation function. Θ_i denotes the cumulative probability of s above a threshold. The activation functions acquire more discrete steps as the maximal firing rate $\nu_{\max}$ increases (assuming constant T). (b) Optimal activation functions for the neurons in (a). Note that all the activation functions have the same number of steps for a given noise level. The values of $R = \nu_{\max}T$ are indicated on the top left of each plot and shown as dashed lines in (a).

Because we assumed that the dynamic ranges of neurons do not overlap (see Sec. II), we have

$$\begin{aligned}\theta_{m+1,2} \leqslant \theta_{m+1,1} < \cdots < \theta_{N,2} \leqslant \theta_{N,1}\\ < \theta_{m,1} \leqslant \theta_{m,2} < \cdots < \theta_{11} \leqslant \theta_{12}.\end{aligned} \tag{57}$$

We transform the firing thresholds $\{\theta_{i1}\}$ and $\{\theta_{i2}\}$ to intervals of stimulus s, as for the binary neurons in Eq. (38), for

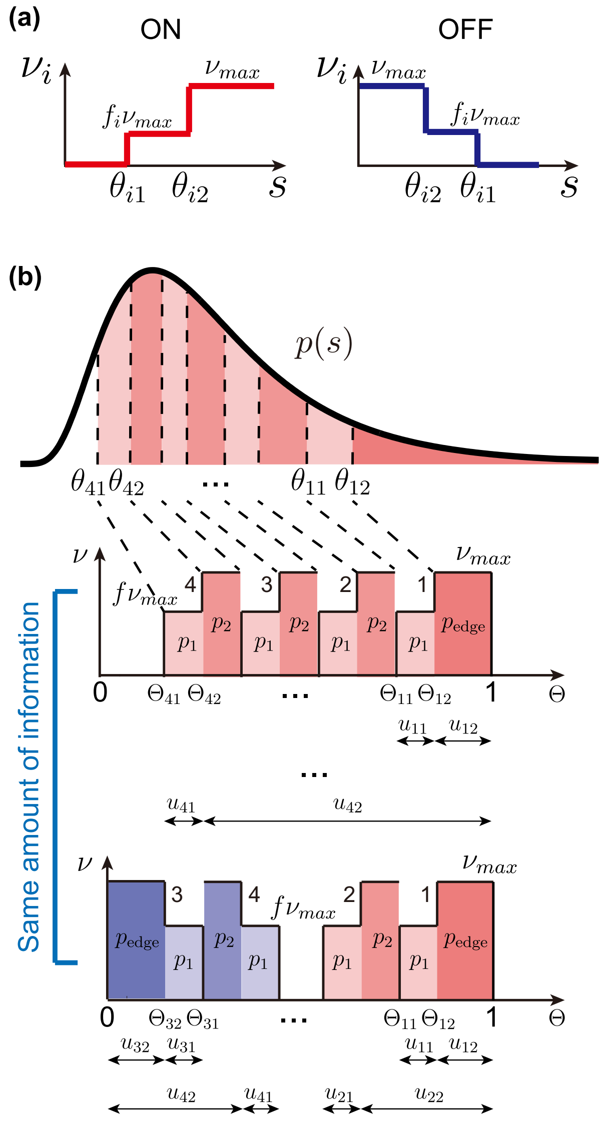

FIG. 5. Efficient population coding of ternary neurons. (a) Activation functions of ON and OFF ternary neurons. Each neuron has the same maximal firing rate $\nu_{\rm max}$, with an activation function described by two thresholds, $\theta_{i,1}$ and $\theta_{i,2}$, and an intermediate firing rate $f_i\nu_{\rm max}$. (b) Optimal configurations of homogeneous populations with only ON neurons and mixed ON and OFF neurons. Θ_{ij} denotes the cumulative probability of s above a threshold θ_{ij} for $j = \{1, 2\}$. The optimal thresholds partition the cumulative stimulus space into regular intervals [Eqs. (60)–(62)]. The optimized mutual information is independent of the ON-OFF mixture for any noise generation function [Eq. (65)].

the ON neurons,

$$u_{i1} = \int_{\theta_{i1}}^{\theta_{i2}} ds\, p(s), \quad u_{i2} = \int_{\theta_{i2}}^{+\infty} ds\, p(s) \tag{58}$$

and for the OFF neurons,

$$u_{i1} = \int_{\theta_{i2}}^{\theta_{i1}} ds\, p(s), \quad u_{i2} = \int_{-\infty}^{\theta_{i2}} ds\, p(s). \tag{59}$$

We can still find direct relations between (u_{i1}, u_{i2}, f_i) and $(u_{i+1,1}, u_{i+1,2}, f_{i+1})$ {Eq. (S7.32) in the SM [31]}. Finally, we can express the optimal thresholds and intermediate firing levels as

$$\begin{aligned} u_{i1} &= \frac{u_1}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]}, \\ u_{i2} &= \frac{u_2 + (i-1)[u_1(1-q_1) + u_2(1-q_2)]}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]} \\ &\quad (\text{ON neurons, } i = 1, \dots, m), \\ u_{i2} &= \frac{u_2 + (i-m-1)[u_1(1-q_1) + u_2(1-q_2)]}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]} \\ &\quad (\text{OFF neurons, } i = m+1, \dots, N), \\ f_i &= f, \end{aligned} \tag{60}$$

where (u_1, u_2, f) is the optimal (u_{11}, u_{12}, f_1) in a population of a single neuron ($N = 1$), $q_1 = L(0, fR)$ and $q_2 = L(0, R)$ with $R = \nu_{\rm max}T$.

Defining the cumulative stimulus intervals as Eq. (44),

$$\begin{aligned} &p_{i1} = u_{i1}, \\ &p_{12} = u_{12}, p_{m+1,2} = u_{m+1,2}, \\ &p_{i2} = u_{i2} - u_{i-1,2} - u_{i-1,1}, \\ &i = 2, \dots, m, m+2, \dots, N, \end{aligned} \tag{61}$$

which are the probabilities of the stimulus intervals that have the same firing rates $\vec{\nu}$. With optimal $\{u_{i1}\}$, $\{u_{i2}\}$, and $\{f_i\}$ [Eq. (60)], these intervals can be expressed as

$$\begin{aligned} p_{11} = p_{21} = \cdots = p_{N1} &= \frac{u_1}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]} \overset{\text{def}}{=} p_1, \\ p_{22} = \cdots = p_{m2} = p_{m+2,2} = \cdots = p_{N2} &= \frac{-u_1 q_1 + u_2(1-q_2)}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]} \overset{\text{def}}{=} p_2, \\ p_{12} = p_{m+1,2} &= \frac{u_2}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]} \overset{\text{def}}{=} p_{\rm edge}, \\ p_1 q_1 + p_2 &= p_{\rm edge}(1-q_2). \end{aligned} \tag{62}$$

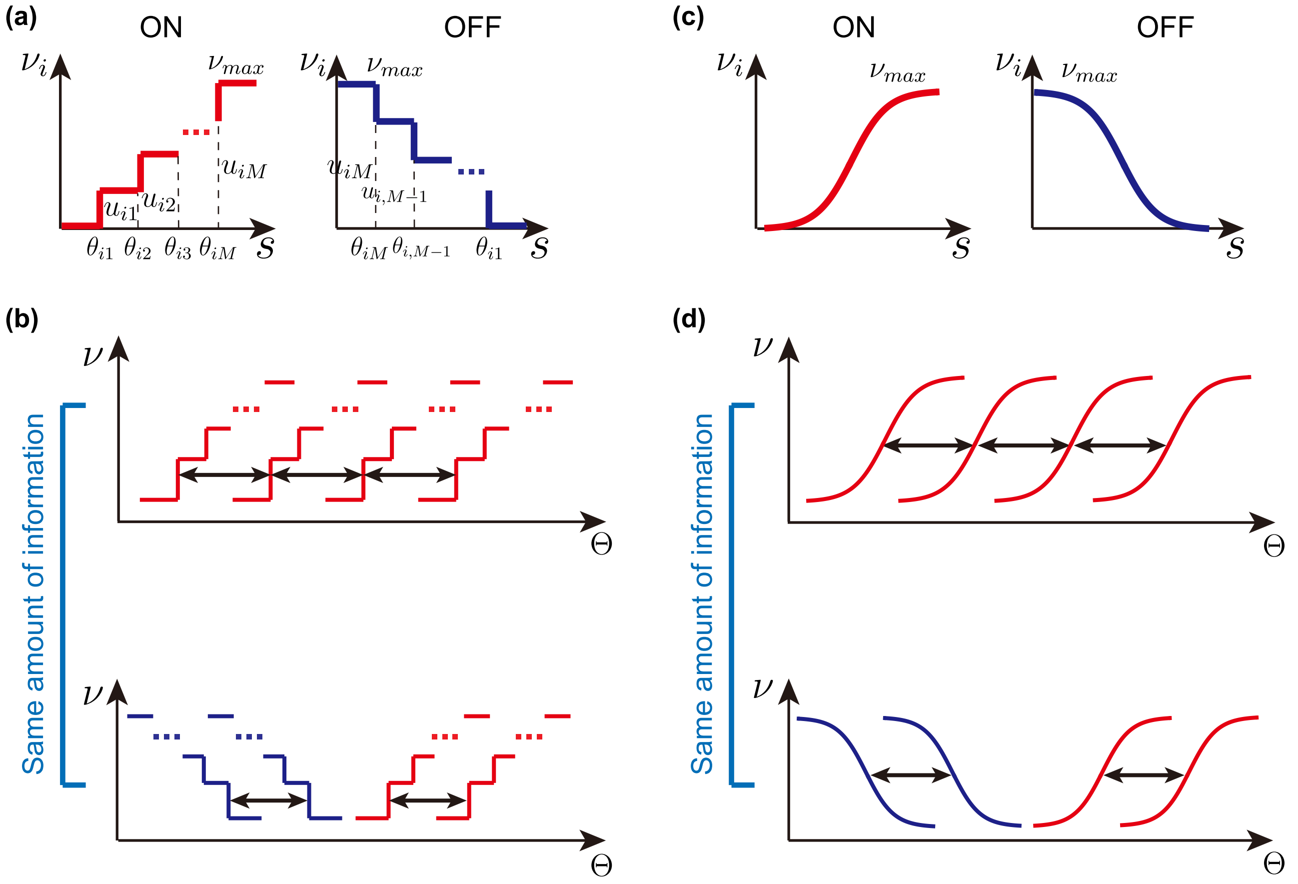

FIG. 6. (a) Efficient population coding of neurons with any activation function. Activation functions of ON and OFF $(M+1)$-ary neurons. Each neuron has the same maximal firing rate $\nu_{\max}$, and $M+1$ possible firing rates. (b) Optimal configurations of homogeneous ON population and ON-OFF mixtures of $(M+1)$-ary neurons. Black arrows indicate that the distances between each two adjacent activation functions are the same in the probability space of the stimuli s. (c) As $M \to +\infty$, the activation functions of $(M+1)$-ary neurons become continuous. (d) Correspondingly, the optimal configuration of activation functions also transforms into the continuous limit as $M \to +\infty$.

This derives the optimal thresholds for a mixed population of ON and OFF cells [Fig. 5(b)].

Similar to binary neurons {Eq. (40) in the SM [31]}, the maximal mutual information can be related to $P(\vec{0})$ as

$$I_N^{\max} = -\log P(\vec{0}). \tag{63}$$

This allows us to derive the relationship between the maximal mutual information and the stimulus interval p as (see the SM [31])

$$I_N^{\max} = -\log\left[1 - N(p_1 + p_2)\right]. \tag{64}$$

Using the optimal values of p_1 and p_2 {Eq. (S7.57) in the SM [31] and Eq. (62)}, we have

$$I_N^{\max} = -\log \frac{1 - u_1(1-q_1) - u_2(1-q_2)}{1 + (N-1)[u_1(1-q_1) + u_2(1-q_2)]}. \tag{65}$$

We can then write the maximal mutual information of an N-neuron ternary population as a function of the maximal mutual information of a single neuron population,

$$I_N^{\max} = \log\left[N\left(\exp\left(I_1^{\max}\right) - 1\right) + 1\right], \tag{66}$$

similar to the case with binary neurons [Eq. (41)].

G. The optimal thresholds and the maximal mutual information for a population of neurons with any shapes of activation functions

After generalizing the optimal thresholds in our efficient population coding framework from binary to ternary neurons, it is now straightforward to generalize them to activation functions with any number of steps. Activation functions with more than three steps can be represented with multiple p_{ij} and f_i. For example, an $(M+1)$-ary neuron i can be described with M thresholds, $\theta_{i1}, \dots, \theta_{iM}$, separating the spontaneous firing rate $\nu = 0$, intermediate firing rate $\nu = f_{i1}\nu_{\max}, \dots, f_{i,M-1}\nu_{\max}$, and the maximal firing rate $\nu = \nu_{\max}$ [Fig. 6(a)]. For simplicity, we also define $f_{iM} = 1$ for any index i.

Defining the cumulative stimulus thresholds similarly to binary neurons [Eqs. (38) and (44)] and ternary neurons [Eqs. (58) and (61)], we have for ON neurons

$$\begin{aligned} u_{ik} &= \int_{\theta_{ik}}^{\theta_{i,k+1}} ds\, p(s),\ k = 1, \dots, M-1, \\ u_{iM} &= \int_{\theta_{iM}}^{+\infty} ds\, p(s), \end{aligned} \tag{67}$$

and for OFF neurons

$$\begin{aligned} u_{ik} &= \int_{\theta_{i,k+1}}^{\theta_{ik}} ds\, p(s), k = 1, \dots, M-1, \\ u_{iM} &= \int_{-\infty}^{\theta_{iM}} ds\, p(s). \end{aligned} \tag{68}$$

Then the cumulative stimulus intervals can be written as

$$\begin{aligned} p_{ik} &= u_{ik}, \quad i = 1, \dots, N; \quad k = 1, \dots, M-1, \\ p_{1M} &= u_{1M}, \quad p_{m+1,M} = u_{m+1,M}, \\ p_{iM} &= u_{iM} - \sum_{k=1}^{M} u_{i-1,k}, \quad i = 2, \dots, m, m+2, \dots, N, \end{aligned} \tag{69}$$

and based on this we can calculate the mutual information and the optimal thresholds for discrete activation functions with any number of steps.

The calculations with $(M+1)$-ary neurons follow very similarly to the calculations with ternary neurons (see the SM [31]). In summary, for a population of m ON and $N-m$ OFF $(M+1)$-ary neurons, when the mutual information $I(s, \vec{n})$ is maximized, we have

$$\begin{aligned} & p_{1k} = p_{2k} = \dots = p_{Nk} \overset{\text{def}}{=} p_k, \quad k = 1, 2, \dots, M-1, \\ & p_{2M} = \dots = p_{mM} = p_{m+2,M} = \dots = p_{NM} \overset{\text{def}}{=} p_M, \\ & p_{1M} = p_{m+1,M} \overset{\text{def}}{=} p_{\text{edge}}, \\ & f_{1k} = f_{2k} = \dots = f_{Nk} \overset{\text{def}}{=} f_k, \quad k = 1, 2, \dots, M-1, \\ & \sum_{k=1}^{M-1} p_k L(0, f_k R) + p_M = p_{\text{edge}}(1 - L(0, R)). \end{aligned} \tag{70}$$

This derives the optimal thresholds for populations of $(M+1)$-ary neurons [Fig. 6(b)].

Also same as before {Eq. (S7.21) in the SM [31] and Eqs. (64), (41), and (66)}, we still have

$$I_N^{\max} = -\log P(\vec{0}), \tag{71}$$

$$I_N^{\max} = -\log\left[1 - N\sum_{k=1}^{M} p_k\right], \tag{72}$$

and

$$I_N^{\max} = \log\left[N\left(\exp\left(I_1^{\max}\right) - 1\right) + 1\right]. \tag{73}$$

We find that the maximal mutual information increases logarithmically with the number of neurons, independent of the specific shape of the activation functions and the type of noise.

Even letting $M \to \infty$, which corresponds to the limiting case in which the activation functions are continuous [Fig. 6(c)], the optimal activation functions have the same shape on the probability space of the stimuli s, and the displacement from one to the next is a constant [Fig. 6(d)].

H. The optimal thresholds and the maximal mutual information for a population of neurons with any shapes of activation functions and heterogeneous maximal firing rates

The optimal thresholds of ternary neurons or any shapes of activation functions that we derived for identical maximal firing rates for all cells can also be generalized to heterogeneous maximal firing rates, which are different across the different cells. We first start with ternary neurons. Similarly as before [Eqs. (57)–(59)], for every neuron i in a population, we define an intermediate firing rate $f_i\, \nu_{\max,i}$, the cumulative stimulus intervals

$$\begin{aligned} u_{i1} &= \text{Prob}\,(\nu_i = f_i\, \nu_{\max,i}), \\ u_{i2} &= \text{Prob}\,(\nu_i = \nu_{\max,i}), \end{aligned} \tag{74}$$

and

$$\begin{aligned} q_{i1} &= L(0, f_i R_i), \\ q_{i2} &= L(0, R_i). \end{aligned} \tag{75}$$

Instead of searching for relations between (u_{i1}, u_{i2}, f_i) and $(u_{i+1,1}, u_{i+1,2}, f_{i+1})$ {Eq. (S7.32) in the SM [31]}, here we need to derive direct relations between neuron i in a population and a population of a single neuron with the same maximal firing rate (also $\nu_{\max,i}$). Therefore, for such a population of a single neuron, given its maximal firing rate $\nu_{\max}$, we denote $R = \nu_{\max} T$, and then

(i) $u_1^*(R)$ as its optimal u_1
(ii) $u_2^*(R)$ as its optimal u_2
(iii) $f^*(R)$ as its optimal f
(iv) $q_1^*(R) = L(0, f^*(R)R)$ as the probability $p(n = 0|\nu = f^*(R)R)$ after optimization
(v) $q_2^*(R) = L(0, R)$ as the probability $p(n = 0|\nu = R)$ after optimization
(vi) $P^*(0, R) = 1 - u_1^*(R)[1 - q_1^*(R)] - u_2^*(R)[1 - q_2^*(R)]$ as the probability of observing no spike after optimization.

With these new definitions, we can derive the optimal thresholds and intermediate firing rates as (see the SM [31])

$$\begin{aligned} u_{i1} &= \frac{u_1^*(R_i)}{P^*(0, R_i)} e^{-I_N}. \\ u_{i2} &= \left[\frac{u_2^*(R_i)}{P^*(0, R_i)} - \sum_{j=i}^{N} \frac{1}{P^*(0, R_j)} + (N - i)\right] e^{-I_N} + 1 \\ &\qquad \text{(ON neurons, } i = 1, \dots, m\text{)}, \\ u_{i2} &= \left[\frac{u_2^*(R_i)}{P^*(0, R_i)} - \sum_{j=i}^{N} \frac{1}{P^*(0, R_j)} \right. \\ &\qquad \left. - \sum_{j=1}^{m} \frac{1}{P^*(0, R_j)} + (N - i + m)\right] e^{-I_N} + 1 \\ &\qquad \text{(OFF neurons, } i = m+1, \dots, N\text{)}, \\ f_i &= f^*(R_i), \end{aligned} \tag{76}$$

and the maximal mutual information as

$$I_N = \log\left[\sum_{j=i}^{N} \frac{1}{P^*(0, R_j)} - (N - 1)\right]. \tag{77}$$

Note that similarly to the case of binary neurons with heterogeneous maximal firing rates [Eq. (53), Fig. 3], this equation indicates that the maximal mutual information is the same as long as all the distribution of maximal firing rates is the same, independent of how the population is mixed, any

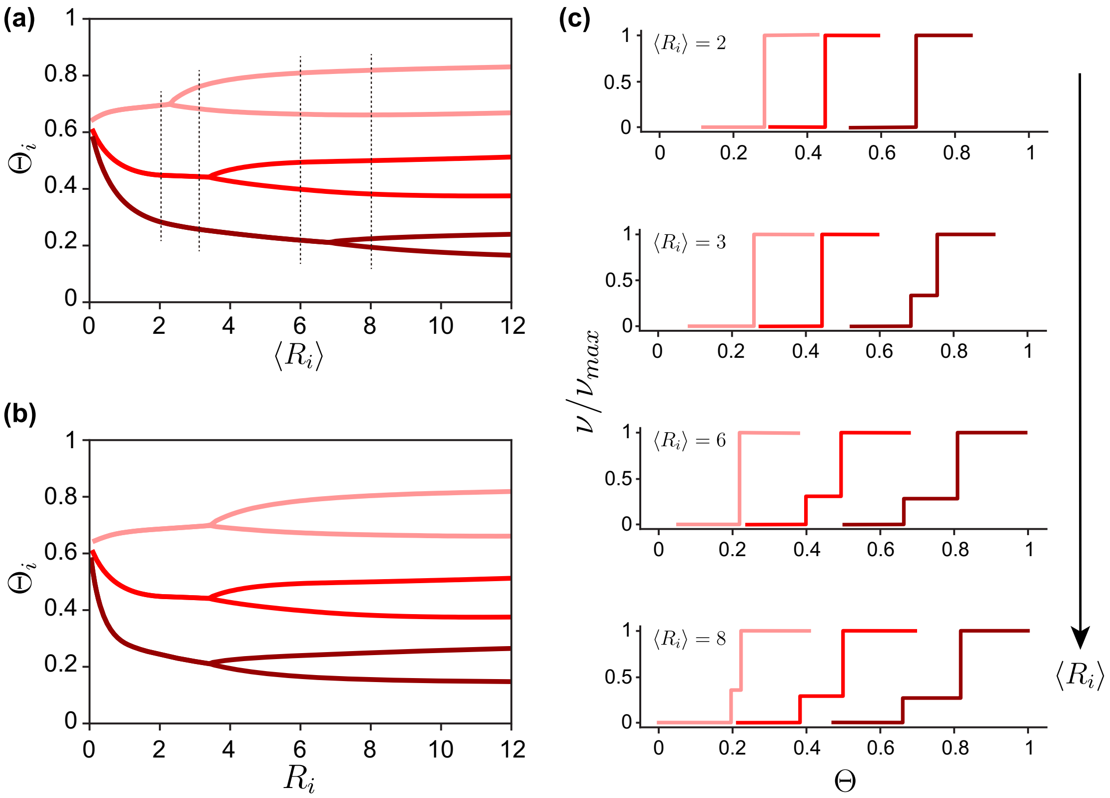

FIG. 7. The optimal activation functions in a neuronal population with heterogeneous maximal firing rate. The population consists of $N = 3$ ternary ON neurons, with $R_1 : R_2 : R_3 = 1 : 2 : 3$. (a) Optimal thresholds as a function of $\langle R_i \rangle = (R_1 + R_2 + R_3)/3$. (b) Optimal thresholds as a function of R_i of individual neurons. (c) Optimal activation functions at different values of $\langle R_i \rangle$ [dashed lines in (a)]. Note that the activation functions can have different numbers of steps in the same population. $\langle R_i \rangle$ values are indicated on the top left corner of each plot.

threshold shuffling within the ON and OFF subpopulations and any replacing of ON with OFF neurons and vice versa.

Defining $p_{i1} = u_{i1}$, $p_{12} = u_{12}$, $p_{i2} = u_{i-1,2} - u_{i-1,1} - u_{i2}$ $(i > 1)$, we can express the stimulus intervals as

$$
\begin{aligned}
p_{i1} &= \frac{u_1^*(R_i)}{P^*(0,R_i)} e^{-I_N}.\\
p_{i2} &= \left[\frac{u_2^*(R_i) - u_1^*(R_i)}{P^*(0,R_i)} - \frac{u_2^*(R_{i-1})\, q_2^*(R_{i-1}) - u_1^*(R_{i-1})(1 - q_1^*(R_{i-1}))}{P^*(0,R_{i-1})}\right] e^{-I_N} \quad (i = 2,\dots,m,m+2,\dots,N),\\
p_{12} &= \left[\frac{u_2^*(R_1)}{P^*(0,R_1)} - \sum_{j=1}^{N} \frac{1}{P^*(0,R_j)} + (N-1)\right] e^{-I_N} + 1,\\
p_{m+1,2} &= \left[\frac{u_2^*(R_{m+1})}{P^*(0,R_{m+1})} - \sum_{j=1}^{N} \frac{1}{P^*(0,R_j)} + (N-1)\right] e^{-I_N} + 1,\\
f_i &= f^*(R_i).
\end{aligned}
\tag{78}
$$

Here, $f_i = f^*(R_i)$ indicates that in a population, the optimal activation functions of different neurons may consist of different numbers of steps depending on the maximal firing rate constraint of those neurons (Fig. 7). This result is unique to the case of heterogeneous maximal firing rates in the population, compared to the case of identical maximal firing rates for all cells where all the optimal activation functions in the population have same number of steps (Fig. 4). The bifurcations where the optimal activation functions acquire more steps are now neuron specific (Fig. 7), only depending on the maximal firing rate of single neurons but not other neurons in the same population. This pattern of bifurcation differs from the case of identical maximal firing rates for all cells where the bifurcations occur at the same maximal firing rates for the whole population (Fig. 4).

Unlike binary neurons [Eq. (55)], here we cannot analytically find how the stimulus intervals $\{p_{i1}, p_{i2}\}$ depend on the maximal firing rates $\nu_{\max,i}$, or equivalently, the maximum spike counts $\{R_i\}$. To obtain some understanding of the optimal thresholds in Eq. (78), we performed numerical calculations for Poisson noise (Figs. S2 and S3 in the SM [31]).

Similar to ternary neurons (see the SM [31]), we can also derive the optimal thresholds and intermediate firing rates for any activation function (with any number of steps) as

$$\begin{aligned}
u_{ik} &= \frac{u_k^*(R_i)}{P^*(0,R_i)} e^{-I_N} \quad k=1,\dots,M-1,\\
u_{iM} &= \left[\frac{u_M^*(R_i)}{P^*(0,R_i)} - \sum_{j=i}^{N}\frac{1}{P^*(0,R_j)} + (N-i)\right] e^{-I_N} + 1 \quad (\text{ON neurons, } i=1,\dots,m),\\
u_{iM} &= \left[\frac{u_M^*(R_i)}{P^*(0,R_i)} - \sum_{j=i}^{N}\frac{1}{P^*(0,R_j)} - \sum_{j=1}^{m}\frac{1}{P^*(0,R_j)} + (N-i+m)\right] e^{-I_N} + 1 \quad (\text{OFF neurons, } i=m+1,\dots,N),\\
f_{ik} &= f_k^*(R_i), \quad k=1,\dots,M-1.
\end{aligned} \tag{79}$$

The maximal mutual information can be expressed in the same way as for ternary neurons [Eq. (77)],

$$I_N = \log\left[\sum_{j=i}^{N}\frac{1}{P^*(0,R_j)} - (N-1)\right], \tag{80}$$

which is again independent of the ON-OFF mixture and the shuffling within the ON or OFF subpopulations. Finally, the optimal stimulus intervals are given by

$$\begin{aligned}
p_{ik} &= \frac{u_1^*(R_i)}{P^*(0,R_i)} e^{-I_N} \quad k=1,\dots,M-1,\\
p_{iM} &= \left[\frac{u_M^*(R_i) - \sum_{k=1}^{M-1} u_k^*(R_i)}{P^*(0,R_i)} - \frac{u_M^*(R_{i-1})\, q_M^*(R_{i-1}) - \sum_{k=1}^{M-1} u_k^*(R_{i-1})\big(1-q_k^*(R_{i-1})\big)}{P^*(0,R_{i-1})}\right] e^{-I_N}\\
&\quad (i=2,\dots,m,m+2,\dots,N),\\
p_{1M} &= \left[\frac{u_M^*(R_1)}{P^*(0,R_1)} - \sum_{j=1}^{N}\frac{1}{P^*(0,R_j)} + (N-1)\right] e^{-I_N} + 1,\\
p_{m+1,M} &= \left[\frac{u_2^*(R_{m+1})}{P^*(0,R_{m+1})} - \sum_{j=1}^{N}\frac{1}{P^*(0,R_j)} + (N-1)\right] e^{-I_N} + 1,\\
f_{ik} &= f_k^*(R_i), \quad k=1,\dots,M-1.
\end{aligned} \tag{81}$$

In sum, we have generalized the efficient coding framework to any spiking noise, any shapes of activation functions, and any distribution of maximal firing rates. Although the solution becomes more complicated as we relax the assumptions, we can analytically link the optimal thresholds and optimal intermediate firing rates to those of a single neuron. Thus, the dependence of the optimal activation functions on the maximal firing rates, or equivalently the noise levels, can be interpreted intuitively. In addition, for any spiking noise, shapes of activation functions, and distribution of maximal firing rates, the maximal mutual information is always independent of the mixture of ON and OFF neurons, and the reordering of firing thresholds within ON or OFF subpopulations.

I. Information per spike is the highest for equal ON-OFF mixtures

While the total mutual information about a stimulus encoded by any mixture of ON and OFF cells is identical, the different ON and OFF populations have different total metabolic costs, measured in terms of the total mean firing rate of the population. Using the derived optimal thresholds, we can calculate the total population mean firing rate for different ON and OFF mixtures. For a population of binary neurons, same as [18], the mean firing rate $\bar{\nu}$ can be calculated as

$$\bar{\nu} = \left[p_{\text{edge}} + \frac{N-1}{2}p + \frac{m}{N}(m-N)p\right]\nu_{\max}, \tag{82}$$

which reaches its maximum when $m=0$ or $m=N$ (a homogeneous ON or homogeneous OFF population), and reaches its minimum when $m \to N/2$ [an equally mixed ON-OFF population, Fig. 8(a)].

Since the maximal mutual information is the same for homogeneous populations (all ON or all OFF) and all ON-OFF mixtures, the information per spike is the highest for equal ON-OFF mixtures and the lowest for homogeneous populations with only ON or OFF neurons [Fig. 8(b)].

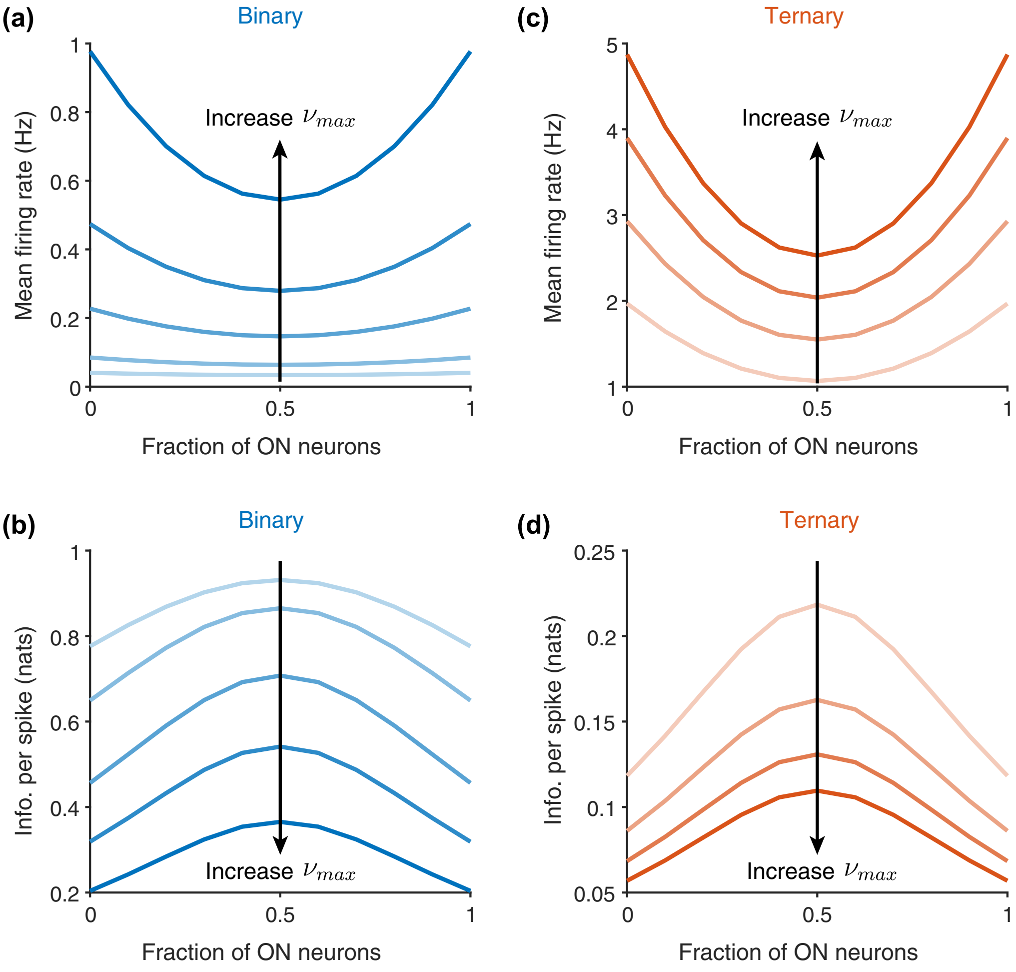

FIG. 8. Information per spike is the highest for equal ON-OFF mixtures. (a) The total mean firing rate $\bar{\nu}$ of neurons in populations with different fractions of ON neurons (m/N). (b) Information per spike ($I/(\bar{\nu}T)$) of neuron populations of different ON neuron ratios (m/N). In (a) and (b), $T = 1$ s and different line colors correspond to $\nu_{\max} = 0.1,\ 0.2,\ 0.5,\ 1,\ 2$. (c) Same as (a) but for ternary neurons. (d) Same as (b) but for ternary neurons. In (c) and (d), $\nu_{\max} = 4,\ 6,\ 8,\ 10$. Population size is $N = 10$ for all the panels.

With ternary neurons, the average firing rate $\bar{\nu}$ in the population with optimal thresholds becomes (see the SM [31])

$$\bar{\nu} = \bigg[p_{\text{edge}} + p_1 f + \frac{N-1}{2}(p_1 + p_2) + \frac{m}{N}(m - N)(p_1 + p_2)\bigg]\nu_{\max}. \tag{83}$$

And generalizing this to $(M+1)$-ary neurons, the mean firing rate of the population $\bar{\nu}$ becomes

$$\bar{\nu} = \bigg[p_{\text{edge}} + \frac{N-1}{2}\sum_{k=1}^{M} p_k + \frac{m}{N}(m - N)\sum_{k=1}^{M} p_k + \sum_{k=1}^{M-1} p_k f_k\bigg]\nu_{\max}. \tag{84}$$

As in the population with binary neurons, also here the highest mean firing rate is achieved for the purely ON or OFF populations with $m = 0$ or $m = N$ and the minimum is reached when $m \to N/2$ (see the SM [31]). Thus, similar to binary or ternary neurons, the information per spike is the highest in the equally mixed ON and OFF population, i.e., when $m \to N/2$.

When the maximal firing rates are heterogeneous, the average firing rate $\bar{\nu}$ does not only depend on the ratio of ON vs OFF neurons, but also depends on the order of neurons, i.e., the shuffling of thresholds in the population. We can demonstrate that for binary neurons (see the SM [31]): (1) Homogeneous ON or OFF populations with only ON or OFF cells generate the highest $\bar{\nu}$. (2) To reach the lowest $\bar{\nu}$, neurons with higher maximal firing rates should have smaller probability to fire (smaller u_i). Mathematically, given the predefined order of thresholds [Eq. (37)], that is,

$$\theta_{m+1} < \cdots < \theta_N < \theta_m < \cdots < \theta_1,$$

there should be

$$\begin{aligned} &\nu_{\max,1} \geqslant \cdots \geqslant \nu_{\max,m} \\ &\nu_{\max,m+1} \geqslant \cdots \geqslant \nu_{\max,N}. \end{aligned} \tag{85}$$

As an intuitive example, in Fig. 3, $\bar{\nu}$ is lower in panel (c) than in (d), and also lower in (a) compared to (b). (3) When the lowest $\bar{\nu}$ is reached, the number of ON neurons m

should satisfy

$$\frac{\sum_{i=m+1}^{N}\log q_i}{\sum_{i=1}^{m-1}(1-q_i)q_i^{q_i/(1-q_i)}} \simeq \frac{\log q_m}{(1-q_m)q_m^{q_m/(1-q_m)}}. \tag{86}$$

Point (3) can be understood as the balance between the activity of ON and OFF neurons. $|\log q_i|$ and $(1-q_i)q_i^{q_i/(1-q_i)}$ both decrease with q_i, and thus, increase with R_i. Therefore, the denominator of the left side of Eq. (86), $\sum_{i=1}^{m-1}(1-q_i)q_i^{q_i/(1-q_i)}$, is approximately proportional to the total firing rate of ON neurons, that is $\sum_{i=1}^{m}\nu_{\max,i}$. Similarly, the numerator $\sum_{i=m+1}^{N}\log q_i$ is approximately proportional to the total firing rate of OFF neurons, i.e., $\sum_{i=m+1}^{N}\nu_{\max,i}$. Therefore, Eq. (86) provides the optimal firing rate ratio of ON vs OFF neurons to minimize the mean firing rate of the mixed population. As a special case, when the maximal firing rates are identical for all cells, i.e., $q_i = q$, Eq. (86) degenerates into $m \simeq (N+1)/2$, which is consistent with the optimal m in Eqs. (82)–(84) above (also see Fig. 8).

In sum, for any shape of activation functions, although the maximal mutual information is independent of the numbers of ON and OFF cells, equal ON-OFF mixtures are more efficient in that their mean firing rate is the lowest, hence producing the highest information per spike. In the case of heterogeneous maximal firing rates, it is more efficient if neurons with higher maximal firing rates fire with lower probabilities. Intuitively, the mean firing rate of a neuron population can be related to the overlap of the neurons' high firing states—when the overlap is bigger, the neurons fire simultaneously, increasing the metabolic cost.

IV. DISCUSSION

Many neural systems process information using multiple neurons in parallel to encode a sensory stimulus in the presence of different biological constraints. In this paper, we have developed an efficient coding theory that maximizes the Shannon mutual information between stimuli and neuronal spikes subject to neuronal noise. We considered several aspects of the neuronal populations to make them more biologically realistic, including ON and OFF neurons, different shapes of neuronal activation functions and a heterogeneity of maximal firing rates, different statistics of noise in the spike generation mechanism, and different numbers of neurons.

We assumed that each neuron has a zero spontaneous firing rate and a fixed maximal firing rate, which can be interpreted as the inverse of the noise level. Generalizing previous results beyond Poisson spiking to any noise statistics, we first showed that the optimal activation function of a single neuron that maximizes the mutual information must be discrete. Additionally, we demonstrated that also in a population of any number of neurons and any noise statistics, the optimal activation functions are all discrete. Intuitively, discrete activation functions could be advantageous because they create distinct firing levels that are more distinguishable in the presence of spiking noise, compared to continuous activation functions. As the maximal firing rate increases, the different firing levels become easier to separate, making room for more steps in the activation functions. Interestingly, for a population of fixed size with identical maximal firing rates for all cells, we found that the number of steps in the population increases with increasing the maximal firing rate constraint (or equivalently, decreasing noise level) simultaneously for all the neurons. This implies that at any noise level, all neurons have the same activation function shape, i.e., an activation function that consists of the same number of steps. When the maximal firing rates of the different cells are heterogeneous, the number of steps in the optimal activation function of each neuron depends on the maximal firing rate of that neuron only. Therefore, neurons in the same population can have different activation function shape.

A second aspect of optimal coding pertains to the optimal activation functions distribution of all neurons in stimulus space. Remarkably, when the maximal firing rates of all cells are the same, the optimal activation functions divide the probability space of the stimulus into equal regions, hence implementing a coding strategy, which emphasizes stimuli that occur with higher probability. While this result has been long known in the limit of low noise (high maximal firing rate) as "histogram equalization" [4], we show that it holds for any noise generation function and any amount of noise. When the maximal firing rates of the different cells are heterogeneous, the organization of thresholds is no longer regular but rather results in smaller (larger) stimulus regions when the firing rates are low (high) to compensate for the higher (lower) noise.

Finally, by considering populations with different proportions of ON and OFF neurons, we demonstrated that the maximal information is independent of the ON-OFF mixture. ON and OFF splitting of a sensory signals has been experimentally observed in various sensory systems, including the mammalian retina [10,11], the medulla of *Drosophila* visual system [36,37], and the mechanosensory neurons in the legs of the adult *Drosophila* [12]. Our theory applies only to populations, which code for the same one-dimensional stimulus variable using just the spike count. However, in the vertebrate retina, diverse types of retinal ganglion cells represent a visual stimulus, and differ in their spatial and temporal processing characteristics [38,39]. Nonetheless, certain types of ganglion cells exist both as ON and OFF, with otherwise similar spatial location and temporal processing features, and hence are consistent with the assumptions in our theory. Considering a theoretical framework for the encoding of a scalar stimulus, which has no spatial or temporal correlations is clearly a limitation of our model, since real-world stimuli possess rich temporal and spatial patterns [11,40–42]. Several frameworks have begun exploring the impact of these complex sensory stimuli on efficient coding [43–45]. For example, mainly using the mammalian retina as a model, the optimal numbers of ON and OFF retinal ganglion cells were derived as a function of the spatial contrast statistics in these natural scenes [11]. Other studies have built more extensive cascade models consisting of receptive fields and nonlinearities, and predicted the structure of retinal mosaics (spatial grids of cells that encode different sensory information) [46,47] and intermosaic relationships [48,49] by considering the spatial structures of natural scenes. To account for the temporally rich dynamics, some models have extended similar cascade models to the temporal domain and predicted the spatial and temporal integration properties of retinal mosaics when training the models on natural videos [48,50,51] or studied adaptation to a changing environment [52,53]. While these are exciting recent developments that capitalize on new technologies to simultaneously record many cells and evaluate the theoretical predictions, they are often focused on particular systems, such as the vertebrate retina. Considering its applicability to the encoding of a one-dimensional static stimulus, our framework allows analytical tractability, and generalization across different dimensions, including noise functions, firing rate distributions, and activation functions, as well as being agnostic to the neural populations being modeled. Our model would need to be revised to account for the encoding of high-dimensional spatial and temporal stimuli.

In addition to sensory stimuli, neural activity also exhibits a high degree of structure, for example, firing with specific temporal dynamics (sustained or transient [10,54]). Structured neural activity significantly impacts neural coding and could alter the conclusions of specific efficient coding theoretical results. For instance, noise correlations can significantly affect coding performance [55,56]. Bursting, another common and functional type of neural activity other than spiking, also has profound coding implications [57]. Burst coding can be more efficient [58], more biological realistic [59], and more multiplexed [60] than rate coding.

The equality of mutual information across different ON-OFF mixtures predicted by our theoretical framework seems to be inconsistent with experimental data in the many sensory systems where different numbers of ON and OFF neurons have been observed (e.g., primate somatosensory cortex, fly visual system, and mammalian olfactory system, summarized in [18]). Nonetheless, in our theoretical results, despite the equality of total information the overall cost of spiking across the different ON-OFF mixtures differs. Hence, populations with equal ON-OFF mixtures have the lowest spiking cost, while homogeneous populations with only ON or OFF neurons have the highest cost. This implies despite the equality of information across the different ON-OFF mixtures, the information per spike is the highest for the equal ON-OFF mixture, and decreases monotonically when a single subpopulation (ON or OFF) begins to dominate the mixture. Taking into account the metabolic cost could be one answer to explaining different numbers of ON and OFF cells across different sensory systems. Other possibilities include different measures of coding efficiency than mutual information that might be used by the different sensory systems [9,25], different sources of noise (input vs output [14,61,62]), as well as including spatiotemporal statistics in natural images.

Although we considered different noise statistics, we assumed that the noise was inherent in the spike generation mechanism. Noise can enter in different places along the encoding pathway, for example, in the activation functions of the neurons [13], and have considerable effects on neuronal coding [14,62–64]. Furthermore, we only focused on stimulus encoding without discussing how the information can be further decoded from the neuronal spikes. The decoding of information could be another key aspect of efficient neuronal coding [18,28,43,65]. Previous study has shown that the measure of efficiency can significantly influence conclusions about the ON-OFF composition of the population and the activation function distribution [18,20,62]. It is possible that different sensory systems have evolved optimal coding strategies by maximizing different objective functions, or may not even be optimal at all. Our paper provides a general theoretical framework that in principle can be applied to various sensory systems without the nuances of each system.

ACKNOWLEDGMENTS

This work is supported by the funding from the Max Planck Society (S.S. and J.G.), the Technical University of Munich (J.G.), and NIH Grant No. R01 NS111477 (M.M.). We also would like thank the "Computation in Neural Circuits" group for comments on this work, especially our colleagues Claudia Cusseddu and Judith Parkinson-Schwarz for providing feedback on the manuscript.

APPENDIX

The variables used in this paper and their definitions are listed in Table I below.

TABLE I. Summary of the parameters in this paper.

Variables	Definitions
N	Number of neurons
s	The scalar stimulus
$p(s)$	The stimulus distribution
$\nu_{\max}$	The maximal firing rate constraint
$\nu_i(s)$	The activation function of neuron i
T	The coding window within which the spike count is computed
R	$\nu_{\max}T$, the expected value of spike count when a neuron fires at the maximal firing rate $\nu_{\max}$
n_i	The spike count of neuron i
$\nu_i T$	The expected value of neuron i's spike count
$L(n_i, \nu_i T)$	The noise generation function, equivalent to $p(n_i\|\nu_i)$
$I(s, n)$	Shannon mutual information between stimulus s and spike count n encoded by the neuronal population
$i(\cdot)$	Density of mutual information
F_ν	The cumulative distribution function of the firing rate ν
m	The number of ON neurons
Q_m	The probability that none of the ON neurons $1, 2, \ldots, m$ fires
$F_i^{(j)}$	cumulative distribution function of ν_i, given that none of the neurons $1, 2, \ldots, j$ fires
θ_{ik}	The kth firing threshold of neuron i whose activation function is discrete
p_{ik}	The kth stimulus interval of neuron i
u_{ik}	The kth cumulative stimulus interval of neuron i
f_{ik}	The ratio of the kth intermediate firing rate
$P_i(0)$	The probability that neuron i does not fire
$u_{ik}^{(j,N)}$	In a group of N neurons, the kth cumulative stimulus interval of neuron i, given that none of the neurons $1, \ldots, j$ $(j < i)$ fires
$P_i^{(k,N)}(0)$	In a group of N neurons, the kth cumulative stimulus interval of neuron i, given that none of the neurons $1, \ldots, j$ $(j < i)$ fires
q	$L(0, \nu T)$, the probability that the spike count is 0 given a nonzero firing rate

[1] S. B. Laughlin, Efficiency and complexity in neural coding, in *Novartis Found Symp: Complexity in Biological Information Processing* (Willey Online Library, Chichester, 2001), Vol. 239, pp. 177–192.

[2] S. Schreiber, C. K. Machens, A. V. Herz, and S. B. Laughlin, Energy-efficient coding with discrete stochastic events, Neural Comput. **14**, 1323 (2002).

[3] J. J. Atick and A. N. Redlich, Towards a theory of early visual processing, Neural Comput. **2**, 308 (1990).

[4] S. Laughlin, A simple coding procedure enhances a neuron's information capacity, Z. Naturforsch. C **36**, 910 (1981).

[5] J. H. van Hateren, Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation, J. Comp. Physiol. A **171**, 157 (1992).

[6] A. L. Hodgkin and A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J. Physiol. **117**, 500 (1952).

[7] S. Weiler, D. Guggiana Nilo, T. Bonhoeffer, M. Hübener, T. Rose, and V. Scheuss, Orientation and direction tuning align with dendritic morphology and spatial connectivity in mouse visual cortex, Curr. Biol. **32**, 1743 (2022).

[8] F. L. Wightman and D. J. Kistler, The dominant role of low-frequency interaural time differences in sound localization, J. Acoust. Soc. Am. **91**, 1648 (1992).

[9] N. S. Harper and D. McAlpine, Optimal neural population coding of an auditory spatial cue, Nature (London) **430**, 682 (2004).

[10] K. Franke, P. Berens, T. Schubert, T. Euler, M. Bethge, and T. Baden, Inhibition decorrelates visual feature representations in the inner retina, Nature (London) **542**, 439 (2017).

[11] 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. Acad. Sci. USA **107**, 17368 (2010).

[12] A. Mamiya, P. Gurung, and J. C. Tuthill, Neural coding of leg proprioception in *Drosophila*, Neuron **100**, 636 (2018).

[13] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, Noise in the nervous system, Nat. Rev. Neurosci. **9**, 292 (2008).

[14] K. Röth, S. Shao, and J. Gjorgjieva, Efficient population coding depends on stimulus convergence and source of noise, PLoS Comput. Biol. **17**, e1008897 (2021).

[15] S. Shamai, Capacity of a pulse amplitude modulated direct detection photon channel, IEE Proc. Commun. Speech Vis. **137**, 424 (1990).

[16] A. P. Nikitin, N. G. Stocks, R. P. Morse, and M. D. McDonnell, Neural population coding is optimized by discrete tuning curves, Phys. Rev. Lett. **103**, 138101 (2009).

[17] X. Pitkow and M. Meister, Decorrelation and efficient coding by retinal ganglion cells, Nat. Neurosci. **15**, 628 (2012).

[18] J. Gjorgjieva, M. Meister, and H. Sompolinsky, Functional diversity among sensory neurons from efficient coding principles, PLoS Comput. Biol. **15**, e1007476 (2019).

[19] A. M. Taberner and M. C. Liberman, Response properties of single auditory nerve fibers in the mouse, J. Neurophysiol. **93**, 557 (2005).

[20] 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 (2015).

[21] M. J. Berry, D. K. Warland, and M. Meister, The structure and precision of retinal spike trains, Proc. Natl. Acad. Sci. USA **94**, 5411 (1997).

[22] P. Kara, P. Reinagel, and R. C. Reid, Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons, Neuron **27**, 635 (2000).

[23] H. S. Seung and H. Sompolinsky, Simple models for reading neuronal population codes, Proc. Natl. Acad. Sci. USA **90**, 10749 (1993).

[24] D. K. Warland, P. Reinagel, and M. Meister, Decoding visual information from a population of retinal ganglion cells, J. Neurophysiol. **78**, 2336 (1997).

[25] M. Bethge, D. Rotermund, and K. Pawelzik, Optimal neural rate coding leads to bimodal firing rate distributions, Netw. Comput. Neural Syst. **14**, 303 (2003).

[26] F. Rieke, D. Warland, R. R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, 1997).

[27] W. Bialek, R. R. de Ruyter van Steveninck, and D. Warland, Reading a neural code, Science **252**, 1854 (1991).

[28] J. Gjorgjieva, H. Sompolinsky, and M. Meister, Benefits of pathway splitting in sensory coding, J. Neurosci. **34**, 12127 (2014).

[29] 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 (London) **454**, 995 (2008).

[30] V. J. Uzzell and E. J. Chichilnisky, Precision of spike trains in primate retinal ganglion cells, J. Neurophysiol. **92**, 780 (2004).

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.043205 for the proof of Eq. (5); the proof of Eq. (8); the proof that the optimal activation functions of a population of neurons are discrete; numerical simulations that demonstrates the number of steps in the optimal activation functions increases with the maximal firing rate; detailed calculations for Sec. III C (a population of binary neurons); detailed calculations for Sec. III D (a population of binary neurons with heterogeneous maximal firing rates); detailed calculations for Sec. III F (a population of ternary neurons); detailed calculations for Sec. III G (a population of neurons with any shapes of activation functions); detailed calculations for Sec. III H (a population of neurons with activation functions of any shape and heterogeneous maximal firing rates).

[32] J. G. Smith, The information capacity of amplitude- and variance-constrained scalar Gaussian channels, Inf. Control **18**, 203 (1971).

[33] R. G. Bartle, *The Elements of Real Analysis* (John Wiley & Sons, Hoboken, NJ, 1964).

[34] K. Knopp, *Theory of Functions* (Dover Publications, New York, 1945).

[35] H.-A. Shen, S. Moser, and J.-P. Pfister, A generalization of the equal coding theorem, in *2023 IEEE Information Theory Workshop (ITW), Saint-Malo, France* (IEEE, PIscataway, NJ, 2023), pp. 329–334.

[36] D. A. Clark, L. Bursztyn, M. A. Horowitz, M. J. Schnitzer, and T. R. Clandinin, Defining the computational structure of the motion detector in *Drosophila*, Neuron **70**, 1165 (2011).

[37] M. D. Ketkar, B. Gür, S. Molina-Obando, M. Ioannidou, C. Martelli, and M. Silies, First-order visual interneurons

distribute distinct contrast and luminance information across on and off pathways to achieve stable behavior, eLife **11**, e74937 (2022).

[38] J. R. Sanes and R. H. Masland, The types of retinal ganglion cells: Current status and implications for neuronal classification, Annu. Rev. Neurosci. **38**, 221 (2015).

[39] T. Baden, P. Berens, K. Franke, M. Román Rosón, M. Bethge, and T. Euler, The functional diversity of retinal ganglion cells in the mouse, Nature (London) **529**, 345 (2016).

[40] C. Kayser, W. Einhäuser, and P. König, Temporal correlations of orientations in natural scenes, Neurocomputing **52-54**, 117 (2003).

[41] D. W. Dong and J. J. Atick, Statistics of natural time-varying images, Netw. Comput. Neural Syst. **6**, 345 (1995).

[42] K. D. Simmons, J. S. Prentice, G. Tkačik, J. Homann, H. K. Yee, S. E. Palmer, P. C. Nelson, and V. Balasubramanian, Transformation of stimulus correlations by the retina, PLoS Comput. Biol. **9**, e1003344 (2013).

[43] M. Park and J. W. Pillow, Bayesian inference for low rank spatiotemporal neural receptive fields, in *Advances in Neural Information Processing Systems*, edited by C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (Curran Associates, Inc., 2013), Vol. 26.

[44] D. Ganguli and E. P. Simoncelli, Efficient sensory encoding and Bayesian inference with heterogeneous neural populations, Neural Comput. **26**, 2103 (2014).

[45] T. Tesileanu, M. M. Conte, J. J. Briguglio, A. M. Hermundstad, J. D. Victor, and V. Balasubramanian, Efficient coding of natural scene statistics predicts discrimination thresholds for grayscale textures, eLife **9**, e54347 (2020).

[46] Y. Karklin and E. Simoncelli, Efficient coding of natural images with a population of noisy linear-nonlinear neurons, in *Advances in Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (Curran Associates, Inc., 2011), Vol. 24.

[47] E. Doi, J. L. Gauthier, G. D. Field, J. Shlens, A. Sher, M. Greschner, T. A. Machado, L. H. Jepson, K. Mathieson, D. E. Gunning *et al.*, Efficient coding of spatial information in the primate retina, J. Neurosci. **32**, 16256 (2012).

[48] N. Y. Jun, G. Field, and J. Pearson, Efficient coding, channel capacity, and the emergence of retinal mosaics, in *Advances in Neural Information Processing Systems*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022), Vol. 35, pp. 32311–32324.

[49] S. Roy, N. Y. Jun, E. L. Davis, J. Pearson, and G. D. Field, Intermosaic coordination of retinal receptive fields, Nature (London) **592**, 409 (2021).

[50] S. Ocko, J. Lindsey, S. Ganguli, and S. Deny, The emergence of multiple retinal cell types through efficient coding of natural movies, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018), Vol. 31.

[51] A. J. Sederberg, J. N. MacLean, and S. E. Palmer, Learning to make external sensory stimulus predictions using internal correlations in populations of neurons, Proc. Natl. Acad. Sci. USA **115**, 1105 (2018).

[52] W. F. Młynarski and A. M. Hermundstad, Adaptive coding for dynamic sensory inference, eLife **7**, e32055 (2018).

[53] W. Młynarski, M. Hledík, T. R. Sokolowski, and G. Tkačik, Statistical analysis and optimality of neural systems, Neuron **109**, 1227 (2021).

[54] D. J. Margolis and P. B. Detwiler, Different mechanisms generate maintained activity in ON and OFF retinal ganglion cells, J. Neurosci. **27**, 5994 (2007).

[55] B. B. Averbeck, P. E. Latham, and A. Pouget, Neural correlations, population coding and computation, Nat. Rev. Neurosci. **7**, 358 (2006).

[56] S. Panzeri, M. Moroni, H. Safaai, and C. D. Harvey, The structures and functions of correlations in neural population codes, Nat. Rev. Neurosci. **23**, 551 (2022).

[57] F. Zeldenrust, W. J. Wadman, and B. Englitz, Neural coding with bursts—Current state and future perspectives, Front. Comput. Neurosci. **12**, 48 (2018).

[58] S. Park, S. Kim, H. Choe, and S. Yoon, Fast and efficient information transmission with burst spikes in deep spiking neural networks, in *Proceedings of the 56th Annual Design Automation Conference, 2019 (DAC '19)* (Association for Computing Machinery, New York, 2019), Article 53, pp. 1–6.

[59] E. Williams, A. Payeur, A. Gidon, and R. Naud, Neural burst codes disguised as rate codes, Sci. Rep. **11**, 15910 (2021).

[60] R. Naud and H. Sprekeler, Sparse bursts optimize information transmission in a multiplexed neural code, Proc. Natl. Acad. Sci. USA **115**, E6329 (2018).

[61] D. B. Kastner and S. A. Baccus, Coordinated dynamic encoding in the retina using opposing forms of plasticity, Nat. Neurosci. **14**, 1317 (2011).

[62] 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 Comput. Biol **12**, e1005150 (2016).

[63] F. Rieke, D. A. Bodnar, and W. Bialek, Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents, Proc. R. Soc. London B **262**, 259 (1995).

[64] E. J. Chichilnisky and F. Rieke, Detection sensitivity and temporal resolution of visual signals near absolute threshold in the salamander retina, J. Neurosci. **25**, 318 (2005).

[65] Z. Wang, A. A. Stocker, and D. D. Lee, Efficient neural codes that minimize lp reconstruction error, Neural Comput. **28**, 2656 (2016).