Supplemental Material
Efficient population coding of sensory stimuli

Shuai Shao[1,2], Markus Meister[2] and Julijana Gjorgjieva[1,2]

1. Computation in Neural Circuits Group, Max Planck Institute for Brain Research, Frankfurt, Germany
2. Donders Institute and Faculty of Science, Radboud University, Nijmegen, Netherlands
3. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
4. School of Life Sciences, Technical University of Munich, Freising, Germany

Contact: gjorgjieva@tum.de

1 The mutual information between stimulus and spikes equals the mutual information between firing rates and spikes

In this section we prove the argument in the main text that the mutual information between the stimuli s and the spike counts $\vec{n}$, equals the mutual information between the firing rates $\vec{\nu}$ and the spike counts $\vec{n}$, i.e., $I(s, \vec{n}) = I(\vec{\nu}, \vec{n})$ (Eq. 5). This was also shown in previous literature [1] but limited to a single neuron (i.e., $N = 1$, when $\vec{n}$ and $\vec{\nu}$ are scalars).

Since the spike counts of different neurons are independent of each other, we can write

$$p(\vec{n}|s) = \prod_i p(n_i|s). \tag{S1.1}$$

Inserting it into the formula of the Mutual Information (Eq. 3), we have

$$
\begin{aligned}
I(s, \vec{n}) &= \sum_{\vec{n}} \int \mathrm{d}s\, p(s)\, p(\vec{n}|s) \log \frac{p(\vec{n}|s)}{P(\vec{n})} \\
&= \sum_{\vec{n}} \int \mathrm{d}s\, p(s)\, p(\vec{n}|s) \log p(\vec{n}|s) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_{\vec{n}} \sum_i \int \mathrm{d}s\, p(s) \log p(n_i|s) \prod_k p(n_k|s) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_i \sum_{\vec{n}} \int \mathrm{d}s\, p(s) \log p(n_i|s) \prod_k p(n_k|s) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}).
\end{aligned}
\tag{S1.2}
$$

Note that $\prod_k p(n_k|s) = p(n_i|s) \prod_{k \neq i} p(n_k|s)$, summing over all $k \neq i$, we have

$$
\begin{aligned}
I(s, \vec{n}) &= \sum_i \sum_{n_i} \int \mathrm{d}s\, p(s)\, p(n_i|s) \log p(n_i|s) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_i \sum_{n_i} \int \mathrm{d}\nu_i\, p(\nu_i)\, p(n_i|\nu_i) \log p(n_i|\nu_i) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}).
\end{aligned}
\tag{S1.3}
$$

Denoting by $\vec{\nu}^{(i)} = (\nu_1, ..., \nu_{i-1}, \nu_{i+1}, ..., \nu_N)$ the vector of all the ν except for ν_i, we have

$$\int \mathrm{d}^{N-1}\vec{\nu}^{(i)}\, p(\vec{\nu}^{(i)}|\nu_i) = 1. \tag{S1.4}$$

Therefore, equation (Eq. S1.3) becomes

$$
\begin{aligned}
I(s, \vec{n}) &= \sum_i \sum_{n_i} \int \mathrm{d}^{N-1}\vec{\nu}^{(i)}\, p(\vec{\nu}^{(i)}|\nu_i) \int \mathrm{d}\nu_i\, p(\nu_i)\, p(n_i|\nu_i) \log p(n_i|\nu_i) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_i \sum_{n_i} \int \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu})\, p(n_i|\nu_i) \log p(n_i|\nu_i) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_i \sum_{n_i} \int \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu})\, p(n_i|\vec{\nu}) \log p(n_i|\vec{\nu}) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_i \sum_{\vec{n}} \int \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu}) \log p(n_i|\vec{\nu}) \prod_k p(n_k|\vec{\nu}) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}).
\end{aligned}
\tag{S1.5}
$$

Similar to Eq. S1.1, we also have

$$
p(\vec{n}|\vec{\nu}) = \prod_i p(n_i|\vec{\nu})
\tag{S1.6}
$$

which leads to

$$
\begin{aligned}
I(s, \vec{n}) &= \sum_i \sum_{\vec{n}} \int \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu})\, p(\vec{n}|\vec{\nu}) \log p(n_i|\vec{\nu}) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_{\vec{n}} \int \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu})\, p(\vec{n}|\vec{\nu}) \sum_i \log p(n_i|\vec{\nu}) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_{\vec{n}} \int \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu})\, p(\vec{n}|\vec{\nu}) \log p(\vec{n}|\vec{\nu}) - \sum_{\vec{n}} P(\vec{n}) \log P(\vec{n}) \\
&= \sum_{\vec{n}} \int_{\vec{\nu}} \mathrm{d}^{N}\vec{\nu}\, p(\vec{\nu})\, p(\vec{n}|\vec{\nu}) \log \frac{p(\vec{n}|\vec{\nu})}{P(\vec{n})} \\
&= I(\vec{\nu}, \vec{n}).
\end{aligned}
\tag{S1.7}
$$

2 Density of the mutual information for a single neuron is constant when optimized

This section works as a first step to prove that the optimal activation function for a single neuron is discrete. We limit our discussion to a single neuron. Without loss of generality, we only consider an ON neuron with an activation function where the firing rate increases with stimulus intensity. Studying an OFF neuron is entirely symmetric. As in the main text, the maximal firing rate of this neuron is constrained to ν_{max} and the spontaneous firing rate is denoted by ν_0.

2.1 A neuron with a discrete activation function

First, we consider a neuron with a discrete activation function. In this case, the firing rate can only be some discrete values between ν_0 and $\nu_{\max}$. Therefore, we denote the probability that the firing rate is ν by p_ν, instead of $p(\nu)$ that we commonly write. The mutual information is then

$$
I(s, n) = I(\nu, n) = \sum_n \sum_\nu p_\nu\, p(n|\nu) \log \frac{p(n|\nu)}{P(n)}.
\tag{S2.1}
$$

We can define the entropy of the spike count at a given firing rate as

$$
h(\nu) = -\sum_{n=0}^{+\infty} p(n|\nu) \log p(n|\nu)
\tag{S2.2}
$$

and note that

$$
P(n) = \sum_\nu p_\nu\, p(n|\nu),
\tag{S2.3}
$$

so that we have

$$I(\nu, n) = -\sum_\nu p_\nu \, h(\nu) - \sum_\nu p_\nu \sum_n p(n|\nu) \log P(n). \tag{S2.4}$$

Since p_ν are probabilities, we have the constraint that $\sum_\nu p_\nu = 1$, hence to optimize the objective function we include a Lagrange multiplier,

$$\widetilde{I} = I(\nu, n) + \lambda(\sum_\nu p_\nu - 1). \tag{S2.5}$$

Assuming optimality,

$$\partial_{p_\nu} \widetilde{I} = -h(\nu) - \sum_n p(n|\nu) \log P(n) - \sum_n p(n|\nu) + \lambda = 0. \tag{S2.6}$$

Absorbing $-\sum_n p(n|\nu) = -1$ into λ, i.e. $\lambda \to \lambda - 1$, we have

$$-h(\nu) - \sum_n p(n|\nu) \log P(n) + \lambda = 0. \tag{S2.7}$$

Multiplying both sides by p_ν and summing over ν, we have

$$I(s, n) + \lambda = 0. \tag{S2.8}$$

We define

$$i(\nu) = \sum_n p(n|\nu) \log \frac{p(n|\nu)}{P(n)}. \tag{S2.9}$$

Multiplying this equation with p_ν and summing over ν, we have

$$I = \sum_\nu p_\nu i(\nu). \tag{S2.10}$$

Therefore, we call $i(\nu)$ "the density of mutual information", which is also defined by Eq. 6 in the main text. According to Eq. S2.7, we can write

$$I(\nu, n) = -\lambda = -\sum_n p(n|\nu) \log P(n) - h(\nu) = \sum_n p(n|\nu) \log \frac{p(n|\nu)}{P(n)} = i(\nu). \tag{S2.11}$$

This means when the mutual information is optimized, $i(\nu)$ is a constant for all possible ν. The convexity of the mutual information ensures that the optimal solution is unique. As a special case, if the spontaneous rate $\nu_0 = 0$, according to Eq. 1 and Eq. 2, we have $p(n = 0|\nu = 0) = 1$ and $p(n \neq 0|\nu = 0) = 0$. As a result,

$$I^{\max} = i(\nu = 0) = -\log P(0). \tag{S2.12}$$

Also, Eq. S2.11 means the mutual information $I(\nu, n)$ is distributed proportionally to the probabilities p_ν when it is maximized. In addition, one can also define

$$i_s(s) = \sum_n p(n|s) \log \frac{p(n|s)}{P(n)}, \tag{S2.13}$$

then we have

$$I = \int \mathrm{d}s \, p(s) \, i_s(s) \tag{S2.14}$$

and

$$i(\nu) = i_s(s). \tag{S2.15}$$

Therefore, the maximal mutual information $I(s, n)$ will be distributed proportionally to the probability density of the stimulus s, denoted by $p(s)$ in the main text and Fig. 1B. The density function $i_s(s)$ is also a constant over the space of stimulus s. For example, if we have a ternary activation function with three possible firing rates 0, $\nu_{max}/2$, and ν_{max}, and the stimulus s follows a standard normal distribution, the input space in terms of ν is $\{0, \nu_{max}/2, \nu_{max}\}$, so we have $i(\nu = 0) = i(\nu = \nu_{max}/2) = i(\nu = \nu_{max})$. Similarly, the input space of s is then the set of all real numbers $\mathbb{R}$, and we have $i_s(s) = \text{const}, s \in \mathbb{R}$.

2.2 A neuron with a continuous activation function

We assume that the neuron has a continuous and smooth (analytic) activation function, with the lowest rate (i.e., the spontaneous firing rate) ν_0 and the maximal firing rate ν_{max}. Then, the mutual information can be written as:

$$I = \sum_{n=0}^{+\infty} \int_{\nu_0}^{\nu_{max}} \mathrm{d}s \, p(s) \, p(n|s) \log \frac{p(n|s)}{P(n)} = \sum_{n=0}^{+\infty} \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, p(\nu) \, p(n|\nu) \log \frac{p(n|\nu)}{P(n)}. \tag{S2.16}$$

Define $\widetilde{I} = I + \lambda \left(\int_{\nu_0}^{\nu_{max}} p(\nu) \mathrm{d}\nu - 1 \right)$, then

$$\widetilde{I} = -\sum_{n=0}^{+\infty} P(n) \log P(n) + \sum_{n=0}^{+\infty} \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, p(n|\nu) \, p(\nu) \log p(n|\nu) + \lambda \left(\int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, p(\nu) - 1 \right). \tag{S2.17}$$

When optimized,

$$\delta\widetilde{I} = -\sum_{n=0}^{+\infty} (\log P(n) + 1) \, \delta P(n) - \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, h(\nu) \, \delta p(\nu) + \lambda \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, \delta p(\nu) = 0. \tag{S2.18}$$

Because

$$\delta P(n) = \delta \left[\int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, p(\nu) \, p(n|\nu) \right] = \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, p(n|\nu) \, \delta p(\nu), \tag{S2.19}$$

we have

$$\delta\widetilde{I} = -\int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \sum_{n=0}^{+\infty} p(n|\nu) \, (\log P(n) + 1) \, \delta p(\nu) - \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, h(\nu) \, \delta p(\nu) + \lambda \int_{\nu_0}^{\nu_{max}} \mathrm{d}\nu \, \delta p(\nu) = 0 \tag{S2.20}$$

which leads to

$$-\sum_{n=0}^{+\infty} p(n|\nu) \, (\log P(n) + 1) - h(\nu) + \lambda = 0. \tag{S2.21}$$

Absorbing $-\sum_{n=0}^{+\infty} p(n|\nu) = -1$ into λ, multiplying by $p(\nu)$, and integrating over ν, we have $I + \lambda = 0$. As a result,

$$I = -\lambda = -\sum_n p(n|\nu) \log P(n) - h(\nu) = \sum_n p(n|\nu) \log \frac{p(n|\nu)}{P(n)} = i(\nu), \quad \text{for } \nu \in [\nu_0, \nu_{max}] \tag{S2.22}$$

which means the density of mutual information $i(\nu)$ is a constant for all firing rates ν. One can still define the density function with stimulus s as Eq. S2.13 and the $i_s(s)$ is still a constant when the mutual information is optimized.

In summary, we have shown that the density of mutual information $i(\nu)$ is a constant for all possible firing rates, independent of whether the activation function is discrete or continuous. We note that this result has also been proven in previous work using a different approach based on the convexity of mutual information [2,3].

3 The optimal activation functions of a population of neurons are discrete

To prove that the optimal activation functions are discrete, we first need to prove that when the mutual information of a population of N neurons is maximized, the density of mutual information $\widetilde{i}(\nu_1)$ that we defined in the main text is a constant and equals to the maximal mutual information $I_N^{\max}$ (Eq. 35). Consistent with the main text, we denote $p(n_i|\nu_i)$ by $L(n_i, \nu_i T)$ from now on.

According to the definition in the main text (Eq. 34), we have

$$\widetilde{i}(\nu_1) = \sum_{n_1} p(n_1|\nu_1) \log \frac{p(n_1|\nu_1)}{P(n)} + p(n_1 = 0|\nu_1) I_{N-1}^{\max}$$

$$= \sum_{n_1} L(n_1, \nu_1 T) \log \frac{L(n_1, \nu_1 T)}{P(n)} + L(0, \nu_1 T) I_{N-1}^{\max} \tag{S3.1}$$

and we can see that when neurons $2, ..., N$ are all optimized,

$$\int d\nu_1 \, i(\nu_1) \, p(\nu_1) = I_N = I(F_1) + P_1(0) I_{N-1}^{\max}. \tag{S3.2}$$

Similar as in the last section, we define $\widetilde{I}_N = I_N + \lambda \left(\int_0^{\nu_{max}} p(\nu_1) d\nu_1 - 1 \right)$, and write

$$\widetilde{I}_N = - \sum_{n_1=0}^{+\infty} P(n_1) \log P(n_1) + \sum_{n_1=0}^{+\infty} \int_0^{\nu_{max}} d\nu_1 \, L(n_1, \nu_1 T) \, p(\nu_1) \log L(n_1, \nu_1 T)$$

$$+ I_{N-1}^{\max} \int_0^{\nu_{max}} d\nu_1 \, p(\nu_1) L(0, \nu_1 T) + \lambda \left(\int_0^{\nu_{max}} d\nu_1 \, p(\nu_1) - 1 \right). \tag{S3.3}$$

When optimized,

$$\delta \widetilde{I}_N = - \sum_{n_1=0}^{+\infty} \left(\log P(n_1) + 1 \right) \delta P(n_1) + \int_0^{\nu_{max}} d\nu_1 \left(I_{N-1}^{\max} L(0, \nu_1 T) - h(\nu_1) \right) \delta p(\nu_1) + \lambda \int_0^{\nu_{max}} d\nu_1 \, \delta p(\nu_1) = 0. \tag{S3.4}$$

Because

$$\delta P(n_1) = \delta \left[\int_0^{\nu_{max}} d\nu_1 \, p(\nu_1) \, L(n_1, \nu_1 T) \right] = \int_0^{\nu_{max}} d\nu_1 \, L(n_1, \nu_1 T) \, \delta p(\nu_1), \tag{S3.5}$$

we have

$$\delta \widetilde{I}_N = - \int_0^{\nu_{max}} d\nu_1 \sum_{n_1=0}^{+\infty} L(n_1, \nu_1 T) \left(\log P(n_1) + 1 \right) \delta p(\nu_1)$$

$$+ \int_0^{\nu_{max}} d\nu_1 \left(I_{N-1}^{\max} L(0, \nu_1 T) - h(\nu_1) \right) \delta p(\nu_1) + \lambda \int_0^{\nu_{max}} d\nu_1 \, \delta p(\nu_1) = 0 \tag{S3.6}$$

which leads to

$$- \sum_{n_1=0}^{+\infty} L(n_1, \nu_1 T) \left(\log P(n_1) + 1 \right) + I_{N-1}^{\max} L(0, \nu_1 T) - h(\nu_1) + \lambda = 0. \tag{S3.7}$$

Absorbing $- \sum_{n_1=0}^{+\infty} L(n_1, \nu_1 T) = -1$ into λ, multiplying by $p(\nu_1)$, and integrating over ν_1, we have $I_N + \lambda = 0$. As a result,

$$I_N^{\max} = -\lambda = - \sum_{n_1} L(n_1, \nu_1 T) \log P(n_1) + I_{N-1}^{\max} L(0, \nu_1 T) - h(\nu_1)$$

$$= \sum_{n_1} L(n_1, \nu_1 T) \log \frac{L(n_1, \nu_1 T)}{P(n_1)} + I_{N-1}^{\max} L(0, \nu_1 T) = \widetilde{i}(\nu_1), \quad \text{for } \nu_1 \in [0, \nu_{max}]. \tag{S3.8}$$

Second, we need to prove that the above result (Eq. S3.8) will lead to a contradiction if the optimal activation function F_1 is continuous. From the discussion in the main text, this is equivalent to finding a paradox in

$$\widetilde{i}(\nu_1) = \sum_{n_1=0}^{+\infty} L(n_1, \nu_1 T) \log \frac{L(n_1, \nu_1 T)}{P(n_1)} + I_{N-1}^{\max} L(0, \nu_1 T) = I_N^{\max} = \text{const.} \tag{S3.9}$$

Following the same procedure as in the main text, we can prove that if we write the Maclaurin series

$$L(n_1, \nu_1 T) = \sum_{k=1}^{+\infty} a_{n_1,k}(\nu_1 T)^k \tag{S3.10}$$

for any $n_1 \geq 1$, the sum of the coefficients of $\log(\nu T)$ terms in the mth derivative of $\widetilde{i}(\nu_1)$ is then $\sum_{n_1=1}^{+\infty} a_{n_1,m} j(n_1)$, where $j(n_1)$ is the minimal index of k that makes $a_{n_1,k} > 0$. This follows the same formalism as Eq. 22 in the main text, because the additional term here, $I_{N-1}^{\max} L(0, \nu_1 T)$, does not contribute to $\log(\nu T)$ terms when it is written as a Maclaurin series.

If Eq. S3.9 were correct, all the derivatives of $\widetilde{i}(\nu_1)$ would be 0, and we would have

$$\sum_{n_1=1}^{+\infty} a_{n_1,m} j(n_1) = 0 \text{ for any } m \geq 1 \tag{S3.11}$$

because $\log(\nu_1 T)$ diverges as $\nu_1 T \to 0$. Again, similar as in the main text, we could show that in this case,

$$L(n_1 = 0, \nu_1 T) = 1 \text{ for any } \nu_1, \tag{S3.12}$$

which cannot be true. Therefore, we have proved that the optimal F_1 being continuous will lead to a paradox. Therefore, in a population of N neurons, given that neurons $2, ..., N$ are all optimized, the optimal activation function of neuron 1 will be discrete.

4   The number of steps in the optimal activation functions increases as a function of the maximal firing rate constraint

Here, we perform extensive numerical calculations on neuronal populations with up to four neurons and any ON-OFF mixture to demonstrate that as the maximal firing rate constraint $\nu_{\max}$ increases, the number of steps in the optimal activation functions increases. We calculated the optimal thresholds numerically using three different noise generation functions $L(n, \nu T)$ (Fig. S1):
(1) Poisson distribution

$$L(n, \nu T) = \frac{(\nu T)^n}{n!} \exp(-\nu T), \tag{S4.1}$$

(2) Binomial distribution

$$L(n, \nu T) = \binom{N}{n} p^n (1 - p)^{N-n}, \tag{S4.2}$$

where $N = 30$ and $p = \nu T / N$, and
(3) Geometric distribution

$$L(n, \nu T) = p^n (1 - p) \tag{S4.3}$$

where $p = \nu T / (1 + \nu T)$.

In detail, we calculated the mutual information as a function of the firing thresholds and intermediate firing rates, based on Eq. 5. These parameters were initialized randomly, and then optimized using SLSQP method [4]. All code was written in Python 2.7. A sample is publicly available at https://zenodo.org/record/8083056.

We find that the number of thresholds increases as the maximal firing rate constraint $\nu_{\max}$ increases. Moreover, for all neurons in the same population, the threshold splitting occurs at the same firing rate, meaning that every neuron in the population has an optimal discrete activation function with the same number of steps. Hence, the optimal neuronal population consists of exclusively binary neurons, or exclusively ternary neurons, or exclusively quaternary neurons, etc. But it can never be a mixture of neurons with different numbers of steps, e.g., binary and ternary.
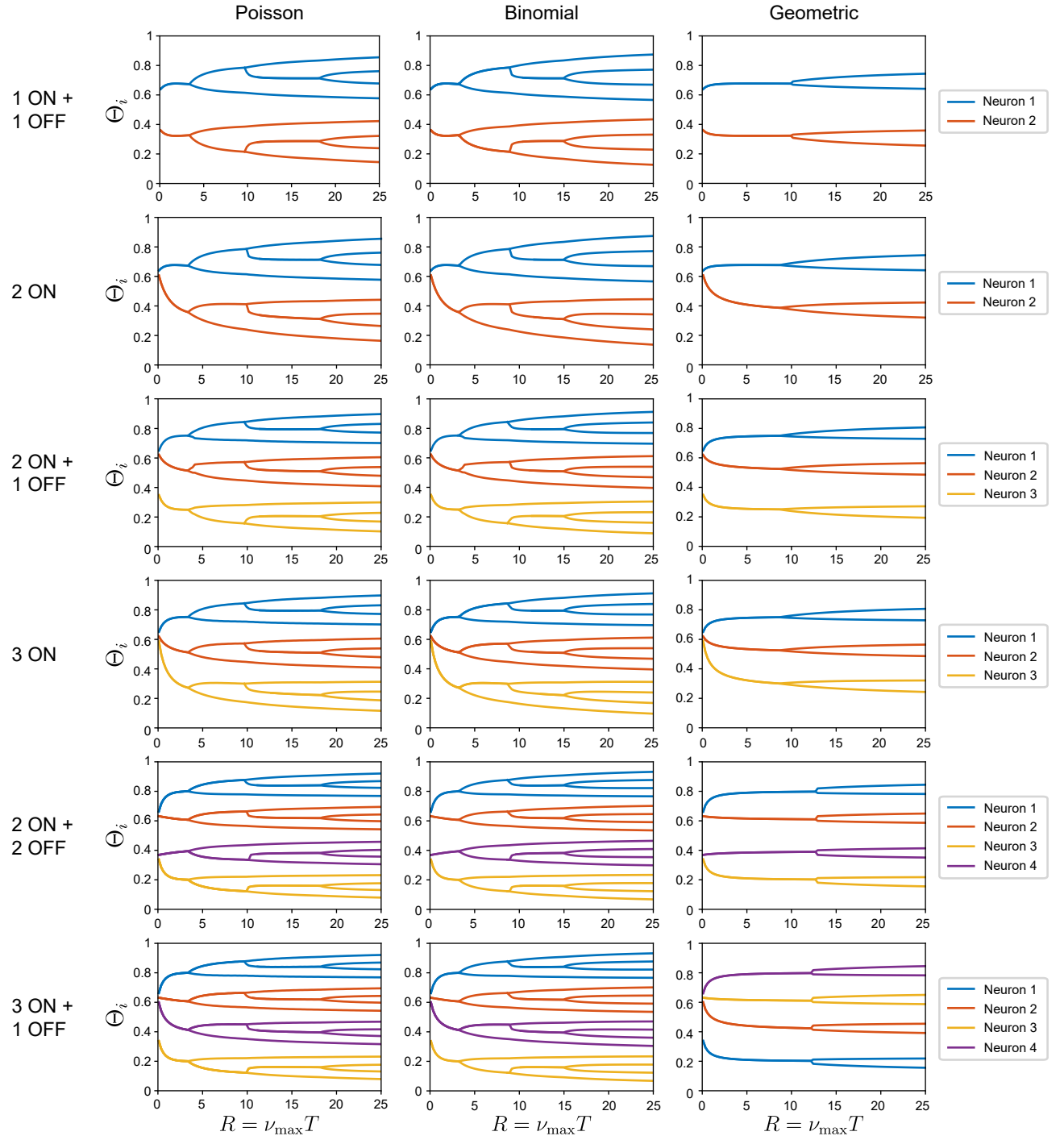
Figure S1. Optimal thresholds in different neuronal populations (1 ON + 1 OFF, 2 ON, 2 ON + 1 OFF, 3 ON, 2 ON + 2 OFF, and 3 ON + 1 OFF) with different noise generation functions (Poisson, Binomial, and Geometric distribution).

5   Population coding of binary neurons with any noise generation function

From now on, we denote the probability function of spike generation as $L(n_i, r_i)$, where n_i is the spike count of neuron i, and r_i is the expected value of n_i. When the firing rate of neuron i is ν_i and the time window is T, we have $r_i = \nu_i T$. Consistent with the main text, we assume the neurons do not have a spontaneous firing rate, i.e., $\nu_0 = 0$. $R = \nu_{\max} T$ is the maximal value of any r_i.

For a binary neuron, we define the interval of stimulus space partitioned by its threshold as $u_i = \text{Prob}(\nu_i = \nu_{max})$, which is the same as Eq. 38 in the main text (Fig. 2B). The mutual information between stimuli and spikes can be formulated as

$$I_1 = g(u_1) = \sum_n \int_s \mathrm{d}s\, p(s)\, p(n|s) \log \frac{P(n|s)}{P(n)}$$

$$= -(1 - u_1) \log P(0) + u_1 L(0, R) \log \frac{L(0, R)}{P(0)} + u_1 \sum_{n=1}^{+\infty} L(n, R) \log \frac{L(n, R)}{u_1 L(n, R)} \tag{S5.1}$$

$$= -P(0) \log P(0) + u_1 L(0, R) \log L(0, R) - u_1 \log u_1 \sum_{n=1}^{+\infty} L(n, R).$$

Define $q = L(0, R) = 1 - \sum_{n=1}^{+\infty} L(n, R)$, we have

$$P(0) = 1 - u_1 + u_1 q \tag{S5.2}$$

$$I_1 = -P(0) \log P(0) + u_1 q \log q - u_1 (1 - q) \log u_1. \tag{S5.3}$$

Here, all the nonzero spike counts have been merged as in previous work with Poisson spike statistics [5]. It is equivalent to only having a firing state $n \neq 0$ and a non-firing state $n = 0$. The only difference from Poisson spike statistics is the exact formulation of the function L and the value of q. This similarity allows us to use some of the results derived in previous literature [5]. For example, given a population of N neurons, its mutual information can be written as

$$I_N = g(u_1) + (1 - u_1(1 - q)) \left[g(u_2^{(1)}) + ... + \left(1 - u_{N-1}^{(N-2)}(1 - q)\right) g(u_N^{(N-1)}) \right] \tag{S5.4}$$

where $u_i^{(j)}$ means the revised probability of u_i after knowing that all neurons $1, ..., j$ $(j < i)$ did not spike, and $g(u) = -(1 - u + uq) \log(1 - u + uq) + uq \log q - u(1 - q) \log u$. The index of neurons follow Eq. 37 in the main text (Fig. 2B).

One can show that $u_i^{(j)}$ follows the rule below [5]:

$$u_i^{(j)} = \frac{u_i^{(j-1)} - \left(1 - P_j^{(j-1)}(0)\right)}{P_j^{(j-1)}(0)} = \frac{u_i^{(j-1)} - u_j^{(j-1)}(1 - q)}{1 - u_j^{(j-1)}(1 - q)} \tag{S5.5}$$

if neuron i and neuron j are both ON, or both OFF, and

$$u_i^{(j)} = \frac{u_i^{(j-1)}}{P_j^{(j-1)}(0)} = \frac{u_i^{(j-1)}}{1 - u_j^{(j-1)}(1 - q)} \tag{S5.6}$$

if neuron i is OFF but neuron j is ON. Here $P_j^{(k)}(0)$ is the probability that neuron j does not fire after knowing that none of the neurons $1, ..., k$ $(k < j)$ spikes, we also have

$$P_j^{(k)}(0) = 1 - u_j^{(k)}(1 - q). \tag{S5.7}$$

Taking derivatives of I_N with respect to $u_N^{(N-1)}$, $u_{N-1}^{(N-2)}$, ..., and u_1 yields [5]

$$u_i^{(i-1)} = \frac{1}{(N - i + 1)(1 - q) + q^{-q/(1-q)}}. \tag{S5.8}$$

Using Eq. S5.8, Eq. S5.5, and Eq. S5.6 iteratively, we can then derive the optimal solution

$$u_i = \frac{1 + (i-1)(1-q)}{N(1-q) + q^{-q/(1-q)}} \tag{S5.9}$$

for ON neurons and

$$u_i = \frac{1 + (m-i+1)(1-q)}{N(1-q) + q^{-q/(1-q)}} \tag{S5.10}$$

for OFF neurons. With the definition in Eq. 44, we instantly get

$$p_1 = p_{m+1} = \frac{1}{N(1-q) + q^{-q/(1-q)}} \overset{\text{def}}{=} p_{\text{edge}}$$

$$p_2 = ... = p_m = p_{m+2} = ... = p_N = \frac{1-q}{N(1-q) + q^{-q/(1-q)}} \overset{\text{def}}{=} p \tag{S5.11}$$

$$p = (1-q)p_{\text{edge}}$$

which is summarized in Eq. 45 of the main text. In this case the maximal mutual information

$$I_N^{\max} = \log\left(1 + N(1-q)q^{q/(1-q)}\right) = -\log P(\vec{0}) \tag{S5.12}$$

and we also have

$$I_N^{\max} = -\log\left(1 - Np\right). \tag{S5.13}$$

Using Eq. S5.12 and comparing $I_1^{\max}$ and $I_N^{\max}$, we have

$$I_N^{\max} = \log\left[N(\exp(I_1^{\max}) - 1) + 1\right]. \tag{S5.14}$$

We can also calculate the overall mean firing rate ($\bar{\nu}$) of a population given the optimal thresholds of Eq. S5.11. If the stimulus s is higher than θ_1 (the threshold of the highest ON neuron), all the m ON neurons fire at $\nu_{\max}$ together. If $s \in (\theta_2, \theta_1)$, $m-1$ ON neurons (neuron $2, ..., m$) fire at $\nu_{\max}$, and so on so forth. Following this idea, we can write

$$\begin{aligned}
\bar{\nu} &= \frac{\nu_{\max}}{N}\left[p_{\text{edge}}m + p\sum_{i=1}^{m-1}(m-i) + p_{\text{edge}}(N-m) + p\sum_{i=1}^{N-m-1}(N-m-i)\right] \\
&= \frac{\nu_{\max}}{N}\left[p_{\text{edge}}N + p\frac{m(m-1)}{2} + p\frac{(N-m)(N-m-1)}{2}\right] \\
&= \left[p_{\text{edge}} + \frac{N-1}{2}p + \frac{m}{N}(m-N)p\right]\nu_{\max},
\end{aligned} \tag{S5.15}$$

which produces Eq. 82 in the main text.

6 Population coding of binary neurons with heterogeneous maximal firing rates

6.1 Maximal mutual information and optimal firing thresholds

As mentioned in the main text, we define $\nu_{\max,i}$ as the maximal firing rate of neuron i, and then

$$R_i = \nu_{\max,i}\, T, \qquad q_i = L(0, R_i), \tag{S6.1}$$

and

$$u_i = \text{Prob}(\nu_i = \nu_{\max,i}) = \begin{cases} \int_{\theta_i}^{+\infty} \mathrm{d}s\, p(s), \text{ for ON} \\ \int_{-\infty}^{\theta_i} \mathrm{d}s\, p(s), \text{ for OFF}. \end{cases} \tag{S6.2}$$

9

Similar to Eq. S5.4, the mutual information of a population of N neurons can be decomposed into N terms,

$$I_N = g_1(u_1) + (1 - u_1(1 - q_1)) \left[g_2(u_2^{(1)}) + \dots + \left(1 - u_{N-1}^{(N-2)}(1 - q_{N-1})\right) g_N(u_N^{(N-1)}) \right] \tag{S6.3}$$

where $u_i^{(j)}$ means the revised probability of u_i after knowing that all neurons $1, \dots, j (j < i)$ did not spike, and $g_i(u)$ denotes the information encoded by neuron i. With binary neurons and heterogeneous maximal firing rates

$$g_i(u) = -(1 - u + uq_i) \log(1 - u + uq_i) + uq_i \log q_i - u(1 - q_i) \log u. \tag{S6.4}$$

Also similarly to the case with identical maximal firing rates for all cells (Eq. S5.5 and Eq. S5.6), $u_i^{(j)}$ relates to $u_i^{(j-1)}$ with

$$u_i^{(j)} = \frac{u_i^{(j-1)} - \left(1 - P_j^{(j-1)}(0)\right)}{P_j^{(j-1)}(0)} = \frac{u_i^{(j-1)} - u_j^{(j-1)}(1 - q_j)}{1 - u_j^{(j-1)}(1 - q_j)} \tag{S6.5}$$

if neuron i and neuron j are both ON, or both OFF, and

$$u_i^{(j)} = \frac{u_i^{(j-1)}}{P_j^{(j-1)}(0)} = \frac{u_i^{(j-1)}}{1 - u_j^{(j-1)}(1 - q_j)} \tag{S6.6}$$

if neuron i is OFF but neuron j is ON. $P_j^{(k)}(0)$ is the probability that neuron j does not fire after knowing that none of the neurons $1, \dots, k \, (k < j)$ spikes, and can be calculated as

$$P_j^{(k)}(0) = 1 - u_j^{(k)}(1 - q_j). \tag{S6.7}$$

Taking derivatives of I_N with respect to $u_N^{(N-1)}$, $u_{N-1}^{(N-2)}$, ..., and u_1 yields

$$u_i^{(i-1)} = \frac{q_i^{q_i/(1-q_i)}}{1 + \sum_{j=i}^{N}(1 - q_j)q_j^{q_j/(1-q_j)}}. \tag{S6.8}$$

Using Eq. S6.5 and Eq. S6.6 iteratively, we can derive the optimal solution

$$u_i = \frac{q_i^{q_i/(1-q_i)} + \sum_{j=1}^{i-1}(1 - q_j)q_j^{q_j/(1-q_j)}}{1 + \sum_{j=1}^{N}(1 - q_j)q_j^{q_j/(1-q_j)}} \tag{S6.9}$$

for ON neurons and

$$u_i = \frac{q_i^{q_i/(1-q_i)} + \sum_{j=m+1}^{i-1}(1 - q_j)q_j^{q_j/(1-q_j)}}{1 + \sum_{j=1}^{N}(1 - q_j)q_j^{q_j/(1-q_j)}}. \tag{S6.10}$$

for OFF neurons. Also, inserting Eq. S6.8 into Eq. S6.3, we can calculate the maximal mutual information as

$$I_N = \log \left[1 + \sum_{i=1}^{N}(1 - q_i)q_i^{q_i/(1-q_i)} \right] \tag{S6.11}$$

With the definition of p_i (Eq. 54), we can calculate the cumulative stimulus intervals partitioned by firing thresholds as

$$p_1 = \frac{q_1^{q_1/(1-q_1)}}{1 + \sum_{j=1}^{N}(1 - q_j)q_j^{q_j/(1-q_j)}} = e^{-I_N} q_1^{q_1/(1-q_1)}$$

$$p_{m+1} = \frac{q_{m+1}^{q_{m+1}/(1-q_{m+1})}}{1 + \sum_{j=1}^{N}(1 - q_j)q_j^{q_j/(1-q_j)}} = e^{-I_N} q_{m+1}^{q_{m+1}/(1-q_{m+1})} \tag{S6.12}$$

$$p_i = \frac{q_i^{q_i/(1-q_i)} - q_{i-1}^{1/(1-q_{i-1})}}{1 + \sum_{j=1}^{N}(1 - q_j)q_j^{q_j/(1-q_j)}} = e^{-I_N} \left[q_i^{q_i/(1-q_i)} - q_{i-1}^{1/(1-q_{i-1})} \right], \quad i \neq 1 \text{ and } i \neq m$$

10

which produces Eq. 55 in the main text. Next, we ask how the optimal stimulus intervals $\{p_i\}$ depend on the noise levels q_i. To proceed, we define

$$f_1(x) = x^{x/(1-x)}, \qquad f_2(x) = x^{1/(1-x)}, \qquad x \in (0,1) \tag{S6.13}$$

which are the two functions appearing in Eq. S6.12 above. One can prove that

$$\begin{aligned}
\frac{\mathrm{d}\log f_1(x)}{\mathrm{d}x} &= \frac{1 - x + \log x}{(1-x)^2} < 0 \\
\frac{\mathrm{d}\log f_2(x)}{\mathrm{d}x} &= \frac{1 - x + x\log x}{(1-x)^2 x} > 0.
\end{aligned} \tag{S6.14}$$

Therefore, given the fixed amount of mutual information I_N, p_1 decreases with q_1, p_{m+1} decreases with q_{m+1}, and any other p_i decreases with both q_i and q_{i-1}. When $q_i \to 1$ and $q_{i-1} \to 1$, p_i approaches its lower limit 0. In contrast, when $q_i \to 0$ and $q_{i-1} \to 0$, p_i approaches its upper limit e^{-I_N}.

6.2 Mean firing rate

The mean firing rate can be calculated as

$$\begin{aligned}
\bar{\nu} &= \frac{1}{N} \sum_{i=1}^{N} \nu_{\max_i} u_i \\
&= \frac{e^{-I_N}}{NT} \left[-\sum_{i=1}^{N} q_i^{q_i/(1-q_i)} \log q_i - \sum_{i=1}^{m} \log q_i \sum_{j=1}^{i-1} (1 - q_j) q_j^{q_j/(1-q_j)} - \sum_{i=m+1}^{N} \log q_i \sum_{j=m+1}^{i-1} (1 - q_j) q_j^{q_j/(1-q_j)} \right]
\end{aligned} \tag{S6.15}$$

Here, we prove that (1) homogeneous populations of ON or OFF neurons generate the highest mean firing rate $\bar{\nu}$; (2) When the lowest $\bar{\nu}$ is reached, Eq. 85 and Eq. 86 must be valid. To proceed, we denote

$$A_i = -\log q_i, \quad B_i = (1 - q_i) q_i^{q_i/(1-q_i)}. \tag{S6.16}$$

Note that for any i, $A_i > 0$ and $B_i > 0$.

To prove the first argument, we rewrite the mean firing rate as a function of m, the number of ON neurons in the population:

$$\bar{\nu}(m) = \frac{e^{-I_N}}{NT} \left[\sum_{i=1}^{N} q_i^{q_i/(1-q_i)} A_i + \sum_{i=1}^{m} A_i \sum_{j=1}^{i-1} B_j + \sum_{i=m+1}^{N} A_i \sum_{j=m+1}^{i-1} B_j \right]. \tag{S6.17}$$

One can then obtain

$$\begin{aligned}
\bar{\nu}(m=N) - \bar{\nu}(m) &= \frac{e^{-I_N}}{NT} \left[\sum_{i=1}^{N} A_i \sum_{j=1}^{i-1} B_j - \sum_{i=1}^{m} A_i \sum_{j=1}^{i-1} B_j - \sum_{i=m+1}^{N} A_i \sum_{j=m+1}^{i-1} B_j \right] \\
&= \frac{e^{-I_N}}{NT} \left[\sum_{i=m+1}^{N} A_i \sum_{j=1}^{i-1} B_j - \sum_{i=m+1}^{N} A_i \sum_{j=m+1}^{i-1} B_j \right] \\
&= \frac{e^{-I_N}}{NT} \sum_{i=m+1}^{N} A_i \sum_{j=1}^{m} B_j \geq 0,
\end{aligned} \tag{S6.18}$$

which means a homogeneous ON population always generates higher mean firing rate than a mixed population with the same maximal firing rates $\{\nu_{\max,i}\}$. Similarly, one can also find $\bar{\nu}(m=0) - \bar{\nu}(m) \geq 0$. Therefore, we have proved the first proposition above, that homogeneous populations of ON or OFF neuron generate the highest mean firing rate $\bar{\nu}$.

To prove Eq. 85, we assume for neuron t and $t+1$ ($1 \leq t < m$), $q_t > q_{t+1}$, which results in $A_t < A_{t+1}$ and $B_t < B_{t+1}$. Then we swap neuron t and $t+1$ and show that $\bar{\nu}$ becomes lower. The change of $\bar{\nu}$ due to the swap can be expressed as

$$
\begin{aligned}
\Delta\bar{\nu} &= \frac{e^{-I_N}}{NT} \left[\sum_{i=1}^{t-1} A_i \sum_{j=1}^{i-1} B_j + A_{t+1} \sum_{j=1}^{t-1} B_j + A_t \left(\sum_{j=1}^{t-1} B_j + B_{t+1} \right) + \sum_{i=t+2}^{m} A_i \sum_{j=1}^{i-1} B_j - \sum_{i=1}^{N} A_i \sum_{j=1}^{i-1} B_j \right] \\
&= \frac{e^{-I_N}}{NT} \left[(A_{t+1} - A_t) \sum_{j=1}^{t-1} B_j + (A_t - A_{t+1}) \sum_{j=1}^{t} B_j + (B_{t+1} - B_t)A_t \right] \\
&= \frac{e^{-I_N}}{NT} \left(B_{t+1}A_t - A_{t+1}B_t \right).
\end{aligned}
$$

(S6.19)

To compare $B_{t+1}A_t$ and $A_{t+1}B_t$, we calculate the derivative $\mathrm{d}(A_i/B_i)/\mathrm{d}q_i$ as follows:

$$
\frac{\mathrm{d}}{\mathrm{d}q_i} \left(\frac{A_i}{B_i} \right) = B_i \left[\left(\frac{\log q_i}{1 - q_i} \right)^2 - \frac{1}{q_i} \right],
$$

(S6.20)

where we have used Eq. S6.14 to acquire $\mathrm{d}B_i/\mathrm{d}q_i = B_i \log q_i/(1 - q_i)^2$. Then we define

$$
f_3(x) = \log x - \frac{x - 1}{\sqrt{x}}
$$

(S6.21)

and show that $f_3'(x) \leq 0$. This leads to

$$
\frac{\mathrm{d}}{\mathrm{d}q_i} \left(\frac{A_i}{B_i} \right) = B_i \left[\left(\frac{\log q_i}{1 - q_i} \right)^2 - \frac{1}{q_i} \right] < 0,
$$

(S6.22)

and then

$$
\frac{A_t}{B_t} < \frac{A_{t+1}}{B_{t+1}}.
$$

(S6.23)

Therefore, if $q_t > q_{t+1}$, swapping neuron t and $t+1$ will reduce the mean firing rate $\bar{\nu}$. This implies when the mean firing rate $\bar{\nu}$ is minimized, there must be

$$
q_1 \leq ... \leq q_m.
$$

(S6.24)

In the same way we can prove that there must also be

$$
q_{m+1} \leq ... \leq q_N,
$$

(S6.25)

which is the equivalency for OFF neurons. These two equations prove Eq. 85 in the main text.

Finally, to prove Eq. 86, we calculate the difference $\bar{\nu}(m) - \bar{\nu}(m - 1)$, which should converge to 0 when $\bar{\nu}(m)$ approaches the minimum. According to Eq. S6.17,

$$
\bar{\nu}(m) - \bar{\nu}(m - 1) = A_m \sum_{j=1}^{m-1} B_j - B_m \sum_{j=m+1}^{N} A_i.
$$

(S6.26)

Then $\bar{\nu}(m) - \bar{\nu}(m - 1) \to 0$ gives rise to Eq. 86 in the main text.

7 Population coding of ternary neurons with any noise generation function

7.1 Maximal mutual information of a population of ternary neurons

Following Eq. 29, we decompose the mutual information encoded by N neurons as

$$
I_N \left(\vec{u_1}, ..., \vec{u_N} \right) = I_m \left(\vec{u_1}, ..., \vec{u_m} \right) + Q_m I_{N-m} \left(\vec{u}_{m+1}^{(m)}, ..., \vec{u}_N^{(m)} \right).
$$

(S7.1)

Here, the cumulative stimulus intervals $\vec{u}_i$ are written as vectors because the activation functions are ternary. Q_m denotes the probability that none of the neurons $1, 2, ..., m$ fires. As before, we define the superscript (m) to denote the 'revised' stimulus intervals assuming that the m ON neurons do not fire.

To generalize the definitions of these revised thresholds for N ternary activation functions, we use $u_{i1}^{(j,N)}$ and $u_{i2}^{(j,N)}$ to denote the cumulative stimulus intervals given the condition that none of the neurons $1, ..., j$ ($j < i$) fires, and

$$P_i^{(j,N)}(0) = 1 - u_{i1}^{(j,N)}(1 - L(0, f_i R)) - u_{i2}^{(j,N)}(1 - L(0, R)) \tag{S7.2}$$

to denote the probability that neuron i does not fire, when none of the neurons $1, ..., j$ ($j < i$) fires. We can decompose the mutual information encoded by a population of N neurons into N single terms, each of which contains the mutual information encoded by one neuron. This allows us to calculate the mutual information and optimize the threshold in a recursive way, first for one neuron, then for two neurons and then generalize for any N neurons (compare to Eq. 32),

$$I_N = g(\vec{u_1}) + P_1(0) \left\{ g\left(\vec{u_2^{(1)}}\right) + P_2^{(1,N)}(0) \left[g\left(\vec{u_3^{(2)}}\right) + ... + P_{N-1}^{(N-2,N)}(0) \, g\left(\vec{u_N^{(N-1)}}\right) \right] \right\}. \tag{S7.3}$$

Note that similar to the case of binary neurons (Eq. S5.1 and Eq. S5.4), the function g here still denotes the mutual information of a single neuron, despite the difference that it takes in a vector instead of a scalar.

The mutual information of one single neuron can be written as

$$I_1 = g(u_{11}, u_{12}) = \sum_n \int_s \mathrm{d}s \, p(s)p(n|s) \log \frac{P(n|s)}{P(n)}$$

$$= -(1 - u_{11} - u_{12}) \log P(0) + u_{11} L(0, fR) \log \frac{L(0, fR)}{P(0)} + u_{12} L(0, R) \log \frac{L(0, R)}{P(0)}$$

$$+ u_{11} \sum_{n=1}^{+\infty} L(n, fR) \log \frac{L(n, fR)}{u_{11} L(n, fR) + u_{12} L(n, R)} + u_{12} \sum_{n=1}^{+\infty} L(n, R) \log \frac{L(n, R)}{u_{11} L(n, fR) + u_{12} L(n, R)}$$

$$= -P(0) \log P(0) + u_{11} L(0, fR) \log L(0, fR) + u_{12} L(0, R) \log L(0, R)$$

$$+ u_{11} \sum_{n=1}^{+\infty} L(n, fR) \log \frac{L(n, fR)}{u_{11} L(n, fR) + u_{12} L(n, R)} + u_{12} \sum_{n=1}^{+\infty} L(n, R) \log \frac{L(n, R)}{u_{11} L(n, fR) + u_{12} L(n, R)}. \tag{S7.4}$$

Its derivatives are

$$\frac{\partial g(u_{11}, u_{12})}{\partial u_{11}} = \frac{\partial P(0)}{\partial u_{11}} [-1 - \log P(0)] + L(0, fR) \log L(0, fR) + \sum_{n=1}^{+\infty} L(n, fR) \left(\log \frac{L(n, fR)}{u_{11} L(n, fR) + u_{12} L(n, R)} - 1 \right) \tag{S7.5}$$

$$\frac{\partial g(u_{11}, u_{12})}{\partial u_{12}} = \frac{\partial P(0)}{\partial u_{12}} [-1 - \log P(0)] + L(0, R) \log L(0, R) + \sum_{n=1}^{+\infty} L(n, R) \left(\log \frac{L(n, R)}{u_{11} L(n, fR) + u_{12} L(n, R)} - 1 \right) \tag{S7.6}$$

and

$$\frac{\partial g(u_{11}, u_{12})}{\partial f} = - u_{11} \frac{\partial L(0, fR)}{\partial f} \log P(0) + u_{11} \sum_{n=0}^{+\infty} \log L(n, fR) \frac{\partial L(n, fR)}{\partial f}$$
$$- \sum_{n=1}^{+\infty} \log[u_{11} L(n, fR) + u_{12} L(n, R)] u_{11} \frac{\partial L(n, fR)}{\partial f}. \tag{S7.7}$$

Note that

$$P(0) = 1 - u_{11} [1 - L(0, fR)] - u_{12} [1 - L(0, R)], \tag{S7.8}$$

$$\frac{\partial P(0)}{\partial u_{11}} = L(0, fR) - 1, \quad \frac{\partial P(0)}{\partial u_{12}} = L(0, R) - 1 \tag{S7.9}$$

we can rewrite Eq. S7.5 and Eq. S7.6 into

$$\frac{\partial g(u_{11}, u_{12})}{\partial u_{11}} = [1 - L(0, fR)] \log P(0) + L(0, fR) \log L(0, fR) + \sum_{n=1}^{+\infty} L(n, fR) \log \frac{L(n, fR)}{u_{11}L(n, fR) + u_{12}L(n, R)}$$

(S7.10)

and

$$\frac{\partial g(u_{11}, u_{12})}{\partial u_{12}} = [1 - L(0, R)] \log P(0) + L(0, R) \log L(0, R) + \sum_{n=1}^{+\infty} L(n, R) \log \frac{L(n, R)}{u_{11}L(n, fR) + u_{12}L(n, R)}.$$

(S7.11)

Combining these two derivatives, and comparing to Eq. S7.4, we can find

$$g(u_{11}, u_{12}) = u_{11}\frac{\partial g}{\partial u_{11}} + u_{12}\frac{\partial g}{\partial u_{12}} - \log P_1(0).$$

(S7.12)

When $g(u_{11}, u_{12})$ is at its maximal value, $\partial g/\partial u_{11} = 0$, $\partial g/\partial u_{12} = 0$, then

$$I_1^{\max} = g(u_{11}^*, u_{12}^*) = -\log P_1(0).$$

(S7.13)

We then examine the mutual information of a population of two neurons,

$$I_2 = g(u_{11}, u_{12}) + P_1(0) g\left(u_{21}^{(1)}, u_{22}^{(1)}\right).$$

(S7.14)

Maximizing $g\left(u_{21}^{(1)}, u_{22}^{(1)}\right)$ gives

$$I_2 = g(u_{11}, u_{12}) - P_1(0) \log P_2^{(1)}(0).$$

(S7.15)

Note that $P_1(0) = 1 - u_{11}(1 - L(0, fR)) - u_{22}(1 - L(0, R))$. When I_2 is at its maximum, we have $\partial I_2/\partial u_{11} = 0$ and $\partial I_2/\partial u_{12} = 0$, which leads to

$$I_2^{\max} = -\log P_2^{(1)}(0) - \log P_1(0).$$

(S7.16)

Next, we use mathematical induction to prove that for arbitrary N (Eq. S7.21), there is

$$I_N^{\max} = -\sum_{j=1}^{N} \log P_j^{(j-1)}(0).$$

(S7.17)

Assume $I_m^{\max} = -\sum_{j=1}^{m} \log P_j^{(j-1)}(0)$, we have

$$I_{m+1} = g(u_{11}, u_{12}) - P_1(0) \sum_{j=2}^{m+1} \log P_j^{(j-1)}(0).$$

(S7.18)

When optimized,

$$\frac{\partial I_{m+1}}{\partial u_{11}} = \frac{\partial g}{\partial u_{11}} + [1 - L(0, fR)] \sum_{j=2}^{m+1} \log P_j^{(j-1)}(0) = 0$$

$$\frac{\partial I_{m+1}}{\partial u_{12}} = \frac{\partial g}{\partial u_{12}} + [1 - L(0, R)] \sum_{j=2}^{m+1} \log P_j^{(j-1)}(0) = 0$$

(S7.19)

$$I_{m+1}^{\max} = u_{11}\frac{\partial g}{\partial u_{11}} + u_{12}\frac{\partial g}{\partial u_{12}} - \log P_1(0) - P_1(0) \sum_{j=2}^{m+1} \log P_j^{j-1}(0)$$

$$= -[u_{11}(1 - L(0, fR)) + u_{12}(1 - L(0, R))] \sum_{j=2}^{m+1} \log P_j^{(j-1)}(0) - \log P_1(0) - P_1(0) \sum_{j=2}^{m+1} \log P_j^{(j-1)}(0)$$

$$= -\sum_{j=1}^{m+1} \log P_j^{(j-1)}(0).$$

(S7.20)

Hence we have verified that if $I_m^{\max} = -\sum_{j=1}^{m} \log P_j^{(j-1)}(0)$, we can show $I_{m+1}^{\max} = -\sum_{j=1}^{m+1} \log P_j^{(j-1)}(0)$. This means we can use mathematical induction to generalize the mutual information from $N = 2$ to arbitrary number of neurons N, and obtain

$$I_N^{\max} = -\sum_{j=1}^{N} \log P_j^{(j-1)}(0) = -\log P(\vec{0}). \tag{S7.21}$$

7.2 Revised probability of ternary neurons

The mutual information of N neurons can be formulated by Eq. S7.3, i.e.,

$$I_N = g(\vec{u_1}) + P_1(0) \left\{ g\left(u_2^{\vec{(1)}}\right) + P_2^{(1,N)}(0) \left[g\left(u_3^{\vec{(2)}}\right) + ... + P_{N-1}^{(N-2,N)}(0) g\left(u_N^{\vec{(N-1)}}\right) \right] \right\}.$$

Assuming we already know that neuron $1, 2, ..., j$ do not fire, we can then derive that for ON neurons, if $j + 1 < i$,

$$u_{i1}^{(j+1,N)} = \int_{\theta_{i1}^{(j,N)}}^{\theta_{i2}^{(j,N)}} p(s|n_{j+1}=0)\,\mathrm{d}s = \int_{\theta_{i1}^{(j,N)}}^{\theta_{i2}^{(j,N)}} \frac{p(s)\,p(n_{j+1}=0|s)}{p(n_{j+1}=0)}\,\mathrm{d}s. \tag{S7.22}$$

Since we already know that neuron $1, 2, ..., j$ do not fire, here we have $p(n_{j+1}=0) = P_{j+1}^{(j,N)}(0)$. Also, given $j + 1 < i$, stimulus s between $\theta_{i1}^{(j,N)}$ and $\theta_{i2}^{(j,N)}$ is too low to trigger a nonzero firing rate of neuron $j + 1$. Therefore, within the interval of the integral of Eq. S7.22, $p(n_{j+1}=0|s) = 1$. Eq. S7.22 then becomes

$$u_{i1}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \int_{\theta_{i1}^{(j,N)}}^{\theta_{i2}^{(j,N)}} p(s)\mathrm{d}s = \frac{u_{i1}^{(j,N)}}{P_{j+1}^{(j,N)}(0)}. \tag{S7.23}$$

Similarly, we have

$$u_{i2}^{(j+1,N)} = \int_{\theta_{i2}^{(j,N)}}^{+\infty} p(s|n_{j+1}=0)\,\mathrm{d}s = \int_{\theta_{i2}^{(j,N)}}^{+\infty} \frac{p(s)\,p(n_{j+1}=0|s)}{p(n_{j+1}=0)}\,\mathrm{d}s. \tag{S7.24}$$

Same as Eq. S7.22, we have $p(n_{j+1}=0) = P_{j+1}^{(j,N)}(0)$. Also because $p(n_{j+1}=0|s) = 1 - p(n_{j+1}\neq0|s)$, we can rewrite Eq. S7.24 as

$$u_{i2}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \int_{\theta_{i2}^{(j,N)}}^{+\infty} p(s)\left[1 - p(n_{j+1}\neq0|s)\right]\,\mathrm{d}s. \tag{S7.25}$$

Because stimulus s lower than $\theta_{i2}^{(j,N)}$ cannot trigger a nonzero firing rate of neuron $j + 1$, we have

$$\int_{\theta_{i2}^{(j,N)}}^{+\infty} p(s)\,p(n_{j+1}\neq0|s)\,\mathrm{d}s = 1 - P_{j+1}^{(j,N)}(0). \tag{S7.26}$$

Substituting back into Eq. S7.25 gives rise to

$$u_{i2}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \left[u_{i2}^{(j,N)} - \left(1 - P_{j+1}^{(j,N)}(0)\right) \right]. \tag{S7.27}$$

Notably, Eq. S7.23 and Eq. S7.27 have similar formulation compared to Eq. S5.5 and Eq. S5.6.
When I_N is maximized,

$$\frac{\partial I_N}{\partial u_N^{(\vec{N-1})}} = \left(\prod_{i=1}^{N-1} P_i^{(i-1,N)}(0) \right) \frac{\partial g\left(u_N^{(\vec{N-1})}\right)}{\partial u_N^{(\vec{N-1})}} = 0. \tag{S7.28}$$

Therefore, optimizing $u_N^{(N-1)}$ is equivalent to maximizing the function $g(\cdot)$, which is the mutual information of a single neuron. Denoting the optimal values using an asterisk, we have

$$u_{N,1}^{(N-1,N)}|_* = u_1^*, \ u_{N,2}^{(N-1,N)}|_* = u_2^*, \tag{S7.29}$$

where u_1^* and u_2^* correspond to the (u_11, u_12) that maximizes $g(\vec{u}_1) = g(u_{11}, u_{12})$. Also, we can see that

$$g\left(u_2^{\vec{(1)}}\right) + P_2^{(1,N)}(0)\left[g\left(u_3^{\vec{(2)}}\right) + ... + P_{N-1}^{(N-2,N)}(0)g\left(u_N^{\vec{(N-1)}}\right)\right] \tag{S7.30}$$

has the exactly same formulation as I_{N-1}, which means

$$u_{i,1}^{(j,N)}|_* = u_{i+1,1}^{(j+1,N+1)}|_*, \ u_{i,2}^{(j,N)}|_* = u_{i+1,2}^{(j+1,N+1)}|_*. \tag{S7.31}$$

7.3 A lemma that connects two adjacent neurons

Next, we seek to find the optimal thresholds by deriving the relationship among $\{u_{i1}\}$ and $\{u_{i2}\}$. We start with proving the following lemma that links two adjacent neurons, i.e., (u_{i1}, u_{i2}, f_i) and $(u_{i+1,1}, u_{i+1,2}, f_{i+1})$.

Lemma: For any N neurons, when I_N is optimized,

$$u_{i1}^{(i-1,N)} = P_i^{(i-1,N)}(0)u_{i+1,1}^{(i,N)}, \quad u_{i2}^{(i-1,N)} = P_i^{(i-1,N)}(0)u_{i+1,2}^{(i,N)}, \quad f_i = f_{i+1}. \tag{S7.32}$$

Remark:

$$I = g(u_{11}, u_{12}) + P_1^{(0,N)}(0)\left[g\left(u_{21}^{(1,N)}, u_{22}^{(1,N)}\right) + ... + P_{N-1}^{(N-2,N)}(0)g\left(u_{N1}^{(N-1,N)}, u_{N2}^{(N-1,N)}\right)\right]. \tag{S7.33}$$

When all the u_{i1}, u_{i2} and f_i are optimized, according to the previous subsection,

$$g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right) + P_i^{(i-1,N)}(0)\left[g\left(u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}\right) + ... + P_{N-1}^{(N-2,N)}(0)g\left(u_{N1}^{(N-1,N)}, u_{N2}^{(N-1,N)}\right)\right]$$
$$= -\sum_{j=i}^{N} \log P_j^{(j-1,N)}(0). \tag{S7.34}$$

This means when $\left(u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}, f_{i+1}\right)$ is optimized,

$$\frac{\partial I_N}{\partial u_{i+1,1}^{(i,N)}} = P_i(0)\left\{\frac{\partial g\left(u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}\right)}{\partial u_{i+1,1}^{(i,N)}} + [1 - L(0, f_{i+1}R)]\sum_{j=i+2}^{N} \log P_j^{(j-1,N)}(0)\right\} = 0 \tag{S7.35}$$

$$\frac{\partial I_N}{\partial f_{i+1}} = P_i(0)\left\{\frac{\partial g\left(u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}\right)}{\partial f_{i+1}} - u_{i+1,1}^{(i,N)}\frac{\partial L(0, f_{i+1}R)}{\partial f_{i+1}}\sum_{j=i+2}^{N} \log P_j^{(j-1,N)}(0)\right\} = 0. \tag{S7.36}$$

Denote

$$\hat{u}_{i1} = P_i^{(i-1,N)}(0)u_{i+1,1}^{(i,N)}, \quad \hat{u}_{i2} = P_i^{(i-1,N)}(0)u_{i+1,2}^{(i,N)}, \quad \hat{f}_i = f_{i+1}, \tag{S7.37}$$

so that with given $\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i$,

$$P_i^{(i-1,N)}(0) = 1 - \hat{u}_{i1}[1 - L(0, f_iR)] - \hat{u}_{i2}[1 - L(0, R)] \tag{S7.38}$$

$$\hat{u}_{i1} = \frac{u_{i+1,1}^{(i,N)}}{1 + u_{i+1,1}^{(i,N)}[1 - L(0, f_iR)] + u_{i+1,2}^{(i,N)}[1 - L(0, R)]} \tag{S7.39}$$

16

$$\hat{u}_{i2} = \frac{u_{i+1,2}^{(i,N)}}{1 + u_{i+1,1}^{(i,N)}[1 - L(0, f_i R)] + u_{i+1,2}^{(i,N)}[1 - L(0, R)]}. \tag{S7.40}$$

We need to prove

$$\frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0, \quad \frac{\partial I_N}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0, \quad \frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0 \tag{S7.41}$$

i.e. the combination of $\hat{u}_{i1}, \hat{u}_{i2}$, and $\hat{f}_i$ is optimal.

Proof:

$$\frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = P_{i-1}(0) \left\{ \frac{\partial g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} + [1 - L(0, \hat{f}_i R)] \sum_{j=i+1}^{N} \log P_j^{(j-1,N)}(0) \right\} \tag{S7.42}$$

$$\frac{\partial I_N}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = P_{i-1}(0) \left\{ \frac{\partial g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} + [1 - L(0, R)] \sum_{j=i+1}^{N} \log P_j^{(j-1,N)}(0) \right\}. \tag{S7.43}$$

According to Eq. S7.35,

$$\frac{1}{P_{i-1}(0)} \frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} - \frac{1}{P_i(0)} \frac{\partial I_N}{\partial u_{i+1,1}^{(i,N)}}\Big|_{u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}, f_{i+1}}$$

$$\overset{Eq.\ S7.35}{=\!=\!=\!=\!=} \frac{\partial g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} - \frac{\partial g\left(u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}\right)}{\partial u_{i+1,1}^{(i,N)}} + \left[1 - L(0, \hat{f}_i R)\right] \log P_{i+1}^{(i,N)}(0)$$

$$\overset{Eq.\ S7.5}{=\!=\!=\!=\!=} [1 - L(0, \hat{f}_i R)] \left(\log P_i^{(i-1,N)}(0) - \log P_{i+1}^{(i,N)}(0) \right) - \sum_{n=1}^{+\infty} L(n, \hat{f}_i R) \log \left[\hat{u}_{i1} L(n, fR) + \hat{u}_{i2} L(n, R) \right] +$$

$$\sum_{n=1}^{+\infty} L(n, \hat{f}_i R) \log \left[u_{i+1,1}^{(i,N)} L(n, fR) + u_{i+1,2}^{(i,N)} L(n, R) \right] + [1 - L(0, \hat{f}_i R)] \log P_{i+1}^{(i,N)}(0)$$

$$= [1 - L(0, \hat{f}_i R)] \left(\log P_i^{(i-1,N)}(0) - \log P_{i+1}^{(i,N)}(0) \right) - \sum_{n=1}^{+\infty} L(n, \hat{f}_i R) \log P_i^{(i-1,N)}(0) + [1 - L(0, \hat{f}_i R)] \log P_{i+1}^{(i,N)}(0). \tag{S7.44}$$

Because

$$L(0, \hat{f}_i R) + \sum_{n=1}^{+\infty} L(n, \hat{f}_i R) = 1 \quad \text{(the sum of probabilities)}, \tag{S7.45}$$

$$\frac{1}{P_{i-1}(0)} \frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} - \frac{1}{P_i(0)} \frac{\partial I_N}{\partial u_{i+1,1}^{(i,N)}} = 0, \quad \frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0. \tag{S7.46}$$

Similarly, we have

$$\frac{\partial I_N}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0. \tag{S7.47}$$

Also,

$$\frac{1}{\hat{u}_{i1}P_{i-1}(0)}\frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{1}{u_{i+1,1}^{(i,N)}P_i(0)}\frac{\partial I_N}{\partial f_{i+1}}\Big|_{u_{i+1,1}^{(i,N)},u_{i+1,2}^{(i,N)},f_{i+1}}$$

$$\overset{Eq. S7.36}{=\!=\!=\!=} -\frac{\partial L(0,\hat{f}_iR)}{\partial \hat{f}_i}\log P_i^{(i-1,N)}(0) + \sum_{n=0}^{+\infty}\log L(n,\hat{f}_iR)\frac{\partial L(n,\hat{f}_iR)}{\partial \hat{f}_i}$$

$$-\sum_{n=1}^{+\infty}\log\left[\hat{u}_{i1}L(n,\hat{f}_iR) + \hat{u}_{i2}L(n,R)\right]\frac{\partial L(n,\hat{f}_iR)}{\partial \hat{f}_i} - \sum_{n=0}^{+\infty}\log L(n,f_{i+1}R)\frac{\partial L(n,f_{i+1}R)}{\partial f_{i+1}} \qquad (S7.48)$$

$$+\sum_{n=1}^{+\infty}\log\left[u_{i+1,1}L(n,f_{i+1}R) + u_{i+1,2}L(n,R)\right]\frac{\partial L(n,f_{i+1}R)}{\partial f_{i+1}}$$

$$-\left(\frac{\partial L(0,\hat{f}_iR)}{\partial \hat{f}_i} - \frac{\partial L(0,f_{i+1}R)}{\partial f_{i+1}}\right)\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0).$$

According to the definition of $\hat{f}_i$, $\hat{u}_{i1}$ and $\hat{u}_{i2}$ (Eq. S7.37), we have

$$\frac{1}{\hat{u}_{i1}P_{i-1}(0)}\frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{1}{u_{i+1,1}^{(i,N)}P_i(0)}\frac{\partial I_N}{\partial f_{i+1}}$$

$$= -\frac{\partial L(0,\hat{f}_iR)}{\partial \hat{f}_i}\log P_i^{(i-1,N)}(0) - \log P_i^{(i-1,N)}(0)\sum_{n=1}^{+\infty}\frac{\partial L(n,\hat{f}_iR)}{\partial \hat{f}_i}$$

$$= -\sum_{n=0}^{+\infty}\frac{\partial L(n,\hat{f}_iR)}{\partial \hat{f}_i}\log P_i^{(i-1,N)}(0) \qquad (S7.49)$$

$$= -\frac{\partial}{\partial \hat{f}_i}\left[\sum_{n=0}^{+\infty}L(n,\hat{f}_iR)\right]\log P_i^{(i-1,N)}(0).$$

Since $\sum_{n=0}^{+\infty}L(n,\hat{f}_iR) = 1$ is a constant, $\frac{\partial}{\partial \hat{f}_i}\left[\sum_{n=0}^{+\infty}L(n,\hat{f}_iR)\right] = 0$, then

$$\frac{1}{\hat{u}_{i1}P_{i-1}(0)}\frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{1}{u_{i+1,1}^{(i,N)}P_i(0)}\frac{\partial I_N}{\partial f_{i+1}} = 0 \qquad (S7.50)$$

which leads to

$$\frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} = 0. \qquad (S7.51)$$

Combined with Eq. S7.46 and Eq. S7.47, the lemma has been proved. ∎

7.4 Optimal thresholds for a homogeneous population of only ON neurons

We first consider a homogeneous population with only ON neurons. All the variables in this subsection are optimized, so we omit the * symbol (that we previously used to indicate optimum, e.g., Eq. S7.31) for clarity. From the lemma (Eq. S7.32) we know that $f_i = f_{i+1}$, which we denote as f. For brevity, we also denote $q_1 = L(0,fR)$ and $q_2 = L(0,R)$.

Combining the lemma (Eq. S7.32) and Eq. S7.2

$$P_i^{(i-1,N)}(0) = 1 - u_{i1}^{(i-1,N)}(1-q_1) - u_{i2}^{(i-1,N)}(1-q_2), \qquad (S7.52)$$

we can write

$$P_i^{(i-1,N)}(0) = \left[1 + u_{i+1,1}^{(i,N)}(1-q_1) + u_{i+1,2}^{(i,N)}(1-q_2)\right]^{-1}$$

$$u_{i1}^{(i-1,N)} = \frac{u_{i+1,1}^{(i,N)}}{1 + u_{i+1,1}^{(i,N)}(1-q_1) + u_{i+1,2}^{(i,N)}(1-q_2)} \tag{S7.53}$$

$$u_{i2}^{(i-1,N)} = \frac{u_{i+1,2}^{(i,N)}}{1 + u_{i+1,1}^{(i,N)}(1-q_1) + u_{i+1,2}^{(i,N)}(1-q_2)}.$$

Eq. S7.53 allows us to recursively calculate $\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)$ from $\left(u_{i+1,1}^{(i,N)}, u_{i+1,2}^{(i,N)}\right)$. Starting from $u_{N1}^{(N-1,N)} = u_1$, $u_{N2}^{(N-1,N)} = u_2$, we can reach

$$P_i^{(i-1,N)}(0) = \frac{1 + (N-i-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}{1 + (N-i)\left[u_1(1-q_1) + u_2(1-q_2)\right]}$$

$$u_{i1}^{(i-1,N)} = \frac{u_1}{1 + (N-i)\left[u_1(1-q_1) + u_2(1-q_2)\right]} \tag{S7.54}$$

$$u_{i2}^{(i-1,N)} = \frac{u_2}{1 + (N-i)\left[u_1(1-q_1) + u_2(1-q_2)\right]}.$$

Using Eq. S7.23 and Eq. S7.27, we can do the inverse calculation of revising the probabilities and obtain

$$u_{i1}^{(j,N)} = \frac{u_1}{1 + (N-j-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}$$

$$u_{i2}^{(j,N)} = \frac{u_2 + (i-j-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}{1 + (N-j-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}. \tag{S7.55}$$

Letting $j = 0$, this becomes the non-revised stimulus intervals (Eq. 58), which is

$$u_{i1} = \frac{u_1}{1 + (N-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}$$

$$u_{i2} = \frac{u_2 + (i-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}{1 + (N-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]}. \tag{S7.56}$$

With the definition of p_{i1} and p_{i2} (Eq. 61), we can write

$$p_{11} = p_{21} = ... = p_{N1} = \frac{u_1}{1 + (N-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]} \stackrel{\text{def}}{=} p_1$$

$$p_{22} = ... = p_{N2} = \frac{-u_1 q_1 + u_2(1-q_2)}{1 + (N-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]} \stackrel{\text{def}}{=} p_2 \tag{S7.57}$$

$$p_{12} = \frac{u_2}{1 + (N-1)\left[u_1(1-q_1) + u_2(1-q_2)\right]} \stackrel{\text{def}}{=} p_{\text{edge}}$$

$$p_1 q_1 + p_2 = p_{\text{edge}}(1-q_2).$$

This derives the optimal thresholds for a homogeneous population of ternary ON cells (Fig. 5B).

7.5 Revised probability in ON-OFF mixtures

For ON-OFF mixtures, the revised probability (Eq. S7.23, Eq. S7.27) needs to be adjusted. We derive the equivalency of Eq. S7.23 and Eq. S7.27 that we derived before for ON neurons.

Assuming we already know that neuron $1, 2, ..., j$ do not fire, for an OFF neuron i, if neuron $j + 1$ is also an OFF neuron, and $j + 1 < i$, we have

$$u_{i1}^{(j+1,N)} = \int_{\theta_{i2}^{(j,N)}}^{\theta_{i1}^{(j,N)}} p(s|n_{j+1} = 0)\,ds = \int_{\theta_{i2}^{(j,N)}}^{\theta_{i1}^{(j,N)}} \frac{p(s)\,p(n_{j+1} = 0|s)}{p(n_{j+1} = 0)}\,ds. \tag{S7.58}$$

Stimulus s within the interval $\left[\theta_{i2}^{(j,N)}, \theta_{i1}^{(j,N)}\right]$ cannot trigger a nonzero firing rate of neuron $j+1$. For this reason, in the integral above, $p(n_{j+1} = 0|s) = 1$. Since we already assumed that neuron $1, 2, ..., j$ do not fire, $p(n_{j+1} = 0) = P_{j+1}^{(j,N)}(0)$. Then we can rewrite Eq. S7.58 as

$$u_{i1}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \int_{\theta_{i2}^{(j,N)}}^{\theta_{i1}^{(j,N)}} p(s)\,\mathrm{d}s = \frac{u_{i1}^{(j,N)}}{P_{j+1}^{(j,N)}(0)}. \tag{S7.59}$$

For $u_{i2}^{(j+1,N)}$, we have

$$u_{i2}^{(j+1,N)} = \int_{-\infty}^{\theta_{i2}^{(j,N)}} p(s|n_{j+1} = 0)\,\mathrm{d}s = \int_{-\infty}^{\theta_{i2}^{(j,N)}} \frac{p(s)\,p(n_{j+1} = 0|s)}{p(n_{j+1} = 0)}\,\mathrm{d}s. \tag{S7.60}$$

Similar to Eq. S7.25 before, we have $p(n_{j+1} = 0) = P_{j+1}^{(j,N)}(0)$ and $p(n_{j+1} = 0|s) = 1 - p(n_{j+1} \neq 0|s)$. We can rewrite Eq. S7.60 as

$$u_{i2}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \int_{-\infty}^{\theta_{i2}^{(j,N)}} p(s)\left[1 - p(n_{j+1} \neq 0|s)\right]\,\mathrm{d}s. \tag{S7.61}$$

Because stimulus s higher than $\theta_{i2}^{(j,N)}$ cannot lead to $n_{j+1} \neq 0$, we have

$$\int_{-\infty}^{\theta_{i2}^{(j,N)}} p(s)\,p(n_{j+1} \neq 0|s)\,\mathrm{d}s = P_{j+1}^{(j,N)}(0). \tag{S7.62}$$

Substituting back into Eq. S7.61 gives rise to

$$u_{i2}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \left[u_{i2}^{(j,N)} - \left(1 - P_{j+1}^{(j,N)}(0)\right)\right]. \tag{S7.63}$$

Here we find Eq. S7.59 and Eq. S7.63 are the same as Eq. S7.23 and Eq. S7.27 we derived before. However, if neuron $j + 1$ is an ON neuron, $u_{i1}^{(j+1,N)}$ will remain the same as above while for $u_{i2}^{(j+1,N)}$, we have $p(n_{j+1} = 0|s) = 1$ for $s \in (-\infty, \theta_{i2}^{(j,N)}]$. As a result, Eq. S7.63 changes into

$$u_{i2}^{(j+1,N)} = \frac{u_{i2}^{(j,N)}}{P_{j+1}^{(j,N)}(0)}. \tag{S7.64}$$

7.6 Optimal thresholds for a population of ON and OFF cells

For ON-OFF mixtures, the mutual information can still be written recursively as Eq. S7.3, hence, the lemma (Eq. S7.32) and Eq.S7.52-S7.54 also pertain to ON-OFF mixed populations. If neuron i and neuron j are both ON neurons ($j < i \leq m$) or both OFF neurons ($i > m$ and $j \geq m$), Eq. S7.55 becomes

$$\begin{aligned}
u_{i1}^{(j,N)} &= \frac{u_1}{1 + (N - j - 1)\left[u_1(1 - q_1) + u_2(1 - q_2)\right]} \\
u_{i2}^{(j,N)} &= \frac{u_2 + (i - j - 1)\left[u_1(1 - q_1) + u_2(1 - q_2)\right]}{1 + (N - j - 1)\left[u_1(1 - q_1) + u_2(1 - q_2)\right]}
\end{aligned} \tag{S7.65}$$

and if neuron i is an OFF neuron but neuron j is an ON neuron ($i > m$, $j \leq m$)

$$\begin{aligned}
u_{i1}^{(j,N)} &= \frac{u_1}{1 + (N - j - 1)\left[u_1(1 - q_1) + u_2(1 - q_2)\right]} \\
u_{i2}^{(j,N)} &= \frac{u_2 + (i - m - 1)\left[u_1(1 - q_1) + u_2(1 - q_2)\right]}{1 + (N - j - 1)\left[u_1(1 - q_1) + u_2(1 - q_2)\right]}.
\end{aligned} \tag{S7.66}$$

Letting $j = 0$, we get the non-revised stimulus intervals for the ON neurons (Eq. 58)

$$u_{i1} = \frac{u_1}{1 + (N-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]},$$
$$u_{i2} = \frac{u_2 + (i-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]}{1 + (N-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]}, \tag{S7.67}$$
$$i \le m$$

and for the OFF neurons (Eq. 59)

$$u_{i2} = \frac{u_2 + (i-m-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]}{1 + (N-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]}, \quad i > m. \tag{S7.68}$$

The cumulative stimulus intervals then become

$$p_{11} = p_{21} = ... = p_{N1} = \frac{u_1}{1 + (N-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]} \overset{\text{def}}{=} p_1$$
$$p_{22} = ... = p_{m2} = p_{m+2,2} = ... = p_{N2} = \frac{-u_1 q_1 + u_2\left(1-q_2\right)}{1 + (N-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]} \overset{\text{def}}{=} p_2 \tag{S7.69}$$
$$p_{12} = p_{m+1,2} = \frac{u_2}{1 + (N-1)\left[u_1\left(1-q_1\right) + u_2\left(1-q_2\right)\right]} \overset{\text{def}}{=} p_{\text{edge}}$$
$$p_1 q_1 + p_2 = p_{\text{edge}}\left(1-q_2\right).$$

This derives the optimal thresholds for a mixed population of ON and OFF cells (Eq. 62, Fig. 5B).

7.7 Mean firing rate of an optimized population

Similar to neuronal populations of binary neurons (Eq. S7.70), here we can still calculate the mean firing rate ($\bar\nu$) in an optimized population of ternary neurons, based on the thresholds of Eq. S7.69. Similar to Eq. S7.70, calculating the mean firing rate yields

$$\bar\nu = \frac{1}{N}\Big[p_{\text{edge}}m + p_2\sum_{i=1}^{m-1}(m-i) + p_1\sum_{i=0}^{m-1}(m-i-1+f) +$$
$$p_{\text{edge}}(N-m) + p_2\sum_{i=1}^{N-m-1}(N-m-i) + p_1\sum_{i=0}^{N-m-1}(N-m-i-1+f)\Big]$$
$$= \frac{1}{N}\Big[p_{\text{edge}}m + (p_1+p_2)\frac{m(m-1)}{2} + p_1\,m - p_1\,m\left(1-f\right) + \tag{S7.70}$$
$$p_{\text{edge}}(N-m) + (p_1+p_2)\frac{(N-m)(N-m-1)}{2} + p_1\left(N-m\right) - p_1\left(N-m\right)\left(1-f\right)\Big]$$
$$= p_{\text{edge}} + \frac{N-1}{2}\left(p_1+p_2\right) + \frac{m}{N}\left(m-N\right)\left(p_1+p_2\right) + p_1 f.$$

Thus, we have Eq. 83 in the main text.

7.8 The maximal mutual information grows logarithmically with the number of neurons

Next, we show that the maximal mutual information for a population of ternary neurons also grows logarithmically with the number of neurons N as for binary neurons (Eq. 41). We first derive a universal relationship between the maximal mutual information $I_N^{\max}$ and the stimulus intervals p_1, p_2 for all mixtures of ON and OFF neurons. Then we apply Eq. S7.57 and Eq. S7.69 to reach the conclusion.

As previously shown (Eq. S7.21), the maximal mutual information is

$$I_N^{\max} = -\log P(\vec{0}). \tag{S7.71}$$

21

This allows us to derive the relationship between the maximal mutual information and the stimulus interval p.

For a homogeneous ON neuron population we define with $p_0 = \text{Prob}(\vec{\nu} = \vec{0})$ the 'silent' interval that is lower than all the thresholds. We also denote $q_1 = L(0, fR)$ and $q_2 = L(0, R)$ for brevity and write:

$$
\begin{aligned}
&P(\vec{0}) \\
&= p_0 + (p_1 q_1 + p_2 q_2) + (p_1 q_1 + p_2 q_2) q_2 + \ldots \\
&+ (p_1 q_1 + p_2 q_2) q_2^{N-2} + (p_1 q_1 + p_{\text{edge}} q_2) q_2^{N-1} \\
&= p_0 + (p_1 q_1 + p_2 q_2) \frac{1 - q_2^{N-1}}{1 - q_2} + p_1 q_1 q_2^{N-1} + p_{\text{edge}} q_2^{N}.
\end{aligned}
\tag{S7.72}
$$

Since all the cumulative stimulus intervals sum up to 1, we get

$$
p_0 = 1 - (N-1)(p_1 + p_2) - p_1 - p_{\text{edge}}.
\tag{S7.73}
$$

Also, from Eq. S7.69 we know

$$
p_{\text{edge}} = \frac{p_1 q_1 + p_2}{1 - q_2}.
\tag{S7.74}
$$

Substituting these two equations back into Eq. S7.72, we can derive

$$
P(\vec{0}) = 1 - N(p_1 + p_2)
\tag{S7.75}
$$

which leads to

$$
I_N^{\max} = -\log\left[1 - N(p_1 + p_2)\right].
\tag{S7.76}
$$

Since a mixed population has the same maximal information as a homogeneous population with the same p_1 and p_2 partitioning the stimulus intervals, this ensures that the maximal mutual information grows exponentially with the number of neurons as in Eq. S7.76 independent of the ON-OFF mixture.

Using the optimal values of p_1 and p_2 (Eq. S7.57 and Eq. S7.69), we have

$$
I_N^{\max} = -\log \frac{1 - u_1 (1 - q_1) - u_2 (1 - q_2)}{1 + (N-1) \left[u_1 (1 - q_1) + u_2 (1 - q_2)\right]}.
\tag{S7.77}
$$

This allows us to write the maximal mutual information of an N-neuron ternary population as a function of the the maximal mutual information of a single neuron population:

$$
I_N^{\max} = \log\left[N(\exp(I_1^{\max}) - 1) + 1\right],
\tag{S7.78}
$$

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

8 Population coding of neurons with any shapes of activation functions and any noise generation function

Here we derive the calculations with $(M+1)$-ary neurons. Because the calculations are similar to the last section, we omit some details and only show the framework of the calculations.

8.1 Maximal mutual information

Eq. S7.1 is still valid for $(M+1)$-ary neurons. The only difference from before is that every $\vec{u_i}$ or $\vec{u_i^{(m)}}$ is now an M-elements vector. Hence, we can still decompose I_N into N terms (Eq. S7.3). Similarly, we use $u_{ik}^{(j,N)}$ to denote the cumulative stimulus interval given the condition that none of the neurons $1, \ldots, j \, (j < i)$ fires (see Fig. 6A and Eq. 37) and $P_i^{(j,N)}(0) = 1 - \sum_{k=1}^{M} u_{ik}^{(j,N)} (1 - L(0, f_{ik} R))$ to denote the probability that neuron i does not fire, when none of the neurons $1, \ldots, j \, (j < i)$ fires. Similar to Eq. S7.29 and Eq. S7.31,

denoting the optimal values using an asterisk, we derive several important relationships among the revised probabilities,

$$u_{Nk}^{(N-1,N)}|_* = u_k^*, \quad u_{ik}^{(j,N)}|_* = u_{i+1,k}^{(j+1,N+1)}|_*. \tag{S8.1}$$

If neuron i and neuron $j+1$ are both ON neurons or both OFF neurons, the revised probabilities follow

$$u_{ik}^{(j+1,N)} = \frac{u_{ik}^{(j,N)}}{P_{j+1}^{(j,N)}(0)}, \quad k < M \tag{S8.2}$$

$$u_{iM}^{(j+1,N)} = \frac{1}{P_{j+1}^{(j,N)}(0)} \left[u_{iM}^{(j,N)} - \left(1 - P_{j+1}^{(j,N)}(0) \right) \right]. \tag{S8.3}$$

If neuron i is an OFF neuron but neuron $j+1$ is an ON neuron, we have

$$u_{ik}^{(j+1,N)} = \frac{u_{ik}^{(j,N)}}{P_{j+1}^{(j,N)}(0)}, \quad k < M; \quad u_{iM} = \frac{u_{iM}^{(j,N)}}{P_{j+1}^{(j,N)}(0)}. \tag{S8.4}$$

Similar to Eq. S7.4, the mutual information of one single neuron can be written as

$$I_1 = g(u_{11}, u_{12}, ..., u_{1M}). \tag{S8.5}$$

We can verify that consistent with Eq. S7.12,

$$g = \sum_{k=1}^{M} u_k \frac{\partial g}{\partial u_k} - \log P(0), \tag{S8.6}$$

which indicates that

$$I_1^{\max} = -\log P(0) = -\log \left[1 - \sum_{k=1}^{M} u_k^* \left(1 - L(0, f_k^* R) \right) \right]. \tag{S8.7}$$

Same as Section 7.2 and Eq. S7.21, we can use mathematical induction to generalize the mutual information from $N = 2$ to arbitrary N. Here, we omit it for simplicity, and give the conclusion directly as

$$I_N^{\max} = -\sum_{j=1}^{N} \log P_j^{(j-1)}(0)|_* = -\log P(\vec{0}). \tag{S8.8}$$

8.2 Optimal thresholds

Next, we seek to find the optimal thresholds by deriving the relationship among all u_{ik}. We start from a lemma that links two adjacent neurons, u_{ik} and $u_{i+1,k}$, as we did for ternary neurons (Eq. S7.32).

Lemma: For any N neurons, when I_N is maximized,

$$u_{ik}^{(i-1,N)} = P_i^{(i-1,N)}(0)u_{i+1,k}^{(i,N)}, \quad f_{ik} = f_{i+1,k}. \tag{S8.9}$$

Since the proof strictly follows those steps in Section 7.3, we do not repeat it here.

We first discuss a homogeneous population with only ON neurons. All the variables in subsequent equations are optimized, so we omit the * symbol for clarity. For simplicity we also denote $q_k = L(0, f_k R)$. For a homogeneous ON population, similar to Eq. S7.54, we have

$$u_{ik}^{(i-1,N)} = \frac{u_k}{1 + (N-i)\sum_{k=1}^{M} u_k(1-q_k)}, \quad k < M$$

$$P_i^{(i-1,N)}(0) = \frac{1 + (N-i-1)\sum_{k=1}^{M} u_k(1-q_k)}{1 + (N-i)\sum_{k=1}^{M} u_k(1-q_k)}. \tag{S8.10}$$

Using Eq. S8.2 and Eq. S8.3, we can do the inverse calculation of revising the probabilities and have

$$u_{ik}^{(j,N)} = \frac{u_k}{1 + (N - j - 1)\sum_{k=1}^{M} u_k(1 - q_k)}, \quad k < M$$

$$u_{iM}^{(j,N)} = \frac{u_M + (i - j - 1)\sum_{k=1}^{M} u_k(1 - q_k)}{1 + (N - j - 1)\sum_{k=1}^{M} u_k(1 - q_k)}.$$

(S8.11)

Letting $j = 0$, these two equations turn to

$$u_{ik}^{(j,N)} = \frac{u_k}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)}, \quad k < M$$

$$u_{iM}^{(j,N)} = \frac{u_M + (i - 1)\sum_{k=1}^{M} u_k(1 - q_k)}{1 + (N - j - 1)\sum_{k=1}^{M} u_k(1 - q_k)}.$$

(S8.12)

With the definition of p_{ik} (Eq. 69), we can write

$$p_{1k} = p_{2k} = ... = p_{Nk} = \frac{u_k}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)} \overset{\text{def}}{=} p_k, \quad k < M$$

$$p_{2M} = ... = p_{NM} = \frac{-\sum_{k=1}^{M} u_k\, q_k + u_M}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)} \overset{\text{def}}{=} p_M$$

$$p_{1M} = \frac{u_M}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)} \overset{\text{def}}{=} p_{\text{edge}}$$

$$\sum_{k=1}^{M-1} p_k\, q_k + p_M = p_{\text{edge}}\, (1 - q_M)$$

(S8.13)

which is summarized in Eq. 70 and Fig. 6B.

The equivalence to ON-OFF mixture has been discussed with ternary neurons in the main text. It still holds for $(M + 1)$-ary neurons, so that we can derive that the optimal thresholds in a mixed population are

$$p_{1k} = p_{2k} = ... = p_{Nk} = \frac{u_k}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)} \overset{\text{def}}{=} p_k, \quad k < M$$

$$p_{2M} = ... = p_{mM} = p_{m+2,M} = ... = p_{NM} = \frac{-\sum_{k=1}^{M} u_k\, q_k + u_M}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)} \overset{\text{def}}{=} p_M$$

$$p_{1M} = p_{m+1,M} = \frac{u_M}{1 + (N - 1)\sum_{k=1}^{M} u_k(1 - q_k)} \overset{\text{def}}{=} p_{\text{edge}}$$

$$\sum_{k=1}^{M-1} p_k\, q_k + p_M = p_{\text{edge}}\, (1 - q_M),$$

(S8.14)

which are also summarized in Eq. 70 and Fig. 6B.

8.3 Mean firing rate of an optimized population

Similar to neuronal populations of ternary neurons (Eq. S7.70), here we can calculate the mean firing rate $\bar{\nu}$ in an optimized population of $(M+1)$-ary neurons, based on the thresholds in Eq. S8.14. We can write

$$
\begin{aligned}
\bar{\nu} &= \frac{1}{N}\Big[p_{\text{edge}} m + p_M \sum_{i=1}^{m-1}(m-i) + \sum_{k=1}^{M-1}\sum_{i=0}^{m-1} p_k\,(m-i-1+f_k)+ \\
&\qquad p_{\text{edge}}(N-m) + p_M \sum_{i=1}^{N-m-1}(N-m-i) + \sum_{k=1}^{M-1}\sum_{i=0}^{N-m-1} p_k\,(N-m-i-1+f_k)\Big] \\
&= \frac{1}{N}\Big[p_{\text{edge}} m + \frac{1}{2}m\,(m-1)\sum_{k=1}^{M} p_k + m\sum_{k=1}^{M-1} p_k\, f_k + \\
&\qquad p_{\text{edge}}(N-m) + \frac{1}{2}(N-m)(N-m-1)\sum_{k=1}^{M} p_k + (N-m)\sum_{k=1}^{M-1} p_k\, f_k\Big] \\
&= p_{\text{edge}} + \frac{1}{2}(N-1)\sum_{k=1}^{M} p_k + \frac{m}{N}(m-N)\sum_{k=1}^{M} p_k + \sum_{k=1}^{M-1} p_k\, f_k,
\end{aligned}
\tag{S8.15}
$$

which gives rise to Eq. 84 in the main text.

8.4 The maximal mutual information grows logarithmically with the number of neurons

Similar to ternary neurons, here we show that the maximal mutual information of a population of M-ary neurons also grows logarithmically with the number of neurons N (Eq. 41). Same as before, we first derive a universal relationship between the maximal mutual information I_N^{max} and the stimulus intervals $p_k\ (k=1,...,M)$ for all mixtures of ON and OFF neurons. Then we apply Eq. S8.13 and Eq. S8.14 to reach the conclusion (Eq. 41). Similarly, we start from

$$
I_N^{\text{max}} = -\log P(\vec{0}) \tag{S8.16}
$$

to derive the relationship between the maximal mutual information and the stimulus interval p. Consider a homogeneous ON neuron population, define $p_0 = \text{Prob}(\vec{\nu}=0)$, i.e. the 'silent' interval that is lower than all the thresholds. Also denote $q_k = L(0, f_k R)$ and $q_M = L(0, R)$ for clarity, we have

$$
\begin{aligned}
P(\vec{0}) &= p_0 + \sum_{k=1}^{M} p_k\, q_k + \left(\sum_{k=1}^{M} p_k\, q_k\right) q_M + ... + \left(\sum_{k=1}^{M} p_k\, q_k\right) q_M^{N-2} + \left(\sum_{k=1}^{M-1} p_k\, q_k + p_{\text{edge}}\, q_M\right) q_M^{N-1} \\
&= p_0 + \left(\sum_{k=1}^{M} p_k\, q_k\right)\frac{1-q_M^{N-1}}{1-q_M} + \left(\sum_{k=1}^{M-1} p_k\, q_k\right) q_M^{N-1} + p_{\text{edge}} q_M^{N}.
\end{aligned}
\tag{S8.17}
$$

Note that

$$
p_0 = 1 - (N-1)\sum_{k=1}^{M} p_k - \sum_{k=1}^{M-1} p_k - p_{\text{edge}}
$$

$$
p_{\text{edge}} = \frac{\sum_{k=1}^{M-1} p_k + p_M}{1-q_M}
\tag{S8.18}
$$

we can derive

$$
P(\vec{0}) = 1 - N\sum_{k=1}^{M} p_k \tag{S8.19}
$$

which leads to

$$
I_N^{\text{max}} = -\log\left[1 - N\sum_{k=1}^{M} p_k\right]. \tag{S8.20}
$$

The ON-OFF mixed population has been shown to have the same maximal information as a homogeneous population and also the same p_k. Hence Eq. S8.20 is valid for all possible ON-OFF neuron populations. Using the optimal values of p_k (Eq. S8.13 and Eq. S8.14), we have

$$I_N^{\max} = -\log \frac{1 - \sum_{k=1}^M u_k(1 - q_k)}{1 + (N - 1)\sum_{k=1}^M u_k(1 - q_k)}. \tag{S8.21}$$

This allows us to relate the maximal mutual information of neuron populations of different sizes. Specifically,

$$I_N^{\max} = \log\left[N(\exp(I_1^{\max}) - 1) + 1\right]. \tag{S8.22}$$

9 Population coding of neurons with activation functions of any shape and heterogeneous maximal firing rates

Here we provide the calculations of the optimal thresholds of neurons with any shape of activation functions and heterogeneous maximal firing rates. Similar to before, here we still start from ternary neurons, which is the simplest case beyond binary neurons. It is straightforward that Eq. S7.3 and Eq. S7.21 still hold, i.e., for heterogeneous maximal firing rates across the cells, we still have

$$I_N = g(\vec{u_1}) + P_1(0)\left\{g\left(u_2^{\vec{(1)}}\right) + P_2^{(1,N)}(0)\left[g\left(u_3^{\vec{(2)}}\right) + ... + P_{N-1}^{(N-2,N)}(0)g\left(u_N^{\vec{(N-1)}}\right)\right]\right\}. \tag{S9.1}$$

and

$$I_N^{\max} = -\sum_{j=1}^N \log P_j^{(j-1)}(0) = -\log P(\vec{0}). \tag{S9.2}$$

In addition, the equations deriving revised probabilities (Eq. S7.58, Eq. S7.60, and Eq. S7.61) are still valid. However, the lemma (Eq. S7.32) becomes invalid due to the heterogeneous maximal firing rates across the cells. Therefore, we need to update the lemma.

9.1 Updated lemma that connects a neuron in a population to a single neuron with the same maximal firing rate

Lemma 2: For any N (ternary) neurons, when I_N is optimized,

$$u_{i1}^{(i-1,N)} = \frac{\prod_{j=i}^N P_j^{(j-1)}(0, R_j)}{P^*(0, R_i)} u_1^*(R_i)$$

$$u_{i2}^{(i-1,N)} = \frac{\prod_{j=i}^N P_j^{(j-1)}(0, R_j)}{P^*(0, R_i)} u_2^*(R_i) \tag{S9.3}$$

$$f_i = f^*(R_i)$$

where $R_i = \nu_{\max,i}T$ is the maximal expected spike count of neuron i in the population of N neurons. $u_1^*(R_i)$, $u_2^*(R_i)$, and $f^*(R_i)$ are defined in the main text as the optimal thresholds and intermediate firing rate of a single neuron with the same maximal firing rate as $\nu_{\max,i}$.

Remark: According to the definition of $u_1^*(R_i)$, $u_2^*(R_i)$, and $f^*(R_i)$, using Eqs. S7.5-S7.7, we can obtain

$$\frac{\partial g\left(u_1^*(R_i), u_2^*(R_i)\right)}{\partial u_1^*(R_i)} = \frac{\partial P^*(0, R_i)}{\partial u_1^*(R_i)}\left[-1 - \log P^*(0, R_i)\right] + L(0, f^*(R_i)R_i)\log L(0, f^*(R_i)R_i)$$

$$+ \sum_{n=1}^{+\infty} L(n, f^*(R_i)R_i)\left(\log \frac{L(n, f^*(R_i)R_i)}{u_1^*(R_i)L(n, f^*(R_i)R_i) + u_2^*(R_i)L(n, R_i)} - 1\right) = 0, \tag{S9.4}$$

$$\frac{\partial g\left(u_1^*(R_i), u_2^*(R_i)\right)}{\partial u_2^*(R_i)} = \frac{\partial P^*(0, R_i)}{\partial u_2^*(R_i)}\left[-1 - \log P^*(0, R_i)\right] + L(0, R_i)\log L(0, R_i)$$

$$+ \sum_{n=1}^{+\infty} L(n, R_i)\left(\log \frac{L(n, R_i)}{u_1^*(R_i)L(n, f^*(R_i)R_i) + u_2^*(R_i)L(n, R_i)} - 1\right) = 0, \tag{S9.5}$$

and

$$\frac{\partial g\left(u_1^*(R_i), u_2^*(R_i)\right)}{\partial f^*(R_i)} = -u_1^*(R_i)\frac{\partial L(0, f^*(R_i)R_i)}{\partial f^*(R_i)}\log P^*(0, R_i) + u_1^*(R_i)\sum_{n=0}^{+\infty}\log L(n, f^*(R_i)R_i)\frac{\partial L(n, f^*(R_i)R_i)}{\partial f^*(R_i)}$$

$$- \sum_{n=1}^{+\infty}\log[u_1^*(R_i)L(n, f^*(R_i)R_i) + u_2^*(R_i)L(n, R_i)]u_1^*(R_i)\frac{\partial L(n, f^*(R_i)R_i)}{\partial f^*(R_i)} = 0. \tag{S9.6}$$

Denote

$$\hat{u_{i1}} = \frac{\prod_{j=i}^{N} P_j^{(j-1)}(0, R_j)}{P^*(0, R_i)}u_1^*(R_i),$$

$$\hat{u_{i2}} = \frac{\prod_{j=i}^{N} P_j^{(j-1)}(0, R_j)}{P^*(0, R_i)}u_2^*(R_i), \tag{S9.7}$$

$$\hat{f}_i = f^*(R_i),$$

we need to prove

$$\frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0, \quad \frac{\partial I_N}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0, \quad \frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0 \tag{S9.8}$$

i.e. the combination of $\hat{u}_{i1}, \hat{u}_{i2},$ and $\hat{f}_i$ is optimal.

Proof:

$$\frac{\partial I_N}{\partial u_{i,1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = P_{i-1}(0)\left\{\frac{\partial g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)}{\partial u_{i1}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} + [1 - L(0, f_iR_i)]\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0)\right\}. \tag{S9.9}$$

$$\frac{\partial I_N}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = P_{i-1}(0)\left\{\frac{\partial g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)}{\partial u_{i2}^{(i-1,N)}}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} + [1 - L(0, R_i)]\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0)\right\}. \tag{S9.10}$$

$$\frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = P_{i-1}(0)\left\{\frac{\partial g\left(u_{i1}^{(i-1,N)}, u_{i2}^{(i-1,N)}\right)}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} - u_{i1}^{(i-1,N)}\frac{\partial L(0, f_iR_i)}{\partial f_i}\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0)\right\}. \tag{S9.11}$$

According to Eq. S9.4,

$$\frac{1}{P_{i-1}(0)}\frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{\partial g\left(u_1^*(R_i),u_2^*(R_i)\right)}{\partial u_1^*(R_i)}$$

$$\stackrel{Eq.\ S9.4}{=\!=\!=\!=\!=} \frac{\partial g\left(u_{i1}^{(i-1,N)},u_{i2}^{(i-1,N)}\right)}{\partial u_{i1}^{(i-1,N)}}\big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{\partial g\left(u_1^*(R_i),u_2^*(R_i)\right)}{\partial u_1^*(R_i)} + \left[1 - L(0,\hat{f}_i R_i)\right]\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0)$$

$$\stackrel{Eq.\ S7.5}{=\!=\!=\!=\!=} [1 - L(0,\hat{f}_i R_i)]\left(\log P_i^{(i-1,N)}(0) - \log P_{i+1}^{(i,N)}(0)\right) - \sum_{n=1}^{+\infty} L(n,\hat{f}_i R_i)\log\left[\hat{u}_{i1}L(n,\hat{f}_i R_i)) + \hat{u}_{i2}L(n,R_i))\right]$$

$$+ \sum_{n=1}^{+\infty} L(n,f^*(R_i)R_i)\log\left[u_1^*(R_i)L(n,f^*(R_i)R_i) + u_2^*(R_i)L(n,R_i)\right]$$

$$+ [1 - L(0,f^*(R_i)R_i)]\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0)$$

$$= [1 - L(0,\hat{f}_i R_i)]\left(\log P_i^{(i-1,N)}(0) - \log P^*(0,R_i)\right) - \sum_{n=1}^{+\infty} L(n,\hat{f}_i R_i)\log\frac{\prod_{j=i}^{N} P_j^{(j-1)}(0,R_j)}{P^*(0,R_i)}$$

$$+ \left[1 - L(0,\hat{f}_i R_i)\right]\sum_{j=i+1}^{N}\log P_j^{(j-1,N)}(0).$$

$$(S9.12)$$

Because

$$L(0,\hat{f}_i R_i) + \sum_{n=1}^{+\infty} L(n,\hat{f}_i R_i) = 1 \quad \text{(the sum of probabilities)}, \tag{S9.13}$$

$$\frac{1}{P_{i-1}(0)}\frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{\partial g\left(u_1^*(R_i),u_2^*(R_i)\right)}{\partial u_1^*(R_i)} = 0, \quad \frac{\partial I_N}{\partial u_{i1}^{(i-1,N)}}\big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} = 0. \tag{S9.14}$$

Similarly, we have

$$\frac{\partial I_N}{\partial u_{i2}^{(i-1,N)}}\big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} = 0. \tag{S9.15}$$

Also,

$$\frac{1}{\hat{u}_{i1}P_{i-1}(0)}\frac{\partial I_N}{\partial f_i}\big|_{\hat{u}_{i1},\hat{u}_{i2},\hat{f}_i} - \frac{1}{u_1^*(R_i)}\frac{\partial g\left(u_1^*(R_i),u_2^*(R_i)\right)}{\partial f^*}$$

$$\stackrel{Eq.\ S9.6}{=\!=\!=\!=\!=} -\frac{\partial L(0,\hat{f}_i R_i)}{\partial \hat{f}_i}\log P_i(0) + \sum_{n=0}^{+\infty}\log L(n,\hat{f}_i R_i)\frac{\partial L(n,\hat{f}_i R_i)}{\partial \hat{f}_i}$$

$$- \sum_{n=1}^{+\infty}\log\left[\hat{u}_{i1}L(n,\hat{f}_i R_i) + \hat{u}_{i2}L(n,R_i)\right]\frac{\partial L(n,\hat{f}_i R_i)}{\partial \hat{f}_i} - \frac{\partial L(0,\hat{f}_i R_i)}{\partial \hat{f}_i}\sum_{j=i+1}^{N}\log P_j^{(j-1)}(0) \qquad (S9.16)$$

$$+ \frac{\partial L(0,f^*(R_i)R_i)}{\partial f^*(R_i)}\log P^*(0,R_i) - \sum_{n=0}^{+\infty}\log L(n,f^*(R_i)R_i)\frac{\partial L(n,f^*(R_i)R_i)}{\partial f^*(R_i)}$$

$$+ \sum_{n=1}^{+\infty}\log\left[u_1^*(R_i)L(n,f^*(R_i)R_i) + u_2^*(R_i)L(n,R_i)\right]\frac{\partial L(n,f^*(R_i)R_i)}{\partial f^*(R_i)}.$$
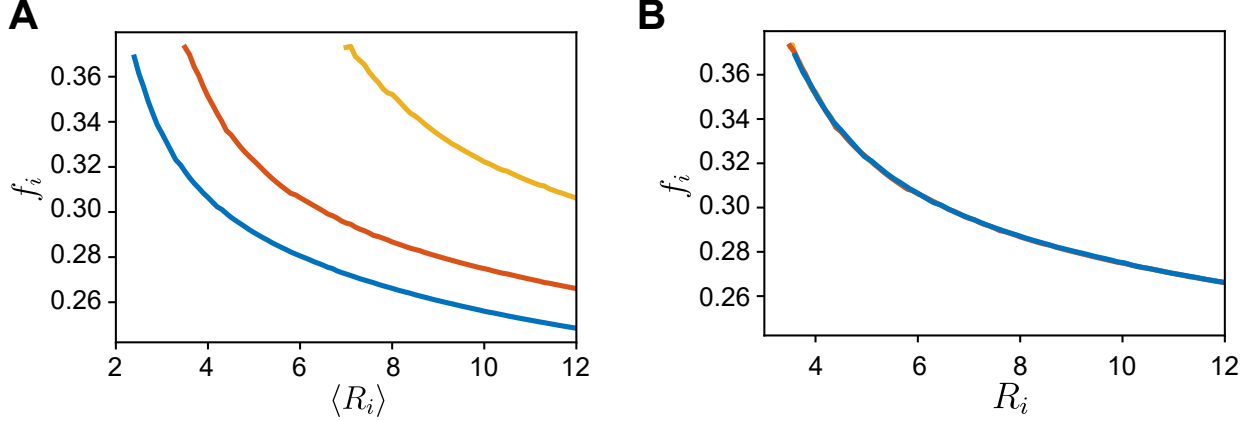
Figure S2. Optimal intermediate firing levels in a population with heterogeneous maximal firing rate across the cells. The population consists of $N = 3$ ternary ON neurons, with $R_1 : R_2 : R_3 = 1 : 2 : 3$. **A.** Optimal intermediate firing levels f_i as a function of $\langle R_i \rangle = (R_1 + R_2 + R_3)/3$. **B.** Optimal intermediate firing levels f_i as a function of R_i of individual neurons.

According to the definition of $\hat{f}_i$, $\hat{u}_{i1}$ and $\hat{u}_{i2}$ (Eq. S7.37), we have

$$
\begin{aligned}
&\frac{1}{\hat{u}_{i1} P_{i-1}(0)} \frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} - \frac{1}{u_1^*(R_i)} \frac{\partial g\left(u_1^*(R_i), u_2^*(R_i)\right)}{\partial f^*} \\
&= \frac{\partial L(0, \hat{f}_i R_i)}{\partial \hat{f}_i} \left(\log P^*(0, R_i) - \log P_i^{(i-1,N)}(0)\right) - \log \frac{\prod_{j=i}^{N} P_j^{(j-1)}(0, R_j)}{P^*(0, R_i)} \sum_{n=1}^{+\infty} \frac{\partial L(n, \hat{f}_i R_i)}{\partial \hat{f}_i} \\
&= \sum_{n=0}^{+\infty} \frac{\partial L(n, \hat{f}_i R_i)}{\partial \hat{f}_i} \left(\log P^*(0, R_i) - \log P_i^{(i-1,N)}(0)\right) \\
&= \left(\log P^*(0, R_i) - \log P_i^{(i-1,N)}(0)\right) \frac{\partial}{\partial \hat{f}_i} \left[\sum_{n=0}^{+\infty} L(n, \hat{f}_i R_i)\right].
\end{aligned}
\tag{S9.17}
$$

Since $\sum_{n=0}^{+\infty} L(n, \hat{f}_i R_i) = 1$ is a constant, $\frac{\partial}{\partial \hat{f}_i}\left[\sum_{n=0}^{+\infty} L(n, \hat{f}_i R_i)\right] = 0$, then

$$
\frac{1}{\hat{u}_{i1} P_{i-1}(0)} \frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} - \frac{1}{u_1^*(R_i)} \frac{\partial g\left(u_1^*(R_i), u_2^*(R_i)\right)}{\partial f^*} = 0
\tag{S9.18}
$$

which leads to

$$
\frac{\partial I_N}{\partial f_i}\Big|_{\hat{u}_{i1}, \hat{u}_{i2}, \hat{f}_i} = 0.
\tag{S9.19}
$$

Combined with Eq. S9.14 and Eq. S9.15, the lemma has been proved. ∎

9.2 Optimal thresholds for a population with heterogeneous maximal firing rates

Using Eq. S7.59, Eq. S7.63, and Eq. S7.64, we can do the inverse calculation of revising the probabilities and obtain

$$
\begin{aligned}
u_{i1}^{(j,N)} &= \frac{\prod_{k=j+1}^{N} P_k^{(k-1)}(0, R_k)}{P^*(0, R_i)} u_1^*(R_i), \\
u_{i2}^{(j,N)} &= \frac{\prod_{k=j+1}^{N} P_k^{(k-1)}(0, R_k)}{P^*(0, R_i)} u_2^*(R_i) + 1 - \prod_{k=j+1}^{i-1} P_k^{(k-1)}(0)
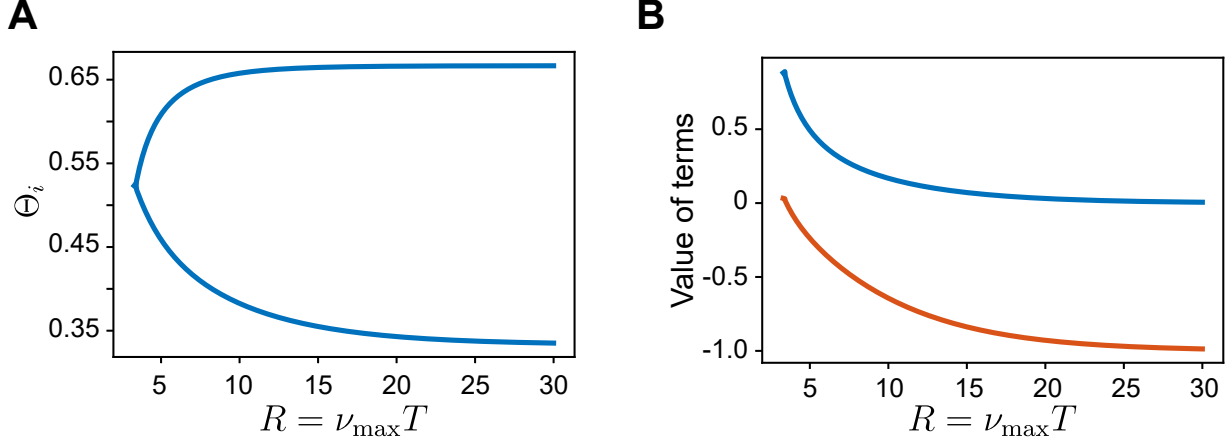\end{aligned}
\tag{S9.20}
$$

29

Figure S3. Numerical calculations for a population of ternary neurons with heterogeneous maximal firing rates. A. Optimal thresholds (u_1, u_2) of a single ternary neuron. **B.** The values of two terms in Eq. 78. Blue: $\left(u_2^*(R) - u_1^*(R)\right)/P^*(0, R)$. Orange: $\left[u_2^*(R)\, q_2^*(R) - u_1^*(R)\left(1 - q_1^*(R)\right)\right]/P^*(0, R)$.

if neuron i and neuron j are both ON neurons ($j < i \le m$) or both OFF neurons ($i > m$ and $j \ge m$), and

$$
\begin{aligned}
u_{i1}^{(j,N)} &= \frac{\prod_{k=j+1}^{N} P_k^{(k-1)}(0, R_k)}{P^*(0, R_i)}\, u_1^*(R_i), \\
u_{i2}^{(j,N)} &= \frac{\prod_{k=j+1}^{N} P_k^{(k-1)}(0, R_k)}{P^*(0, R_i)}\, u_2^*(R_i) + \prod_{k=j+1}^{m} P_k^{(k-1)}(0) - \prod_{k=j+1}^{i-1} P_k^{(k-1)}(0)
\end{aligned}
\tag{S9.21}
$$

if neuron i is an OFF neuron but neuron j is an ON neuron ($i > m$, $j \le m$).

Letting $j = 0$, this becomes the non-revised stimulus intervals (Eq. 74), which is

$$
\begin{aligned}
u_{i1} &= \frac{u_1^*(R_i)}{P^*(0, R_i)} \prod_{k=1}^{N} P_k^{(k-1)}(0, R_k) \\
u_{i2} &= \frac{u_2^*(R_i)}{P^*(0, R_i)} \prod_{k=1}^{N} P_k^{(k-1)}(0, R_k) + 1 - \prod_{k=1}^{i-1} P_k^{(k-1)}(0, R_k) \\
&\quad \text{(ON neurons, } i = 1, ..., m) \\
u_{i2} &= \frac{u_2^*(R_i)}{P^*(0, R_i)} \prod_{k=1}^{N} P_k^{(k-1)}(0, R_k) + \prod_{k=1}^{m} P_k^{(k-1)}(0, R_k) - \prod_{k=1}^{i-1} P_k^{(k-1)}(0, R_k) \\
&\quad \text{(OFF neurons, } i = m+1, ..., N) \\
f_i &= f^*(R_i).
\end{aligned}
\tag{S9.22}
$$

Note that with

$$
I_N = -\log P(\vec{0}) = \prod_{j=1}^{N} P_j^{(j-1)}(0, R_j)
\tag{S9.23}
$$

and

$$
\begin{aligned}
P_i^{(i-1)}(0, R_i) &= 1 - u_{i1}^{(i-1)}\left(1 - q_1(R_i)\right) - u_{i2}^{(i-1)}\left(1 - q_2(R_i)\right) \\
&= 1 - \left[\frac{u_1^*(R_i)}{P(0, R_i)}\left(1 - q_1(R_i)\right) + \frac{u_2^*(R_i)}{P(0, R_i)}\left(1 - q_2(R_i)\right)\right] \prod_{j=i}^{N} P_j^{(j-1)}(0, R_j),
\end{aligned}
\tag{S9.24}
$$

we can obtain

$$P_i^{(i-1)}(0, R_i) = \left[1 + \frac{u_1^*(R_i)}{P(0, R_i)}\left(1 - q_1(R_i)\right) + \frac{u_2^*(R_i)}{P(0, R_i)}\left(1 - q_2(R_i)\right)\right]^{-1},$$ (S9.25)

and then

$$\left(\prod_{j=i}^{N} P_j^{(j-1)}(0, R_j)\right)^{-1} = \sum_{j=i}^{N} \frac{1}{P(0, R_j)} - (N - i).$$ (S9.26)

With Eq. S9.26, we can rewrite the optimal thresholds (Eq. S9.22) as

$$u_{i1} = \frac{u_1^*(R_i)}{P^*(0, R_i)} e^{-I_N}$$

$$u_{i2} = \left[\frac{u_2^*(R_i)}{P^*(0, R_i)} - \sum_{j=i}^{N} \frac{1}{P^*(0, R_j)} + (N - i)\right] e^{-I_N} + 1 \qquad \text{(ON neurons, } i = 1, ..., m)$$

$$u_{i2} = \left[\frac{u_2^*(R_i)}{P^*(0, R_i)} - \sum_{j=i}^{N} \frac{1}{P^*(0, R_j)} - \sum_{j=1}^{m} \frac{1}{P^*(0, R_j)} + (N - i + m)\right] e^{-I_N} + 1 \quad \text{(OFF neurons, } i = m+1, ..., N)$$

$$f_i = f^*(R_i).$$
(S9.27)

According to the definition of p_{i1} and p_{i2}, we can write

$$p_{i1} = \frac{u_1^*(R_i)}{P^*(0, R_i)} e^{-I_N}$$

$$p_{i2} = \left[\frac{u_2^*(R_i) - u_1^*(R_i)}{P^*(0, R_i)} - \frac{u_2^*(R_{i-1}) q_2^*(R_{i-1}) - u_1^*(R_{i-1})\left(1 - q_1^*(R_{i-1})\right)}{P^*(0, R_{i-1})}\right] e^{-I_N} \quad (i = 2, ..., m, m+2, ..., N)$$

$$p_{12} = \left[\frac{u_2^*(R_1)}{P^*(0, R_1)} - \sum_{j=1}^{N} \frac{1}{P^*(0, R_j)} + (N - 1)\right] e^{-I_N} + 1$$

$$p_{m+1,2} = \left[\frac{u_2^*(R_{m+1})}{P^*(0, R_{m+1})} - \sum_{j=1}^{N} \frac{1}{P^*(0, R_j)} + (N - 1)\right] e^{-I_N} + 1$$

$$f_i = f^*(R_i).$$
(S9.28)

This derives the optimal thresholds for a mixed population of ON and OFF ternary cells with heterogeneous maximal firing rates (Eq. 62, Fig. 5B).

Here, $f_i = f^*(R_i)$ indicates that the optimal intermediate firing level of a neuron in a population only depends on the maximal firing rate of that specific neuron (Fig. S2). Consequently, 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).

We performed numerical calculations for Poisson noise to better understand this optimal thresholds structure (Fig. S3). We found that $u_1^*(R)$ increases with R (Fig. S3A). Since $P^*(0, R)$ decreases with R, within a population, p_{i1} is larger for neurons with higher R_i. To analyze how p_{i2} depends on R_i and R_{i-1}, we plotted the two terms in Eq. S9.28 (second line), as functions of R (Fig. S3B). We found that both of them decrease with R (Fig. S3B), which indicates that within a neuron population, p_{i2} is smaller for higher R_i and lower R_{i-1}.

Again, using Eq. S9.26, the maximal mutual information can be rewritten as

$$I_N = \log\left[\sum_{j=i}^{N} \frac{1}{P^*(0, R_j)} - (N - 1)\right],$$ (S9.29)

9.3 $(M+1)$-ary neurons

Above we derived the maximal mutual information solution for ternary neurons with heterogeneous maximal firing rates. Generalizing this from ternary neurons to activation functions with any any shape is straightforward (see Section 8). For $(M+1)$-ary neurons, the optimal thresholds are

$$u_{ik} = \frac{u_k^*(R_i)}{P^*(0,R_i)}e^{-I_N} \quad k = 1,...,M-1$$

$$u_{iM} = \left[\frac{u_M^*(R_i)}{P^*(0,R_i)} - \sum_{j=i}^{N}\frac{1}{P^*(0,R_j)} + (N-i)\right]e^{-I_N} + 1 \quad \text{(ON neurons, } i=1,...,m)$$

$$u_{iM} = \left[\frac{u_M^*(R_i)}{P^*(0,R_i)} - \sum_{j=i}^{N}\frac{1}{P^*(0,R_j)} - \sum_{j=1}^{m}\frac{1}{P^*(0,R_j)} + (N-i+m)\right]e^{-I_N} + 1 \quad \text{(OFF neurons, } i=m+1,...,N)$$

$$f_{ik} = f_k^*(R_i). \quad k = 1,...,M-1,$$

(S9.30)

the maximal mutual information is:

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

(S9.31)

and the optimal stimulus intervals p are given by

$$p_{ik} = \frac{u_1^*(R_i)}{P^*(0,R_i)}e^{-I_N} \quad k = 1,...,M-1$$

$$p_{iM} = \left[\frac{u_M^*(R_i) - \sum_{k=1}^{M-1}u_k^*(R_i)}{P^*(0,R_i)} - \frac{u_M^*(R_{i-1})q_M^*(R_{i-1}) - \sum_{k=1}^{M-1}u_k^*(R_{i-1})(1-q_k^*(R_{i-1}))}{P^*(0,R_{i-1})}\right]e^{-I_N}$$

$$(i = 2,...,m,m+2,...,N)$$

$$p_{1M} = \left[\frac{u_M^*(R_1)}{P^*(0,R_1)} - \sum_{j=1}^{N}\frac{1}{P^*(0,R_j)} + (N-1)\right]e^{-I_N} + 1$$

$$p_{m+1,M} = \left[\frac{u_2^*(R_{m+1})}{P^*(0,R_{m+1})} - \sum_{j=1}^{N}\frac{1}{P^*(0,R_j)} + (N-1)\right]e^{-I_N} + 1$$

$$f_{ik} = f_k^*(R_i), \quad k = 1,...,M-1.$$

(S9.32)

References

[1] A. P. Nikitin, N. G. Stocks, R. P. Morse, and M. D. McDonnell, "Neural population coding is optimized by discrete tuning curves," *Physical Review Letters*, vol. 103, no. 13, pp. 138101:1–4, 2009.

[2] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[3] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proceedings I Communications, Speech and Vision*, vol. 137, no. 6, pp. 424–430, 1990.

[4] D. Kraft, *A Software Package for Sequential Quadratic Programming*. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht, Wiss. Berichtswesen d. DFVLR, 1988.

[5] J. Gjorgjieva, M. Meister, and H. Sompolinsky, "Functional diversity among sensory neurons from efficient coding principles," *PLoS Computational Biology*, vol. 15, no. 11, pp. e1007476:1–38, 2019.