

## Capacity of a Single Spiking Neuron Channel

**Shiro Ikeda**

*shiro@ism.ac.jp*

*Department of Mathematical Analysis and Statistical Inference, Institute of Statistical Mathematics, Tokyo 106-8569, Japan*

**Jonathan H. Manton**

*jmanton@unimelb.edu.au*

*Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia*

Information transfer through a single neuron is a fundamental component of information processing in the brain, and computing the information channel capacity is important to understand this information processing. The problem is difficult since the capacity depends on coding, characteristics of the communication channel, and optimization over input distributions, among other issues. In this letter, we consider two models. The temporal coding model of a neuron as a communication channel assumes the output is  $\tau$  where  $\tau$  is a gamma-distributed random variable corresponding to the interspike interval, that is, the time it takes for the neuron to fire once. The rate coding model is similar; the output is the actual rate of firing over a fixed period of time. Theoretical studies prove that the distribution of inputs, which achieves channel capacity, is a discrete distribution with finite mass points for temporal and rate coding under a reasonable assumption. This allows us to compute numerically the capacity of a neuron. Numerical results are in a plausible range based on biological evidence to date.

### 1 Introduction ---

It is widely believed that neurons send information to other neurons in the form of spike trains. Although precise timings of spikes are important for information transfer, it appears that spike patterns are not deterministic but noisy (Mainen & Sejnowski, 1995). Information theory shows that when a communication channel is corrupted with noise, the rate at which the information can be transmitted reliably through the channel is limited. The upper bound on the rate is known as the channel capacity (Shannon, 1948) (in the rest of the letter, it is referred to simply as “capacity”). When a single neuron is considered as a channel, the capacity is one of the fundamental problems in neuroscience.

The problem has been studied theoretically (MacKay & McCulloch, 1952; Rapoport & Horvath, 1960; Stein, 1967) and biologically (Borst & Theunissen, 1999). Computing capacity is difficult since it depends on multiple factors—type of coding, characteristics of the channel, and input distributions. The type of coding has long been a subject of discussion (MacKay & McCulloch, 1952; Baker & Lemon, 2000; Rullen & Thorpe, 2001). Mainly two types of coding, temporal and rate coding, have been considered. Temporal coding uses interspike intervals (ISIs) to code information, and rate coding uses the number of spikes in a fixed interval. This letter examines both of them.

The channel model is deeply related to the noise of ISIs. Baker and Lemon (2000) reported that the statistical properties of ISIs recorded from primary motor cortex and supplementary motor area (SMA) of monkeys are similar to the gamma distribution. Shinomoto, Shima, and Tanji (2003) and Shinomoto, Miyazaki, Tamura, and Fujita (2005) studied spike trains from multiple areas and proposed a statistical index that describes the randomness of ISIs.<sup>1</sup> The index is deeply related to the gamma distribution (Shinomoto et al., 2003; Ikeda, 2005). In this letter, ISIs are modeled with a gamma distribution. The model is different from the channel model in MacKay and McCulloch (1952), where spikes are assumed to be aligned within a fixed time precision.

The capacity is defined as the supremum of mutual information over possible input distributions. In this letter, a natural assumption is posed, that is, the average firing rate of a single neuron is restricted in an interval. Under this assumption, we consider all possible input distributions and prove that the capacity of each coding is achieved by a discrete distribution that has only finite mass points. The proof of the discreteness of capacity-achieving distributions for each coding shares the steps with other studies of information theory (Smith, 1971; Shamai (Shitz), 1990; Abou-Faycal, Trott, & Shamai (Shitz), 2001; Gursoy, Poor, & Verdú, 2002, 2005). These studies have shown the discreteness for some channels with appropriate assumptions on the input distributions. Our result shows that the information is maximally transmitted through a single neuron when the inputs to the neuron have only a fixed number of “modes.” This is important for biological experiments, since if the input distribution is discrete, the experimentalists have to consider only discrete and finite modes of inputs or stimuli. After the proof, the capacity and the capacity-achieving distribution for each coding are computed. Unfortunately we have not obtained any analytical solution, and they are computed numerically. The results show that the capacity is around 15 to 50 bits

---

<sup>1</sup>Data from pre-SMA, SMA, rostral cingulate motor area (CMar), and prefrontal cortical area (PF) of monkeys are studied in Shinomoto et al. (2003), while data from different layers of area TE of monkeys are studied in Shinomoto et al. (2005).

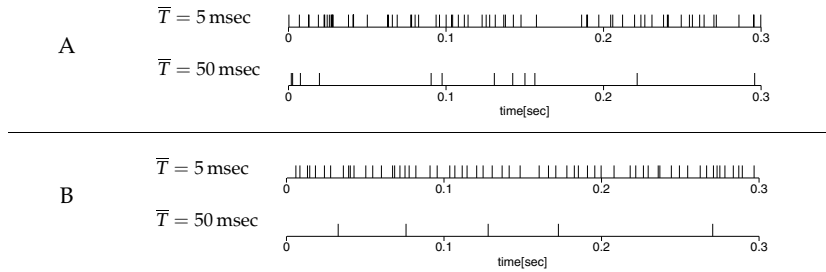


Figure 1: Simulated spike trains. ISIs follow a gamma distribution, where the shape parameter  $\kappa$  is 0.75 for *A* and 4.5 for *B*. The expected values of ISI are 5 msec in the upper trains and 50 msec in the lower trains for both *A* and *B*.

per sec, the same order with the values reported in Borst and Theunissen (1999).

The problem is formulated mathematically in section 2, and the discreteness for each coding is proved in section 3. Section 4 shows numerical studies, and the final section concludes with some discussion. Most of the mathematical proofs are summarized in the appendix.

## 2 Single Neuron Channel

**2.1 ISIs and Communication Channel.** It has been reported that a gamma distribution is a suitable model to describe the stochastic nature of ISIs (Baker & Lemon, 2000; Shinomoto et al., 2003). The gamma distribution has two parameters: the shape parameter  $\kappa$  and the scale parameter  $\theta$ . From some studies,  $\kappa$  of individual neuron appears to be constant (the value of  $\kappa$  may depend on the type of neuron), while  $\theta$  changes dynamically over time.

Figure 1 shows simulated spike trains with two different shape parameter  $\kappa$ 's. It is 0.75 in Figure 1A and 4.5 in Figure 1B. When  $\kappa$  is small, spike trains become more irregular. Ikeda (2005) and Miura, Okada, and Amari (2006) studied the estimation methods of  $\kappa$  from spike trains. Estimation of  $\kappa$  is regarded as the semiparametric statistical estimation (Bickel, Klaassen, Ritov, & Wellner, 1993).

In this letter, we focus not on the estimation but on the information processing of a single neuron. Based on the gamma distribution model, the capacity of a neuron is investigated in the following sections.

**2.2 Communication Channel and Capacity.** Let  $X$  be the input to a noisy channel and  $Y$  be the output. In the following, we assume  $X \in \mathcal{X} \subseteq \mathbb{R}$  is a one-dimensional stochastic variable, and let  $F(\cdot)$  be a cumulative

distribution function of  $X$ . Communication channel is defined as a stochastic model described as  $p(y | x)$ , and the mutual information is defined as

$$I(X; Y) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} p(y | x) \log \frac{p(y | x)}{p(y)} d\mu(y) dF(x), \quad (2.1)$$

where  $p(y) = \int_{x \in \mathcal{X}} p(y | x) dF(x)$ .

Here,  $\mu(y)$  denotes the measure of  $y \in \mathcal{Y}$ . Since the channel is defined as  $p(y | x)$ ,  $I(X; Y)$  is a functional of  $F(\cdot)$ , and we denote it as  $I(F)$ .

Let  $\mathcal{F}$  be the set of cumulative distribution functions of  $X$ . The channel capacity is defined as

$$C = \sup_{F \in \mathcal{F}} I(F). \quad (2.2)$$

For a noisy channel, one interesting fundamental problem is to compute the capacity  $C$ . Another interesting problem is to obtain the distribution, if it exists, which achieves the capacity.

**2.3 Single Neuron: Channel and Coding.** Let us discuss a neuron model. First, we have to define  $X$  and  $Y$  of a neuron communication channel.

The distribution of each ISI is assumed to be independent and to follow a gamma distribution. Let  $T$  denote an ISI, a stochastic variable following a gamma distribution, that is,  $T \sim \Gamma(\kappa, \theta)$ , where  $\kappa > 0$  and  $\theta > 0$  are the shape and the scale parameter, respectively.

We assume  $\kappa$  of each neuron is fixed and known. Shinomoto et al. (2003) define a statistical index  $L_V$  (local index), to characterize each neuron. For a  $T \sim \Gamma(\kappa, \theta)$ ,  $\bar{L}_V = 3/(2\kappa + 1)$  holds. From their investigation with biological data, it seems most of the cells'  $L_V$  are lying in an interval (0.3–1.2), and  $\kappa$  is thus assumed to be in an interval  $\kappa \in [\kappa_m, \kappa_M]$  ( $\kappa_m$  and  $\kappa_M$  are set to 0.75 and 4.5, respectively, in section 4).

Under the assumption, the scale parameter  $\theta$  is the only variable parameter that plays the role of input, that is,  $X$  in section 2.2. The density function of  $t$  is

$$p(t | \theta; \kappa) = \left( \frac{t^{\kappa-1}}{\theta^\kappa} \right) \frac{\exp[-t/\theta]}{\Gamma(\kappa)}, \quad \kappa, \theta > 0, t \geq 0,$$

where we denote it as  $p(t | \theta; \kappa)$  to show  $\theta$  is a stochastic variable and  $\kappa$  is a parameter. The gamma distribution is an exponential family:

$$p(t | \theta; \kappa) = \exp \left[ -\frac{1}{\theta} t + (\kappa - 1) \log t - \log \Gamma(\kappa) - \kappa \log \theta \right]. \quad (2.3)$$

The sufficient statistics are  $T$  and  $\log T$ . The expectations of them are

$$\overline{T} = \kappa\theta, \quad \overline{\log T} = \psi(\kappa) + \log \theta,$$

where  $\psi(\cdot)$  is the digamma function defined as  $\psi(x) = \Gamma'(x)/\Gamma(x)$  for  $x > 0$ . The conditional entropy becomes

$$\begin{aligned} H(T \mid \theta; \kappa) &= - \int_0^\infty p(t \mid \theta; \kappa) \log p(t \mid \theta; \kappa) dt \\ &= \kappa - (\kappa - 1)\psi(\kappa) + \log \Gamma(\kappa) + \log \theta. \end{aligned}$$

Next, let us consider the family of all the possible distributions of input  $\theta$ . Noting that ISI is positive and is not infinite if the neuron is active, it is natural to assume that the average ISI, which depends on  $\theta$  and  $\kappa$ , is limited between  $a_0$  and  $b_0$  ( $a_0$  and  $b_0$  are set to 5 msec and 50 msec, respectively, in section 4), that is,

$$a_0 \leq \overline{T} = \kappa\theta \leq b_0, \quad \text{where } 0 < a_0 < b_0 < \infty.$$

Thus,  $\theta$  is bounded in  $\Theta(\kappa) = \{\theta \mid a(\kappa) \leq \theta \leq b(\kappa)\}$ , where  $a(\kappa)$  and  $b(\kappa)$  are defined as

$$a(\kappa) = a_0/\kappa, \quad b(\kappa) = b_0/\kappa.$$

In the following,  $a(\kappa)$ ,  $b(\kappa)$ , and  $\Theta(\kappa)$  are denoted as  $a$ ,  $b$ , and  $\Theta$ , respectively, as far as no confusion arises. Let us define  $F(\theta)$  as the cumulative distribution function of  $\theta$  and  $\mathcal{F}$  as the set of all possible  $F(\theta)$ , that is,

$$\begin{aligned} \mathcal{F} = \{F : \mathbb{R} \rightarrow [0, 1] \mid &F(\theta) = 0, \quad (\forall \theta < a), \\ &F(\theta) = 1, \quad (\forall \theta > b)\}. \end{aligned} \quad (2.4)$$

Note that every  $F \in \mathcal{F}$  is right-continuous and nondecreasing on  $\Theta$ , and  $\mathcal{F}$  includes continuous and discrete distributions of  $\theta$ .<sup>2</sup>

Next, let us consider  $Y$ , the output of the channel of a neuron communication channel. There are mainly two different ideas in neuroscience. One idea is that  $Y$  is ISI,  $T$ , itself (see MacKay & McCulloch, 1952, for example). This is called temporal coding (see Figure 2). The other is that  $Y$  is the rate, which is the number of spikes in fixed time intervals (see Stein, 1967). This is called rate coding (see Figure 2). In communication theory, "coding" is often used

---

<sup>2</sup>In Stein (1967), the distribution of  $\theta$  was assumed to be discrete. We do not assume it in this letter.

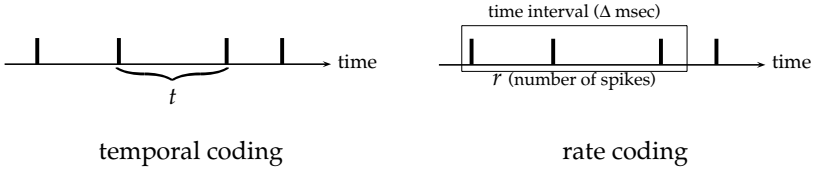


Figure 2: Two types of coding: temporal coding and rate coding.

for “source coding,” “error-control coding,” and “cryptography coding.” It seems that *modulation* is a more suitable term for the above definition. However, we follow the standard usage of the neuroscience community. How to encode (or to modulate) the input  $\theta$  to the neuron channel depends on which coding is used. For temporal coding,  $\theta$  is fixed during the interval  $t$ , while  $\theta$  is fixed during  $\Delta$  for the rate coding. We discuss this in section 5.

Mutual information and the capacity also depend on coding. The capacity of each coding is formally defined in the following.

**2.3.1 Temporal Coding.** In temporal coding, received information is  $T$ . For  $F \in \mathcal{F}$ , we define the marginal distribution as

$$p(t; F, \kappa) = \int_a^b p(t | \theta; \kappa) dF(\theta), \tag{2.5}$$

where  $p(t | \theta; \kappa)$  is defined in equation 2.3. The existence of  $p(t; F, \kappa)$  follows from the existence of  $p(t | \theta; \kappa)$ . The mutual information of  $T$  and  $\theta$  is defined as

$$I_T(F) = \int_a^b i_T(\theta; F) dF(\theta), \quad \text{where} \\ i_T(\theta; F) = \int_0^\infty p(t | \theta; \kappa) \log \frac{p(t | \theta; \kappa)}{p(t; F, \kappa)} dt. \tag{2.6}$$

Let us define  $g(t; F, \kappa)$  and rewrite  $p(t; F, \kappa)$  as

$$g(t; F, \kappa) = \int_a^b \frac{\exp[-t/\theta]}{\theta^\kappa} dF(\theta), \quad p(t; F, \kappa) = \frac{t^{\kappa-1}}{\Gamma(\kappa)} g(t; F, \kappa). \tag{2.7}$$

The mutual information  $I_T(F)$  is rewritten as

$$I_T(F) = h_T(F; \kappa) - \kappa h_{T|\theta}(F; \kappa) - \kappa,$$

where

$$h_T(F; \kappa) = - \int_0^\infty p(t; F, \kappa) \log g(t; F, \kappa) dt,$$

$$h_{T|\theta}(F; \kappa) = \int_a^b \log \theta dF(\theta).$$

Hence, the capacity per channel use or equivalently per spike is defined as

$$C_T = \sup_{F \in \mathcal{F}} I_T(F) = \sup_{F \in \mathcal{F}} (h_T(F; \kappa) - \kappa h_{T|\theta}(F; \kappa)) - \kappa.$$

The capacity  $C_T$  and the distribution that achieves  $C_T$  are studied in the next section.

**2.3.2 Rate Coding.** In rate coding, a time window is set, and the spikes in the interval are counted. Let us denote the interval and the rate as  $\Delta$  and  $R$ , respectively, and define the distribution of  $R$  as  $p(r | \theta; \kappa, \Delta)$ . The form of the distribution of  $R$  is shown in the following lemma:

**Lemma 1.** *The distribution  $p(r | \theta; \kappa, \Delta)$  has the following form,*

$$p(r | \theta; \kappa, \Delta) = P(r\kappa, \Delta/\theta) - P((r+1)\kappa, \Delta/\theta), \quad r \in \mathbb{Z}^*, \quad (2.8)$$

where  $\mathbb{Z}^*$  denotes the set of nonnegative integers and  $P(\alpha, x)$  is the regularized incomplete gamma function:

$$P(0, x) = 1, \quad P(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt, \quad \text{for } \alpha, x > 0.$$

**Proof.** See Appendix A. The same distribution is discussed in Pawlas, Klevanov, and Prokop (2008).

When  $\kappa = 1$ , a gamma distribution is an exponential distribution, and the distribution of  $R$  becomes a Poisson distribution:

$$p(r | \theta; 1, \Delta) = \left(\frac{\Delta}{\theta}\right)^r \frac{\exp[-\Delta/\theta]}{r!}.$$

For an  $F \in \mathcal{F}$ , let us define the following marginal distribution  $p(r; F, \kappa, \Delta)$ :

$$p(r; F, \kappa, \Delta) = \int_a^b p(r | \theta; \kappa, \Delta) dF(\theta).$$

The existence of the integral follows from the existence of  $p(r | \theta; \kappa, \Delta)$ . The mutual information of  $R$  and  $\theta$  is defined as

$$I_R(F) = \int_a^b i_R(\theta, F) dF(\theta), \quad \text{where}$$

$$i_R(\theta, F) = \sum_{r=0}^{\infty} p(r | \theta; \kappa, \Delta) \log \frac{p(r | \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)}. \quad (2.9)$$

Hence, the capacity per channel use or, equivalently, per  $\Delta$  is defined as

$$C_R = \sup_{F \in \mathcal{F}} I_R(F).$$

The capacity  $C_R$  and the distribution that achieves  $C_R$  are studied in the next section.

### 3 Theoretical Studies

---

The cumulative distribution  $F \in \mathcal{F}$  is a right-continuous nondecreasing function on a interval  $\Theta$ . Thus,  $\theta$  can be a discrete or continuous random variable over  $\Theta$ . In this section, the capacity-achieving distribution of a single neuron channel is proved to be a discrete distribution with finite mass points for both temporal and rate coding.

For some channels, the capacity-achieving distributions have been shown to be discrete under some conditions (Smith, 1971; Shamai (Shitz), 1990; Abou-Faycal et al., 2001; Gursoy et al., 2002; Tchamkerten, 2004). The neuron channel with temporal coding is different from those because it does not have an additive noise and the proof must be provided independently. The rate coding with  $\kappa = 1$  is equivalent to the Poisson channel, and the discreteness of the capacity-achieving distribution is proved in Shamai (Shitz) (1990). The proof is easily extended to the case where  $\kappa$  is a positive integer. But we need to prove it for positive real  $\kappa$ 's. Note that although the proofs of the discreteness in this letter are original, they follow the same steps of those papers.

**3.1 Steps to Prove the Discreteness of the Capacity-Achieving Distribution.** The common steps of the proof for the discreteness of the capacity-achieving distributions are shown in this section. In the following, the results of optimization theory and probability theory will be used. Suppose  $X$  is a normed linear space. In optimization theory, the space of all bounded linear functionals of  $X$  is called the normed dual of  $X$  and is denoted  $X^*$ . The weak\* convergence is defined as follows:

**Definition 1.** A sequence  $\{x_n^*\}$  in  $X^*$  is said to converge weak\* to the element  $x^*$  if for every  $x \in X$ ,  $x_n^*(x) \rightarrow x^*(x)$ . In this case, we write  $x_n^*(x) \xrightarrow{w^*} x^*(x)$  (Luenberger, 1969, 5.10).



If  $X$  is the real normed linear space of all bounded continuous functions on  $\mathbb{R}$ ,  $X^*$  includes the set of all probability measures, and it is clear that “weak convergence” of probability measures is “weak\* convergence” on  $X^*$ . The results of optimization theory are applied to probability measures with this equivalence. The following theorem is used to prove the existence and the uniqueness of the capacity-achieving distribution:

**Theorem 1.** *Let  $J$  be a weak\* continuous real-valued functional on a weak\* compact subset  $S$  of  $X^*$ . Then  $J$  is bounded on  $S$  and achieves its maximum on  $S$ . If  $S$  is convex and  $J$  is strictly concave, then the maximum,*

$$C = \max_{x^* \in S} J(x^*), \quad (3.1)$$

*is achieved by a unique  $x^*$  in  $S$ .*

**Proof.** See Luenberger (1969, 5.10), Abou-Faycal et al. (2001), and Gursoy et al. (2002).

From the above discussion,  $\mathcal{F}$  in equation 2.4 is a subset of  $X^*$ . It is clear that  $\mathcal{F}$  is convex. Thus, if  $\mathcal{F}$  is weak\* compact and  $I_T(F)$  (or  $I_R(F)$ ) is a weak\* continuous function on  $F \in \mathcal{F}$  and strictly concave in  $\mathcal{F}$ , the capacity is achieved by a unique distribution  $F_0$  in  $\mathcal{F}$ . This is the first step of the proof. The following proposition states  $\mathcal{F}$  is compact.

**Proposition 1.**  *$\mathcal{F}$  in equation 2.4 is compact in the Lévy metric topology.*

**Proof.** For the proof of compactness, see Smith (1971) (proof of proposition 1). The proof is a direct application of Helly’s compactness theorem (Doob, 1994, sec. X).

The Kuhn-Tucker (K-T) condition on the mutual information is used for the next step of the proof. Before showing the condition, we define the weak differentiability:

**Definition 2.** *Let  $J$  be a function on a convex set  $\mathcal{F}$ . Let  $F_0$  be a fixed element of  $\mathcal{F}$  and  $\eta \in [0, 1]$ . Suppose there exists a map  $J'_{F_0} : \mathcal{F} \rightarrow \mathbb{R}$  such that*

$$J'_{F_0}(F) = \lim_{\eta \downarrow 0} \frac{J((1-\eta)F_0 + \eta F) - J(F_0)}{\eta}, \quad F \in \mathcal{F}.$$

*Then  $J$  is said to be weakly differentiable in  $\mathcal{F}$  at  $F_0$  and  $J'_{F_0}(F)$  is the weak derivative in  $\mathcal{F}$  at  $F_0$ . If  $J$  is weakly differentiable in  $\mathcal{F}$  at  $F_0$  for all  $F \in \mathcal{F}$ ,  $J$  is said to be weakly differentiable in  $\mathcal{F}$ .*

The K-T condition is described as follows:

**Proposition 2.** *Assume  $J$  is a weakly differentiable, concave functional on a convex set  $\mathcal{F}$ . If  $J$  achieves its maximum on  $\mathcal{F}$  at  $F_0$ , then a necessary and sufficient condition for  $F_0$  to attain the maximum is to satisfy the following inequality for all  $F \in \mathcal{F}$ :*

$$J'_{F_0}(F) \leq 0.$$

**Proof.** See proposition 1 in Smith (1971).

If  $I_T(F)$  (or  $I_R(F)$ ) is weakly differentiable, the K-T condition is derived immediately with the theorem. Finally, the discreteness is proved by deriving a contradiction based on the K-T condition and the assumption that  $F_0$  has infinite mass points as its support. Thus, in order to show the discreteness of the capacity-achieving distribution for temporal and rate codings, the following properties must be shown:

1.  $I_T(F)$  and  $I_R(F)$  are weak\* continuous on  $\mathcal{F}$  and strictly concave.
2.  $I_T(F)$  and  $I_R(F)$  are weakly differentiable.

After these are shown, the K-T condition is derived, and the discreteness and the finiteness will be checked.

**3.2 Discreteness of the Capacity-Achieving Distribution for Temporal Coding.** In this section, the capacity-achieving distribution for temporal coding is shown to be a discrete distribution with a finite number of points. We start with the following lemma:

**Lemma 2.**  *$I_T(F)$  in equation 2.6 is a weak\* continuous function on  $F \in \mathcal{F}$  and strictly concave in  $\mathcal{F}$ .*

**Proof.** Section B.1 proves  $I_T(F)$  is a weak\* continuous function.  $I_T(F)$  can be proved to be strictly concave following the proof of lemma 2 in Abou-Faycal et al. (2001).

Lemma 2 and theorem 1 imply that the capacity for temporal coding is achieved by a unique distribution in  $\mathcal{F}$ . In order to show it is a discrete distribution, the following lemma and corollary are used:

**Lemma 3.**  *$I_T(F)$  in equation 2.6 is weakly differentiable in  $\mathcal{F}$ . The weak derivative at  $F_0 \in \mathcal{F}$  has the form*

$$I'_{T,F_0}(F) = \int_a^b i_T(\theta; F_0) dF - I_T(F_0) - \kappa. \quad (3.2)$$

**Proof.** See section B.2.

**Corollary 1.** *Let  $E_0$  denote the points of increase of  $F_0(\theta)$  on  $\theta \in [a, b]$ .  $F_0$  is optimal if and only if*

$$\begin{aligned} i_T(\theta; F_0) &\leq I_T(F_0) + \kappa, \quad \forall \theta \in \Theta \\ i_T(\theta; F_0) &= I_T(F_0) + \kappa, \quad \forall \theta \in E_0. \end{aligned} \tag{3.3}$$

**Proof.** This is proved following the same steps in Smith (1971, corollary 1), with equation 3.2.

The main result of this section is summarized in the following theorem:

**Theorem 2.** *Under the constraint  $\theta \in \Theta$ , the channel capacity of a single neuron channel with temporal coding is achieved by a discrete distribution with a finite number of mass points.*

**Proof.** The extension of  $i_T(\theta; F_0)$  to the complex plain  $z$  is analytic for  $\text{Re } z > 0$ , which is defined as

$$i_T(z; F_0) = -\kappa \log z - \int_0^\infty p(t | z; \kappa) g(t; F_0, \kappa) dt.$$

If  $E_0$  in corollary 1 has infinite points, since  $\Theta$  is bounded and closed,  $E_0$  has a limit point. Hence, from corollary 1, the identity theorem implies  $i_T(z; F_0) = I_T(F_0) + \kappa$  for the region  $\text{Re } z > 0$ . This region includes positive real line  $\mathbb{R}^+$ , and

$$-\int_0^\infty p(t | \theta; \kappa) \log g(t; F_0, \kappa) dt = \kappa \log \theta + I_T(F_0) + \kappa, \quad \theta \in \mathbb{R}^+ \tag{3.4}$$

is implied. The left-hand side of equation 3.4 is bound as follows (see section B.1, equation B.4):

$$\begin{aligned} \frac{1}{b} \int_0^\infty t p(t | \theta; \kappa) dt + \kappa \log a &\leq - \int_0^\infty p(t | \theta; \kappa) \log g(t; F_0, \kappa) dt \\ &\leq \frac{1}{a} \int_0^\infty t p(t | \theta; \kappa) dt + \kappa \log b. \end{aligned} \tag{3.5}$$

Since the expectation of  $T$  with regard to  $p(t | \theta; \kappa)$  is  $\kappa\theta$ , equation 3.5 shows that the left-hand side of equation 3.4 grows linearly with  $\theta$ . Since the right-hand side increases only with  $\log \theta$ , equation 3.4 cannot hold for all  $\theta \in \mathbb{R}^+$ . This is the contradiction, and the optimal distribution has a finite number of mass points.

**3.3 Discreteness of the Capacity-Achieving Distribution for Rate Coding.** The capacity-achieving distribution for rate coding is shown to be a discrete distribution with a finite number of points. Shamaï (Shitz) (1990) proved that the capacity-achieving distribution of a Poisson channel under peak and average power constraints is a discrete distribution with a finite point of supports. Since  $\theta \in \Theta$  is a peak constraint, this directly proves the case  $\kappa = 1$ . For  $\kappa \neq 1$ , further study is needed.

**Lemma 4.**  $I_R(F)$  in equation 2.9 is a weak\* continuous function on  $F \in \mathcal{F}$  and strictly concave in  $\mathcal{F}$ .

**Proof.** Section C.1 proves  $I_R(F)$  is a weak\* continuous function. The concavity of  $I_R(F)$  can be proved as in Abou-Faycal et al. (2001). The proof for the strict concavity follows the proof in section 7.2 of Shamaï (Shitz) (1990), which is an application of Carleman's theorem (Akhiezer, 1965).

Lemma 4 and theorem 1 imply that the capacity for rate coding is achieved by a unique distribution in  $\mathcal{F}$ :

**Lemma 5.**  $I_R(F)$  in equation 2.9 is weakly differentiable in  $\mathcal{F}$ . The weak derivative at  $F_0 \in \mathcal{F}$  has the form

$$I'_{R,F_0}(F) = \int_a^b i_R(\theta; F_0) dF - I_R(F_0). \quad (3.6)$$

**Proof.** The proof is identical to the proof of lemma 3 in section B.2.

**Corollary 2.** Let  $E_0$  denote the points of increase of  $F_0(\theta)$  on  $\theta \in [a, b]$ .  $F_0$  is optimal if and only if

$$\begin{aligned} i_R(\theta; F_0) &\leq I_R(F_0), \quad \forall \theta \in \Theta \\ i_R(\theta; F_0) &= I_R(F_0), \quad \forall \theta \in E_0. \end{aligned} \quad (3.7)$$

**Proof.** This is proved following the same steps in Smith (1971, corollary 1) with equation 3.6.

Finally, the following theorem proves that the capacity-achieving distribution is a discrete distribution with a finite number of mass points:

**Theorem 3.** Under a peak constraint, the channel capacity of a single neuron channel with the rate coding is achieved by a discrete distribution with a finite number of mass points.

**Outline of proof.** The proof follows the same steps of theorem 2. The extension of  $i_R(\theta; F)$  to the complex plain  $z$  is defined as

$$i_R(z; F) = \sum_{r=0}^{\infty} p(r | z; \kappa, \Delta) \log \frac{p(r | z; \kappa, \Delta)}{p(r; F, \kappa, \Delta)},$$

$$p(r | z; \kappa, \Delta) = P(r\kappa, \Delta/z) - P((r+1)\kappa, \Delta/z).$$

Since  $P(\alpha, z)$  and  $\log z$  is analytic for  $\text{Re } z > 0$ ,  $i_R(z; F_0)$  is analytic for  $\text{Re } z > 0$ .

If  $E_0$  in corollary 2 has infinite points, since  $\Theta$  is bounded and closed,  $E_0$  has a limit point and, hence, from equation 3.7, the identity theorem implies  $i_R(z; F_0) = I_R(F_0)$  for the region  $\text{Re } z > 0$ . This region includes positive real line  $\mathbb{R}^+$ , and

$$\sum_{r=0}^{\infty} p(r | \theta; \kappa, \Delta) \log \frac{p(r | \theta; \kappa, \Delta)}{p(r; F_0, \kappa, \Delta)} = I_R(F_0), \quad \theta \in \mathbb{R}^+ \quad (3.8)$$

is implied. The proof (see section C.2) is completed by deriving a contradiction for equation 3.8. The contradiction is derived for  $\kappa \geq 1$  and  $\kappa < 1$  separately.

## 4 Numerical Studies

---

Although the capacity-achieving distribution of each coding has been proved to be discrete with a finite number of mass points, position and probability of each point are not provided. Unfortunately, we do not have an analytic solution. This is also the case for related work (Smith, 1971; Shamai (Shitz), 1990; Abou-Faycal et al., 2001; Gursoy et al., 2005). In this section, the capacity and the capacity-achieving distribution are computed numerically for temporal and rate coding.

**4.1 Common Steps of Numerical Experiments.** Computing the capacity and the capacity-achieving distribution of the neuron channel is difficult since the closed-form expression of  $i_T(\theta; F)$  in equation 2.6 and  $i_R(\theta; F)$  in equation 2.9 is not provided for a general discrete  $F(\theta)$ . Instead, we need to evaluate integrals for  $i_T(\theta; F)$  and summations of infinite series for  $i_R(\theta; F)$ . For the numerical studies, integrals for  $i_T(\theta; F)$  are evaluated with the Gauss-Laguerre quadrature, and infinite series for  $i_R(\theta; F)$  are truncated to sufficiently long finite series.

The strategy to compute the capacity and the capacity-achieving distributions for temporal and rate coding is as follows. Note that related work uses similar methods (Smith, 1971; Abou-Faycal et al., 2001; Gursoy et al., 2005).

1. Initialize  $N$ , the number of points, as 2.
2. Set the position and probability of each point as  $\theta_j$  and  $\pi_j$ , ( $j = \{1, \dots, N\}$ ), respectively, where

$$a(\kappa) \leq \theta_1 < \dots < \theta_N \leq b(\kappa),$$

$$\sum_{j=1}^N \pi_j = 1, \quad \pi_j > 0, \quad j \in \{1, \dots, N\}.$$

3. Starting from some initial values, maximize the corresponding mutual information ( $I_T(F)$  or  $I_R(F)$ ) with respect to  $\{\theta_i\}$  and  $\{\pi_i\}$  until convergence with a gradient method.
4. When it converges, check the corresponding K-T condition in (equations 3.3 or 3.7) to see if it is the capacity-achieving distribution.
5. If the K-T condition is satisfied, the capacity and the capacity-achieving distribution are obtained. Otherwise, increase  $N$  by 1 and go to step 2.

The range of  $\theta$  must be specified for the numerical studies. The range of the expected firing rate is defined as from 5 msec to 50 msec and  $5/\kappa \leq \theta \leq 50/\kappa$ . The choice of the range is discussed in section 5.

The capacity and the capacity-achieving distribution for temporal and rate coding are computed for multiple values of  $\kappa$ . As described in section 2.3, a statistical index  $L_V$  has been proposed that characterizes spike trains (Shinomoto et al., 2003). Its expectation is related to  $\kappa$  as  $\overline{L_V} = 3/(2\kappa + 1)$ . In the following numerical studies, we vary  $\kappa$  from 0.75 to 4.5 (the corresponding  $\overline{L_V}$  is from 0.3 to 1.2) for every 0.05. The range corresponds to the most of the cells'  $L_V$  in Shinomoto et al. (2003, 2005).

**4.2 Temporal Coding.** Figure 3A shows the computed capacity for each  $\kappa$ . The capacity  $C_T$  (bit per channel use) increases monotonically as  $\kappa$  increases.<sup>3</sup> This is natural since as  $\kappa$  increases, ISIs become more regular, and more information can be transferred. The capacity becomes larger than 1 bit when  $\kappa$  becomes 3.85.

The capacity-achieving distributions are shown in Figures 3C and 3D. For each  $\kappa$ , the distribution has only two or three points. Moreover, two of them are ends of the range  $\Theta(\kappa)$  ( $a_0/\kappa$  and  $b_0/\kappa$ ). If  $\kappa$  is smaller than 2.10, there are only two points. When it is equal to 2.10, the number of points becomes three. The position of the third point is very stable for different  $\kappa$ 's. The probability of each point is shown in Figure 3D. The probabilities of both ends tend to be similar, while the probability of the third point increases gradually as  $\kappa$  increases.

The capacity  $C_T$ , is the maximum information transferred per spike. It is also important to show the information rate. Since the capacity-achieving

---

<sup>3</sup>We used bit instead of nat by dividing capacity defined in equation 2.2 by  $\log 2$ .

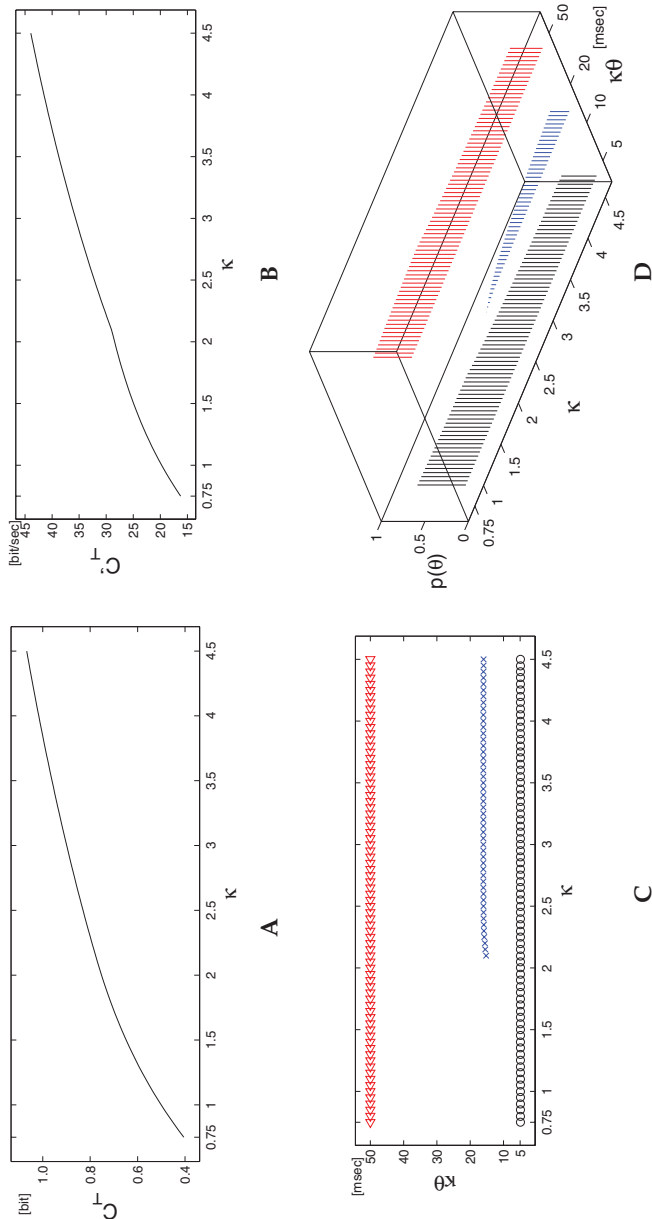


Figure 3: Numerical results of temporal coding. (A) Capacity  $C_T$  (bit per channel use) for each  $\kappa$ . (B) Information rate  $C_r$  (bit per sec) for each  $\kappa$  (see equation 4.1). C and D show the capacity-achieving distribution computed for each  $\kappa$ . (C) The probability mass points. For every  $\kappa$ , two points are on the edges of  $\Theta(\kappa)$  ( $b_0/\kappa$  and  $b_0/\kappa$ , shown as  $\circ$  and  $\diamond$ , respectively). The third point  $\times$  appears as  $\kappa$  becomes 2.10. (D) Probability of each point shown as the height. The axis for  $\kappa\theta$  is logarithmically scaled for visual clarity.

distribution is computed, the following  $C'_T$  (bit per sec) is defined:

$$C'_T = C_T / \bar{\tau}, \quad \text{where } \bar{\tau} = \kappa \sum_{j=1}^N \pi_j \theta_j. \quad (4.1)$$

Note that  $\bar{\tau}$  is around 25 msec for all  $\kappa$  in the experiments. The information rate is shown as the function of  $\kappa$  in Figure 3B. Further discussion is provided in section 5.

**4.3 Rate Coding.** In rate coding, the time window  $\Delta$  must be defined. Since the average time for sending a symbol with temporal coding is around 25 msec,  $\Delta$  is set to 25 msec in the numerical experiment.

Figure 4A shows the computed channel capacity for each  $\kappa$ .  $C_R$  increases monotonically as  $\kappa$  increases. The value is larger than  $C_T$  for the same  $\kappa$ . It becomes larger than 1 bit when  $\kappa$  becomes 2.15.

The capacity-achieving distributions are shown in Figures 4C and 4D. For each  $\kappa$ , the distribution has two to four discrete points, and two of them are ends of the range  $\Theta(\kappa)$  ( $a_0/\kappa$  and  $b_0/\kappa$ ). For  $\kappa < 1.25$ , there are only two points. For  $1.25 \leq \kappa < 4$ , there are three points, and it becomes four for  $\kappa \geq 4.0$ . The probability of each point is shown in Figure 3D. The probabilities of both ends tend to be similar, while the probability of the third point increases gradually as  $\kappa$  increases. When the number of mass points is four, two middle points have similar probability.

In rate coding, the information rate is easily computed. Since  $\Delta$  is fixed, the rate is computed as  $C'_R = C_R / \Delta$  (bit per sec), which is shown in Figure 4B.

## 5 Discussion and Conclusion

---

We have proved the channel capacities of a single neuron with temporal and rate coding are achieved with discrete distributions. Numerical studies show that the number of mass points is from two to four depending on coding and  $\kappa$ . The capacity of a single neuron evaluated in this letter is lower than what has been reported in MacKay and McCulloch (1952) and Rapoport and Horvath (1960) (1000 to 4000 bits per sec), and its order is similar to biologically measured capacities of sensory neurons (Borst & Theunissen, 1999). However, this does not mean the capacity can be achieved biologically. The problem has been simplified in our study, and the details should be discussed. Since channel capacity depends on various factors, each factor is discussed separately in the rest of this section.

**5.1 Encoding: Input Distribution of  $\theta$ .** First, we discuss the input  $\theta$ . Since the ISI is positive and is not infinite if the neuron is active, the constraint ( $\alpha \leq \theta \leq \beta$ ) seems to be natural. The range of  $\theta$  has been set to [5 msec, 50 msec] throughout the letter. The firing rate of each neuron



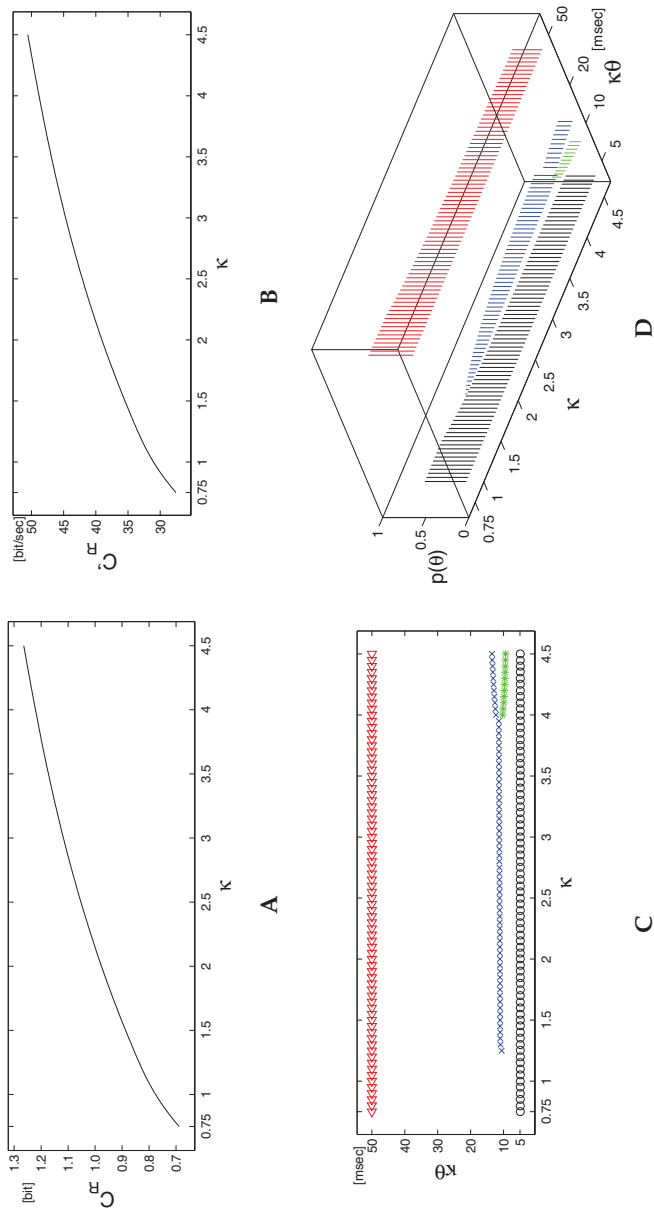


Figure 4: Numerical results of rate coding. (A) Capacity  $C_R$  (bit per channel use) for each  $\kappa$ . (B) Information rate  $C'_R$  (bit per sec) for each  $\kappa$ . C and D show the capacity-achieving distribution computed for each  $\kappa$ . (C) The probability mass points. For every  $\kappa$ , two points are on the ends of the range ( $a_0$  and  $b_0$ , shown as  $\circ$  and  $\times$ , respectively). The third point  $\times$  appears as  $\kappa$  becomes larger than 1.20. The fourth point  $*$  appears as  $\kappa$  becomes 4.0. (D) Probability of each point shown as the height. The axis for  $\kappa\theta$  is logarithmically scaled for visual clarity.

depends on its type, and this range may not be plausible for some neurons. Note that for temporal coding, if the “dynamic range” of the firing rate is 10 dB, the capacity per channel use is identical to the result of this letter. The capacity of rate coding depends on the dynamic range and  $\Delta$ ; therefore, the capacity result of this letter may not be appropriate for some neurons.

In the range  $\Theta(\kappa)$ , the distribution of  $\theta$  has been assumed to be memoryless, that is,  $\theta$  can be different for every channel use. Scale parameter  $\theta$  must be changed every 5 msec at most in temporal coding and 25 msec in rate coding. Biologically speaking,  $\theta$  corresponds to the input to a neuron, and it cannot be changed quickly since the neuron has capacitance. Thus, the source would have memory. This implies the biologically achievable rate should be smaller than the capacity obtained in the numerical studies.

Another problem is the duration to keep the input  $\theta$ , especially for temporal coding. When  $\theta$  is fixed for some duration, the neuron fires according to the gamma distribution; however, the “sender” cannot know when the “receiver” receives the spike. In order to detect an ISI, the receiver must receive two spikes, and it is not clear how the sender can be synchronized with the receiver. One idea is to have a common clock and fix  $\theta$  in an interval. This situation turns out to be rate coding. Another idea is to fix  $\theta$  for a time proportional to the expected ISI,  $\kappa\theta$ . In this case, the receiver may miss some spikes. In either case, the transmitted information will be lower than the numerically computed capacity.

When  $\kappa = 1$ , the rate coding becomes identical to the “Poisson channel” (Bar-David, 1969; Shamai (Shitz), 1990; Guo, Shamai (Shitz), & Verdú, 2008). There is a great deal of work on the Poisson channel communication, and many types of constraints on the input distributions have been considered (Verdú, 1999, provides a summary of Poisson channel communications). Our constraint is a memoryless peak energy constraint, and other constraints can be added. One of the commonly used constraints is the average energy constraint, that is,  $\overline{\theta^2} \leq C$ . Even if we add an average energy constraint to a peak power constraint, we believe the optimal distribution is still discrete for each coding. This has been proved for Poisson channel in Shamai (Shitz) (1990), and its extension to general values of  $\kappa$  seems possible. For the temporal coding, the proof can be straightforwardly extended, as in Smith (1971). However, we do not know how to set  $C$ , which prevents us from employing an average energy constraint. Note that adding an average energy constraint possibly makes the set  $\mathcal{F}$ , and thus the capacity, smaller, and our result is the upper bound of the capacity with an average energy constraint.

The capacity-achieving distributions are discrete distributions with finite mass points. Although this is good in the sense that neurons can transfer information maximumly with discrete numbers of “firing modes,” this does not imply neurons are using only discrete modes. The input of each neuron may vary continuously. The result in this letter shows that even if the input has rich information, the sender cannot send more information than a Markovian source with finite discrete states.

**5.2 Noisy Channel Model.** Characteristics of neurons strongly depend on their types. MacKay and McCulloch (1952) assumed that a neuron is able to fire within a fixed time precision. They have concluded that the each spike can carry up to 9 bits of information, and approximately 1000 to 3000 bits per second could be transferred theoretically. Compared to some biological studies summarized in Borst and Theunissen (1999), this observation might be optimistic. We modeled the stochastic property of them with a gamma function. This is quite different from the model in MacKay and McCulloch (1952).

We set the value of  $\kappa$  between 0.75 to 4.5, which has been indicated in Shinomoto et al. (2003); however, in Baker and Lemon (2000),  $\kappa$  is set to 16, which is much larger than our choice.<sup>4</sup> As  $\kappa$  increases, the capacity and the number of mass points of the capacity achieving distribution increase; therefore, the capacity and the number of mass points for  $\kappa = 16$  would be much larger than our numerical results. We have not shown numerical results for  $\kappa = 16$ , since it is difficult to carry out numerical experiments with a large  $\kappa$  because of numerical precision. This may be solved in the future.

It is also interesting to consider the communication channel with multiple neurons. If there are  $m$  neurons, which follows the same gamma distribution  $\Gamma(\kappa, \theta)$ , the sum of ISIs follows  $\Gamma(m\kappa, \theta)$  and the average of ISIs follows  $\Gamma(m\kappa, \theta/m)$ . Since the channel capacity  $C_T$  and  $C_R$  increases as  $\kappa$  increases, the channel capacity will be larger with multiple neurons. Note that the capacity-achieving distribution is still a discrete distribution with finite probability mass points.

**5.3 Decoding.** The capacity is the maximum of transferred information. In order to achieve the capacity, the receiver must act as an optimal decoder. Let us define the position and probability of the capacity-achieving distribution for temporal and rate codings as  $\{\theta_{T,i}, \pi_{T,i}\}$  and  $\{\theta_{R,i}, \pi_{R,i}\}$ ,  $i = 1, \dots, N$ , respectively. The optimal decoder for the temporal decoder computes the following posterior probability when  $t$  is observed,

$$\varpi_{T,i} = \frac{\pi_{T,i} p(t | \theta_{T,i}; \kappa)}{\sum_{i'} \pi_{T,i'} p(t | \theta_{T,i'}; \kappa)}, \quad i \in \{1, \dots, N\},$$

while the optimal decoder of the rate coding computes the following posterior probability when  $r$  is observed:

$$\varpi_{R,i} = \frac{\pi_{R,i} p(r | \theta_{R,i}; \kappa, \Delta)}{\sum_{i'} \pi_{R,i'} p(r | \theta_{R,i'}; \kappa, \Delta)}, \quad i \in \{1, \dots, N\}.$$

---

<sup>4</sup>One of the reviewers indicated the value of  $\kappa$  might be much larger than that given by current references in the literature.

The discrete distributions  $\varpi_{T,i}$  and  $\varpi_{R,i}$  are the posterior distributions of the input  $\theta$  conditioned on the observations. This “soft decoding” is natural from a mathematical viewpoint; however, it may not be plausible to assume that the postsynaptic neuron is computing  $\varpi_{T,i}$  and  $\varpi_{R,i}$  since the computation is complicated and the value of  $\kappa$  must be known by the neuron.

Another natural decoding is hard decoding; depending on  $t$  or  $r$ , only a single  $\theta$  is considered as the decoding result. The Bayes optimal hard decoding is to choose the  $\theta_i$  which maximizes the posterior distribution. In the case of single-neuron information channels, the hard coding results for temporal and rate coding are defined as, respectively,

$$\hat{\theta}_T = \underset{\theta_{T,i}}{\operatorname{argmax}} \varpi_{T,i} = \underset{\theta_{T,i}}{\operatorname{argmax}} \pi_{T,i} p(t | \theta_{T,i}; \kappa),$$

$$\hat{\theta}_R = \underset{\theta_{R,i}}{\operatorname{argmax}} \varpi_{R,i} = \underset{\theta_{R,i}}{\operatorname{argmax}} \pi_{R,i} p(r | \theta_{R,i}; \kappa, \Delta).$$

Each decoder becomes a simple threshold function. Figure 5 shows the hard decoding boundaries for temporal and rate coding. In temporal coding,  $t$  is a nonnegative real number, and decision boundaries are shown in Figure 5A. In rate coding,  $r$  is a nonnegative integer, and decisions for integers are shown in Figure 5B.

Note that the boundary in Figure 5A is stable between  $\kappa = 0.75$  to 2.6, and even if the capacity-achieving distribution has three states for  $\kappa \geq 2.10$  (see Figure 3), the third point does not appear in decisions until  $\kappa > 2.6$ . Similar results are observed for rate coding. Although the number of points is more than 3 if  $\kappa > 1.20$  (see Figure 4), the decision becomes three points only when  $\kappa > 1.55$ . Even if the number of points is four for  $\kappa \geq 4.00$ , it does not appear as the hard decision. Decision boundaries are not sensitive to small changes of  $\kappa$ .

When hard decoding is employed, both input and output are discrete, and transferred information can be computed easily. The transferred information with the capacity-achieving distribution and the optimal hard decoders are shown in Figure 6. It shows that the transferred information is degraded from the optimal soft decoder; however, the lost information is not very large.

**5.4 Related Work.** Stein (1967) has discussed the channel capacity of the rate coding where a gamma distribution with a fixed  $\kappa$  was the ISI model. The input was assumed to be a discrete distribution of the scale parameter  $\theta$  on an interval, which happened to be optimal, and the capacity was computed numerically in a similar manner.

Although the assumption corresponds to the optimal distribution, the discreteness had not been proved. We believe this letter is the first to prove the discreteness of the optimal distribution for general  $\kappa$ , not only for rate coding but also for temporal coding.

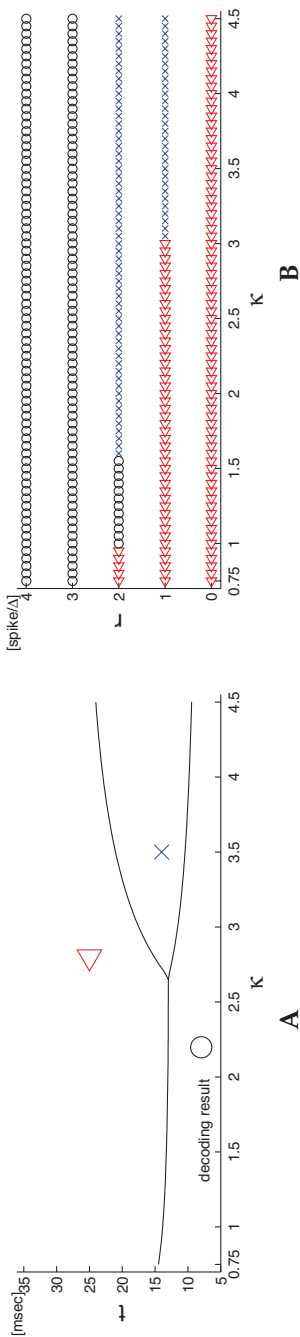


Figure 5: Hard decoding. (A) Hard decoding boundaries for temporal coding. The decision depends on the received ISI and  $\kappa$ .  $\circ$ ,  $\triangleleft$ , and  $\times$  correspond to  $\theta$ 's in Figure 3. (B) Hard decoding for rate coding. The decision depends on the received rate and  $\kappa$ .  $\circ$ ,  $\triangleleft$ , and  $\times$  correspond to  $\theta$ 's in Figure 4.

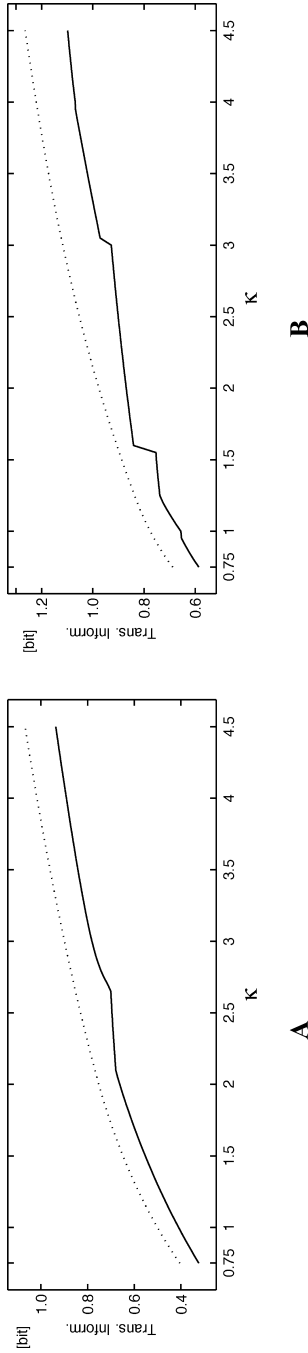


Figure 6: Transferred information per channel use with hard decoders. (A) Transferred information with the capacity-achieving distribution and the optimal hard decoder for temporal coding. The dotted line shows the capacity in Figure 3A. (B) Transferred information with the capacity-achieving distribution and the optimal hard decoder for rate coding. The dotted line shows the capacity in Figure 4A.

Finally we note that similar work has emerged recently in a slightly different context (McDonnell & Stocks, 2008; Nikitin, Stocks, Morse, & McDonnell, 2008).

**5.5 Conclusion.** The channel capacity and the capacity-achieving distribution are obtained for a single neuron information channel. ISIs are modeled with a gamma distribution, and two types of coding, temporal and rate, are considered. Capacity-achieving distributions are proved to be discrete distributions with a finite number of points. Numerical studies show that the number of the points is relatively small for a moderate choice of  $\kappa$ . It should also be noted that neurons may not use efficient error-control codes, which requires a fairly long delay. Instead, the actual encoding and decoding may be very simple and far from optimal as far as the rate is concerned.

The result does not necessarily imply that the neuron is using discrete states as ISIs or that the decoding is soft decoding. However, the information capacity gives the upper bound of the information, which can be transferred through a single neuron. This limit has implications. If the input is a continuous distribution, the transferred information is lower than the capacity, and if hard decoding is employed, the transferred information is lower than the capacity.

In neurophysiological experiments, many trials are accumulated because signals are generally noisy. The results of this letter provide a general guide for how much information could be obtained through a single recording. Also it gives suggestions for the field of brain-machine interface or brain-computed interface (BCI), which tries to extract information from neurons' spikes.

## Appendix A: Proof of Lemma 1

---

The lemma is proved by induction.

$p(0 | \theta; \kappa, \Delta)$  is the probability that  $T$  is larger than  $\Delta$ . Since  $T \sim \Gamma(\kappa, \theta)$

$$\begin{aligned} p(0 | \theta; \kappa, \Delta) &= 1 - \int_0^\Delta \left( \frac{t^{\kappa-1}}{\theta^\kappa} \right) \frac{\exp[-t/\theta]}{\Gamma(\kappa)} dt \\ &= 1 - \frac{1}{\Gamma(\kappa)} \int_0^{\Delta_\theta} u^{\kappa-1} e^{-u} du = 1 - P(\kappa, \Delta_\theta), \end{aligned}$$

where  $\Delta_\theta = \Delta/\theta$ . Assuming equation 2.8 is true for a  $m \in \mathbb{Z}^*$ ,  $p(m+1 | \theta; \kappa, \Delta)$  is written as follows:

$$\begin{aligned} p(m+1 | \theta; \kappa, \Delta) &= \int_0^\Delta \left( \frac{t^{\kappa-1}}{\theta^\kappa} \right) \frac{\exp[-t/\theta]}{\Gamma(\kappa)} p(m | \theta; \kappa, (\Delta - t)) dt \\ &= \frac{1}{\Gamma(\kappa)} \int_0^{\Delta_\theta} u^{\kappa-1} e^{-u} [P(m\kappa, (\Delta_\theta - u)) \\ &\quad - P((m+1)\kappa, (\Delta_\theta - u))] du. \end{aligned} \tag{A.1}$$

If the following relation holds for  $m \in \mathbb{Z}^*$ , it completes the proof:

$$\frac{1}{\Gamma(\kappa)} \int_0^{\Delta_\theta} u^{\kappa-1} e^{-u} P(m\kappa, (\Delta_\theta - u)) du = P((m+1)\kappa, \Delta_\theta). \quad (\text{A.2})$$

Equation A.2 is easily checked for  $m = 0$ :  $m \in \mathbb{Z}^+$  ( $\mathbb{Z}^+$  denotes the set of positive integers) is justified as follows:

$$\begin{aligned} & \frac{1}{\Gamma(\kappa)} \int_0^{\Delta_\theta} u^{\kappa-1} e^{-u} P(m\kappa, (\Delta_\theta - u)) du \\ &= \frac{1}{\Gamma(\kappa)} \int_0^{\Delta_\theta} u^{\kappa-1} e^{-u} \left[ e^{-(\Delta_\theta - u)} \sum_{i=0}^{\infty} \frac{(\Delta_\theta - u)^{m\kappa+i}}{\Gamma(m\kappa + i + 1)} \right] du \\ &= \frac{1}{\Gamma(\kappa)} e^{-\Delta_\theta} \sum_{i=0}^{\infty} \frac{1}{\Gamma(m\kappa + i + 1)} \int_0^{\Delta_\theta} u^{\kappa-1} (\Delta_\theta - u)^{m\kappa+i} du \\ &= e^{-\Delta_\theta} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{(m+1)\kappa+i}}{\Gamma((m+1)\kappa + i + 1)} = P((m+1)\kappa, \Delta_\theta), \end{aligned}$$

where the following relations of  $P(\alpha, x)$  and the beta function have been used:

$$\begin{aligned} P(\alpha, x) &= e^{-x} \sum_{i=0}^{\infty} \frac{x^{\alpha+i}}{\Gamma(\alpha + i + 1)}, \quad \alpha, x > 0, \\ B(\beta, \gamma) &= \int_0^1 t^{\beta-1} (1-t)^{\gamma-1} dt = \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta + \gamma)}, \quad \beta, \gamma > 0. \end{aligned} \quad (\text{A.3})$$

Note that equation A.3 follows from the following equation:

$$\begin{aligned} P(\alpha, x) &= P(\alpha + n, x) + e^{-x} \sum_{i=0}^{n-1} \frac{x^{\alpha+i}}{\Gamma(\alpha + i + 1)}, \\ \lim_{n \rightarrow \infty} P(\alpha + n, x) &= 0, \quad \alpha, x > 0. \end{aligned} \quad (\text{A.4})$$

Since equation A.2 holds for  $m \in \mathbb{Z}^*$ , equation A.1 becomes

$$p(m+1 | \theta; \kappa, \Delta) = P((m+1)\kappa, \Delta_\theta) - P((m+2)\kappa, \Delta_\theta).$$

Equation 2.8 holds for every  $r \in \mathbb{Z}^*$ .



**Appendix B: Capacity-Achieving Distribution for Temporal Coding**

---

**B.1 Proof of Lemma 2:  $I_T(F)$  is Weak\* Continuous.**  $I_T(F)$  is weak\* continuous if the following relation holds,

$$F_n \xrightarrow{w^*} F \implies I_T(F_n) \rightarrow I_T(F), \tag{B.1}$$

since  $I_T(F) = h_T(F; \kappa) - \kappa h_{T|\theta}(F; \kappa) - \kappa$ ; more precisely,

$$F_n \xrightarrow{w^*} F \implies h_T(F_n; \kappa) \rightarrow h_T(F; \kappa) \quad \text{and} \quad h_{T|\theta}(F_n; \kappa) \rightarrow h_{T|\theta}(F; \kappa).$$

$h_{T|\theta}(F_n; \kappa) \rightarrow h_{T|\theta}(F; \kappa)$  holds since  $h_{T|\theta}(F_n; \kappa) = \int_a^b \log \theta \, dF_n(\theta)$  and  $\log \theta$  is a bounded continuous function for  $\theta \in \Theta$ .

Next we show the following equalities:

$$\begin{aligned} \lim_n h_T(F_n; \kappa) &= - \lim_n \int_0^\infty p(t; F_n, \kappa) \log g(t; F_n, \kappa) \, dt \\ &= - \int_0^\infty \lim_n p(t; F_n, \kappa) \log g(t; F_n, \kappa) \, dt \end{aligned} \tag{B.2}$$

$$= - \int_0^\infty p(t; F, \kappa) \log g(t; F, \kappa) \, dt = h_T(F; \kappa). \tag{B.3}$$

The interchange of integral and limit in equation B.2 is justified as follows. From equations 2.5 and 2.7,  $p(t; F, \kappa)$  and  $g(t; F, \kappa)$  are bounded as follows:

$$\begin{aligned} \frac{t^{\kappa-1} \exp[-t/a]}{b^\kappa \Gamma(\kappa)} < p(t; F_n, \kappa) < \frac{t^{\kappa-1} \exp[-t/b]}{a^\kappa \Gamma(\kappa)} \\ -\frac{t}{a} - \kappa \log b < \log g(t; F_n, \kappa) < -\frac{t}{b} - \kappa \log a. \end{aligned} \tag{B.4}$$

From these bounds,  $p(t; F_n, \kappa) \log g(t; F_n, \kappa)$  is bounded for all  $F_n$  with finite  $A_1$  and  $A_2$  as follows:

$$|p(t; F_n, \kappa) \log g(t; F_n, \kappa)| < A_1 t^{\kappa-1} e^{-t/b} + A_2 t^\kappa e^{-t/b}. \tag{B.5}$$

The right-hand side of equation B.5 is integrable as

$$\int_0^\infty [A_1 t^{\kappa-1} e^{-t/b} + A_2 t^\kappa e^{-t/b}] \, dt = \Gamma(\kappa) b^\kappa (A_1 + \kappa b A_2).$$

Since equation B.5 is bounded from above with an integrable function, equation B.2 is justified by the Lebesgue-dominated convergence theorem. Since  $p(t | \theta; \kappa)$  and  $\exp[-t/\theta]/\theta^\kappa$  are continuous bounded functions of  $\theta \in \Theta$ ,  $p(t; F, \kappa)$  and  $g(t; F, \kappa)$  are a continuous function on  $F$ ,  $p(t; F_n, \kappa) \log g(t; F_n, \kappa)$  is also continuous for every  $F_n \in \mathcal{F}$ . These arguments justify equation B.3, and equation B.1 is justified.

**B.2 Proof of Lemma 3.** Let us define  $F_\eta$  and rewrite  $i_T(\theta; F)$  in equation 2.6 as follows:

$$F_\eta = (1 - \eta)F_0 + \eta F,$$

$$i_T(\theta; F) = -\kappa \log \theta - \int_0^\infty p(t | \theta; \kappa) \log g(t; F, \kappa) dt.$$

Then

$$I_T(F_\eta) - I_T(F_0) = \int_a^b i_T(\theta; F_\eta) dF_\eta - \int_a^b i_T(\theta; F_0) dF_0$$

$$= \eta \left[ \int_a^b i_T(\theta; F_\eta) dF - \int_a^b i_T(\theta; F_\eta) dF_0 \right] \quad (\text{B.6})$$

$$+ \int_a^b [i_T(\theta; F_\eta) - i_T(\theta; F_0)] dF_0. \quad (\text{B.7})$$

The weak derivative of  $I_T(F)$  at  $F_0$  is defined as  $I'_{T,F_0}(F) = \lim_{\eta \downarrow 0} (I_T(F_\eta) - I_T(F_0))/\eta$ . By dividing the term in equation B.6 with  $\eta$  and by taking  $\eta \downarrow 0$ , it becomes

$$\int_a^b i_T(\theta; F_0) dF - \int_a^b i_T(\theta; F_0) dF_0 = \int_a^b i_T(\theta; F_0) dF - I_T(F_0) - \kappa.$$

By noting  $g(t; F_\eta, \kappa) = (1 - \eta)g(t; F_0, \kappa) + \eta g(t; F, \kappa)$ , the term in equation B.7 becomes 0. Thus, the weak derivative becomes

$$I'_{T,F_0}(F) = \int_a^b i_T(\theta; F_0) dF - I_T(F_0) - \kappa,$$

which does exist and  $I_T(F)$  is weakly differentiable.

## Appendix C: Capacity-Achieving Distribution for Rate Coding

---

**C.1 Proof of Lemma 4.** First, the following proposition is shown:

**Proposition 3.** *The expectation of  $R$  with respect to  $p(r | \theta; \kappa, \Delta)$  is finite.*

**Proof.** The expectation of  $R$  is

$$\bar{R}_{\kappa, \Delta_\theta} = \sum_{r=0}^{\infty} r p(r \mid \theta; \kappa, \Delta) = \sum_{r=1}^{\infty} P(r\kappa, \Delta_\theta) = e^{-\Delta_\theta} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa + i + 1)}.$$

Since  $P(\alpha, x)$  is a strictly decreasing function of  $\alpha$  for  $\alpha > 0, x > 0$ , if  $\kappa \geq 1$

$$P(r\kappa, \Delta_\theta) \leq P(r \lfloor \kappa \rfloor, \Delta_\theta), \quad r \in \mathbb{Z}^+.$$

Thus, the upper bound is given as

$$\begin{aligned} \bar{R}_{\kappa, \Delta_\theta} &\leq e^{-\Delta_\theta} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{r \lfloor \kappa \rfloor + i}}{\Gamma(r \lfloor \kappa \rfloor + i + 1)} \leq e^{-\Delta_\theta} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{r+i}}{\Gamma(r + i + 1)} \\ &= \bar{R}_{1, \Delta_\theta} = \Delta_\theta, \end{aligned}$$

where  $R_{1, \Delta_\theta} = \Delta_\theta$  holds from the fact that  $p(r \mid \theta; 1, \Delta_\theta)$  is a Poisson distribution. For  $\kappa < 1$ ,  $P(r\kappa, \Delta_\theta) \leq P(\lfloor r\kappa \rfloor, \Delta_\theta)$  holds, and  $\bar{R}_{\kappa, \Delta_\theta}$  is bounded as follows:

$$\begin{aligned} \bar{R}_{\kappa, \Delta_\theta} &= e^{-\Delta_\theta} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa + i + 1)} < e^{-\Delta_\theta} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{\lfloor r\kappa \rfloor + i}}{\Gamma(\lfloor r\kappa \rfloor + i + 1)} \\ &\leq \left\lceil \frac{1}{\kappa} \right\rceil e^{-\Delta_\theta} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{\Delta_\theta^{r+i}}{\Gamma(r + i + 1)} + \left\lfloor \frac{1}{\kappa} \right\rfloor \\ &\leq \left\lceil \frac{1}{\kappa} \right\rceil (\Delta_\theta + 1). \end{aligned}$$

$I_R(F)$  is weak\* continuous if the following relation holds:

$$F_n \xrightarrow{w^*} F \implies I_R(F_n) \rightarrow I_R(F). \tag{C.1}$$

From the definitions of  $I_R(F)$  and  $i_R(\theta, F)$  in equation 2.9,

$$I_R(F) = \int_a^b i_R(\theta, F) dF(\theta).$$

Since  $i_R(\theta, F)$  is a positive continuous function of  $\theta$ , if it is bounded from above, this is justified from the Helly-Bray theorem. It will be shown separately for  $\kappa \geq 1$  and  $\kappa < 1$ .

**For  $\kappa \geq 1$ :** Since  $P(\alpha, \Delta_\theta)$  is a decreasing function of  $\alpha$ , the following inequality holds from equation A.4.

$$\begin{aligned} P(r\kappa, \Delta_\theta) - P(r\kappa + \lfloor \kappa \rfloor, \Delta_\theta) &\leq p(r \mid \theta; \kappa, \Delta) \\ &\leq P(r\kappa, \Delta_\theta) - P(r\kappa + \lceil \kappa \rceil, \Delta_\theta) \\ e^{-\Delta_\theta} \sum_{i=0}^{\lfloor \kappa \rfloor - 1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa + i + 1)} &\leq p(r \mid \theta; \kappa, \Delta) \\ &\leq e^{-\Delta_\theta} \sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa + i + 1)}. \end{aligned} \quad (\text{C.2})$$

With the above equation,  $p(r; F, \kappa, \Delta)$  is bounded from below as follows:

$$p(r; F, \kappa, \Delta) = \int_a^b p(r \mid \theta; \kappa, \Delta) dF(\theta) > e^{-\Delta_M} \sum_{i=0}^{\lfloor \kappa \rfloor - 1} \frac{\Delta_m^{r\kappa+i}}{\Gamma(r\kappa + i + 1)},$$

where  $\Delta_m = \Delta/b$  and  $\Delta_M = \Delta/a$  are the minimum and the maximum of  $\Delta_\theta$ , respectively. Thus,

$$\frac{p(r \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} < e^{\Delta_M - \Delta_\theta} \frac{\sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lfloor \kappa \rfloor - 1} \frac{\Delta_m^{r\kappa+i}}{\Gamma(r\kappa+i+1)}} < B e^{\Delta_M - \Delta_\theta} \left( \frac{\Delta_\theta}{\Delta_m} \right)^{r\kappa},$$

where  $B$  is the following upper bound:

$$\frac{\sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_\theta^i}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lfloor \kappa \rfloor - 1} \frac{\Delta_m^i}{\Gamma(r\kappa+i+1)}} = \frac{1 + \sum_{i=1}^{\lceil \kappa \rceil - 1} \Delta_\theta^i \frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}}{1 + \sum_{i=1}^{\lfloor \kappa \rfloor - 1} \Delta_m^i \frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}} < 1 + \sum_{i=1}^{\lceil \kappa \rceil - 1} \Delta_m^i = B.$$

With the result of proposition 3,  $i_R(\theta, F)$  is bounded from above:

$$\begin{aligned} i_R(\theta, F) &= \sum_{r=0}^{\infty} p(r \mid \theta; \kappa, \Delta) \log \frac{p(r \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} \\ &< \kappa \Delta_\theta \log \frac{\Delta_\theta}{\Delta_m} - \Delta_\theta + \Delta_M + \log B. \end{aligned}$$

**For  $\kappa < 1$ :** When  $\kappa < 1$ , the following relation holds from equation A.4:

$$p(r \mid \theta; \kappa, \Delta) = P(r\kappa, \Delta_\theta) - P((r+1)\kappa + 1, \Delta_\theta) - \frac{e^{-\Delta_\theta} \Delta_\theta^{(r+1)\kappa}}{\Gamma((r+1)\kappa + 1)}.$$

Since  $P(\alpha, x)$  is a decreasing function, the following relation holds:

$$\begin{aligned}
 P(r\kappa, \Delta_\theta) - P(r\kappa + 1, \Delta_\theta) &= \frac{e^{-\Delta_\theta} \Delta_\theta^{(r+1)\kappa}}{\Gamma((r+1)\kappa + 1)} \\
 &< p(r \mid \theta; \kappa, \Delta) < P(r\kappa, \Delta_\theta) - P(r\kappa + 1, \Delta_\theta) \\
 \frac{e^{-\Delta_\theta} \Delta_\theta^{r\kappa}}{\Gamma(r\kappa + 1)} \left( 1 - \Delta_\theta^\kappa \frac{\Gamma(r\kappa + 1)}{\Gamma(r\kappa + \kappa + 1)} \right) &< p(r \mid \theta; \kappa, \Delta) \\
 &< \frac{e^{-\Delta_\theta} \Delta_\theta^{r\kappa}}{\Gamma(r\kappa + 1)}. \tag{C.3}
 \end{aligned}$$

The above equation gives the following bound of  $p(r; F, \kappa, \Delta)$ :

$$\begin{aligned}
 p(r; F, \kappa, \Delta) &= \int_a^b p(r \mid \theta; \kappa, \Delta) dF(\theta) \\
 &> \frac{e^{-\Delta_M} \Delta_M^{r\kappa}}{\Gamma(r\kappa + 1)} \left( 1 - \Delta_M^\kappa \frac{\Gamma(r\kappa + 1)}{\Gamma(r\kappa + \kappa + 1)} \right).
 \end{aligned}$$

From the property of the gamma function,  $\Gamma(r\kappa + 1)/\Gamma(r\kappa + \kappa + 1)$  decreases as  $r$  increases for  $r > 1/\kappa$ , and there exists a finite positive integer  $r_0 \geq 1/\kappa$  such that, for all  $r \geq r_0$ , the following inequality holds for a positive real number  $C_1$ :

$$1 - \Delta_M^\kappa \frac{\Gamma(r\kappa + 1)}{\Gamma(r\kappa + \kappa + 1)} > C_1.$$

Thus,

$$\frac{p(r \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} < e^{\Delta_M - \Delta_\theta} \left( \frac{\Delta_\theta}{\Delta_m} \right)^{r\kappa} \frac{1}{1 - \Delta_M^\kappa \frac{\Gamma(r\kappa + 1)}{\Gamma(r\kappa + \kappa + 1)}} < \frac{e^{\Delta_M - \Delta_\theta}}{C_1} \left( \frac{\Delta_\theta}{\Delta_m} \right)^{r\kappa}.$$

With the result of proposition 3,

$$\begin{aligned}
 S_1 &= \sum_{r=r_0}^\infty p(r \mid \theta; \kappa, \Delta) \log \frac{p(r \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} \\
 &< \kappa \left\lceil \frac{1}{\kappa} \right\rceil (\Delta_\theta + 1) \log \frac{\Delta_\theta}{\Delta_m} - \Delta_\theta + \Delta_M - \log C_1,
 \end{aligned}$$

where  $S_1$  is finite. It can be shown that there exists a real number  $C_2 > 0$ , such that  $p(r \mid \theta; \kappa, \Delta) > C_2$  for all  $\theta \in \Theta, r \in \{0, \dots, r_0 - 1\}$  and the following

sum is finite:

$$S_2 = \sum_{r=0}^{r_0-1} p(r \mid \theta; \kappa, \Delta) \log \frac{p(r \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)}.$$

Thus  $i_R(\theta, F) = S_1 + S_2$  is bounded from above.

**C.2 Proof of Theorem 3.** First, the following proposition is shown:

**Proposition 4.** *As  $x \rightarrow \infty$  ( $x \in \mathbb{R}^+$ ), the following equation holds:*

$$\lim_{x \rightarrow \infty} \frac{\sum_{r=1}^{\infty} P(rm, x)}{x} = \frac{1}{m}, \quad m \in \mathbb{Z}^+. \tag{C.4}$$

**Proof of proposition 4.** From proposition 3 in section C.1,  $\sum_{r=1}^{\infty} P(rm, x)$  is bounded from above with a linear function of  $x$ . Let us define the sum as  $S_m(x)$ . From equation A.3,

$$S_m(x) = \sum_{r=1}^{\infty} P(rm, x) = e^{-x} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{x^{rm+i}}{\Gamma(rm+i+1)}.$$

It is easily checked that

$$\left(\frac{d}{dx} + 1\right)^k S_m(x) = e^{-x} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{x^{rm-k+i}}{\Gamma(rm-k+i+1)}, \quad k \in \{0, \dots, m-1\}.$$

Thus, the following linear differential equation is derived:

$$\sum_{k=0}^{m-1} \left(\frac{d}{dx} + 1\right)^k S_m(x) = e^{-x} \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \frac{x^{r+i}}{\Gamma(r+i+1)} = x.$$

When the differential equation is solved, the general solution gives the following form of  $S_m(x)$ :

$$S_m(x) = \frac{x}{m} + \sum_{k=1}^{m-1} C_k e^{(-1+\alpha_k)x} - \frac{m-1}{m^2}, \quad \alpha_k = \exp\left[\frac{2\pi k\sqrt{-1}}{m}\right]. \tag{C.5}$$

Since  $|\operatorname{Re} \alpha_k| < 1$ ,  $\operatorname{Re}(-1 + \alpha_k) < 0$  holds for  $k \in \{1, \dots, m-1\}$ , and  $\lim_{x \rightarrow \infty} S_m(x)/x = 1/m$ .

**Corollary 3.** *As  $\theta \downarrow 0$ , the expectation of  $R$  with respect to  $p(r \mid \theta; \kappa, \Delta)$  grows proportional to  $\Delta_\theta = \Delta/\theta$ .*

**Proof of corollary 3.** The expectation  $\bar{R}_{\kappa, \Delta_\theta} = \sum_{r=0}^\infty r p(r | \theta; \kappa, \Delta)$  is bounded as follows:

$$\sum_{r=1}^\infty P(r \lceil \kappa \rceil, \Delta_\theta) \leq \bar{R}_{\kappa, \Delta_\theta} = \sum_{r=1}^\infty P(r \kappa, \Delta_\theta) \leq \sum_{r=1}^\infty P(r \lfloor \kappa \rfloor, \Delta_\theta).$$

From proposition 4,  $\sum_{r=1}^\infty P(r \lceil \kappa \rceil, \Delta_\theta)$  and  $\sum_{r=1}^\infty P(r \lfloor \kappa \rfloor, \Delta_\theta)$  grows proportional to  $\Delta_\theta$ , which proves the corollary.

Let us prove theorem 3.

**For  $\kappa \geq 1$ :** From equation C.2,  $p(r; F, \kappa, \Delta)$  is bounded from above as follows:

$$p(r; F, \kappa, \Delta) = \int_a^b p(r | \theta; \kappa, \Delta) dF(\theta) < e^{-\Delta_m} \sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_M^{r\kappa+i}}{\Gamma(r\kappa+i+1)}$$

and

$$\frac{p(r | \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} > e^{\Delta_m - \Delta_\theta} \frac{\sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_M^{r\kappa+i}}{\Gamma(r\kappa+i+1)}} > D e^{\Delta_m - \Delta_\theta} \left( \frac{\Delta_\theta}{\Delta_M} \right)^{r\kappa},$$

where  $D$  is the following lower bound:

$$\frac{\sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_\theta^i}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lceil \kappa \rceil - 1} \frac{\Delta_M^i}{\Gamma(r\kappa+i+1)}} = \frac{\sum_{i=0}^{\lceil \kappa \rceil - 1} \Delta_\theta^i \frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lceil \kappa \rceil - 1} \Delta_M^i \frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}} > \frac{1}{1 + \sum_{i=1}^{\lceil \kappa \rceil - 1} \Delta_M^i} = D.$$

This shows  $i_R(\theta, F)$  is bounded from below as

$$i_R(\theta, F) > \kappa \bar{R}_{\kappa, \Delta_\theta} \log \frac{\Delta_\theta}{\Delta_M} - \Delta_\theta + \Delta_m + \log D.$$

Since  $\bar{R}_{\kappa, \Delta_\theta}$  grows with  $\Delta_\theta$  as  $\theta \downarrow 0$ , the lower bound of  $i_R(\theta, F)$  grows with  $\Delta_\theta \log \Delta_\theta$ . Thus,  $i_R(\theta, F)$  cannot be finite and constant for  $\forall \theta \in \mathbb{R}^+$ , which brings the contradiction.

**For  $\kappa < 1$ :** From equation C.3,  $p(r; F, \kappa, \Delta)$  is bounded from above as follows:

$$p(r; F, \kappa, \Delta) < \frac{e^{-\Delta_m} \Delta_M^{r\kappa}}{\Gamma(r\kappa+1)} \tag{C.6}$$

Let us denote  $r$  as

$$r = r'K + b, \quad \text{where } K = \left\lceil \frac{1}{\kappa} \right\rceil, \quad r' \in \mathbb{Z}^*, \quad b \in \{0, \dots, K-1\}.$$

$r'$  and  $b$  can be considered as stochastic variables, and the following relation holds:

$$\begin{aligned} p(r \mid \theta; \kappa, \Delta) &= p(r'K + b \mid \theta; \kappa, \Delta) = q(r' \mid \theta; \kappa, \Delta)q(b \mid r', \theta; \kappa, \Delta), \\ q(r' \mid \theta; \kappa, \Delta) &= \sum_{b=0}^{K-1} p(r'K + b \mid \theta; \kappa, \Delta), \\ q(b \mid r', \theta; \kappa, \Delta) &= \frac{p(r'K + b \mid \theta; \kappa, \Delta)}{q(r' \mid \theta; \kappa, \Delta)}. \end{aligned}$$

Let  $H_R$  and  $H_{R'}$  be the entropy of  $R$  and  $R'$ , respectively and  $H_{B|R'}$  be the conditional entropy of  $B$  given  $R'$ . The following relation holds:

$$H_R = H_{R'} + H_{B|R'} \leq H_{R'} + \log K,$$

which is justified from  $0 \leq H_{B|R'} \leq \log K$ . With this result,

$$\begin{aligned} &\sum_{r=0}^{\infty} p(r \mid \theta; \kappa, \Delta) \log p(r \mid \theta; \kappa, \Delta) \\ &\geq \sum_{r'=0}^{\infty} q(r' \mid \theta; \kappa, \Delta) \log q(r' \mid \theta; \kappa, \Delta) - \log K. \end{aligned}$$

Since  $\lfloor \kappa K \rfloor = 1$  holds, the probability  $q(r' \mid \theta; \kappa, \Delta)$  is bounded as follows:

$$q(r' \mid \theta; \kappa, \Delta) = P(r'K\kappa, \Delta_\theta) - P((r'+1)K\kappa, \Delta_\theta) \geq e^{-\Delta_\theta} \frac{\Delta_\theta^{r'K\kappa}}{\Gamma(r'K\kappa + 1)}. \quad (\text{C.7})$$

With equations C.6 and C.7,

$$\begin{aligned} \frac{q(r' \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} &= \frac{q(r' \mid \theta; \kappa, \Delta)}{p(r'K + b; F, \kappa, \Delta)} \\ &> e^{\Delta_m - \Delta_\theta} \left( \frac{\Delta_\theta}{\Delta_M} \right)^{r'K\kappa} \Delta_M^{-b\kappa} \frac{\Gamma((r'K + b)\kappa + 1)}{\Gamma(r'K\kappa + 1)} \\ &\geq E e^{\Delta_m - \Delta_\theta} \left( \frac{\Delta_\theta}{\Delta_M} \right)^{r'K\kappa}, \end{aligned}$$



where  $E$  is the following lower bound:

$$\begin{aligned} & \Delta_M^{-b\kappa} \frac{\Gamma((r'K + b)\kappa + 1)}{\Gamma(r'K\kappa + 1)} \\ & \geq \min \{1, \Delta_M^{-\kappa} \Gamma(\kappa + 1), \dots, \Delta_M^{-(K-1)\kappa} \Gamma((K - 1)\kappa + 1)\} = E. \end{aligned}$$

This shows  $i_R(\theta, F)$  is bounded from below as

$$\begin{aligned} i_R(\theta, F) &= \sum_{r=0}^{\infty} p(r \mid \theta; \kappa, \Delta) \log \frac{p(r \mid \theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)} \\ &\geq \sum_{r'=0}^{\infty} \sum_{b=0}^{K-1} p(r'K + b \mid \theta; \kappa, \Delta) \log \frac{q(r' \mid \theta; \kappa, \Delta)}{p(r'K + b; F, \kappa, \Delta)} - \log K \\ &> \sum_{r'=0}^{\infty} q(r' \mid \theta; \kappa, \Delta) \log \left( E e^{\Delta_m - \Delta_\theta} \left( \frac{\Delta_\theta}{\Delta_M} \right)^{r'K\kappa} \right) - \log K \\ &= \left[ \sum_{r'=0}^{\infty} r' q(r' \mid \theta; \kappa, \Delta) \right] K \kappa \log \left( \frac{\Delta_\theta}{\Delta_M} \right) \\ &\quad - \Delta_\theta + \Delta_m + \log E - \log K. \end{aligned}$$

Since  $\sum_{r'=0}^{\infty} r' q(r' \mid \theta; \kappa, \Delta)$  is equivalent to  $\bar{R}_{K\kappa, \Delta_\theta}$ , proposition 4 shows that it grows proportional to  $\Delta_\theta$  as  $\theta \downarrow 0$ . Thus,  $i_R(\theta, F)$  is lower bounded with a term that grows with  $\Delta_\theta \log \Delta_\theta$  and  $i_R(\theta, F)$  cannot be finite and constant for  $\forall \theta \in \mathbb{R}^+$ , which brings the contradiction.

**Acknowledgments**

---

We are grateful for helpful discussions with Mark D. McDonnell. We also thank the anonymous reviewers for valuable feedback. This work was supported by Grant-in-Aid for Scientific Research No. 18079013, MEXT, Japan.

**References**

---

Abou-Faycal, I. C., Trott, M. D., & Shamai (Shitz), S. (2001). The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Transactions on Information Theory*, 47, 1290–1301.

Akhiezer, N. (1965). *The classical moment problem*. (N. Kemmer, Trans.). Edinburgh: Oliver & Boyd.

Baker, S. N., & Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J. Neurophysiol.*, 84, 1770–1780.

- Bar-David, I. (1969). Communication under the Poisson regime. *IEEE Transactions on Information Theory*, *15*, 31–37.
- Bickel, P. J., Klaassen, C., Ritov, Y., & Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore, MD: Johns Hopkins University Press.
- Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature Neuroscience*, *2*, 947–957.
- Doob, J. (1994). *Measure theory*. Berlin: Springer-Verlag.
- Guo, D., Shamai (Shitz), S., & Verdú, S. (2008). Mutual information and conditional mean estimation in Poisson channels. *IEEE Transactions on Information Theory*, *54*, 1837–1849.
- Gursoy, M. C., Poor, H. V., & Verdú, S. (2002). *The capacity of the noncoherent Rician fading channel* (Tech. Rep.). Princeton, NJ: Princeton University.
- Gursoy, M. C., Poor, V., & Verdú, S. (2005). The noncoherent Rician fading channel—Part I: Structure of the capacity-achieving input. *IEEE Transactions on Wireless Communications*, *4*, 2193–2206.
- Ikeda, K. (2005). Information geometry of interspike intervals in spiking neurons. *Neural Computation*, *17*, 2719–2735.
- Luenberger, D. G. (1969). *Optimization by vector space method*. Hoboken, NJ: Wiley.
- MacKay, D. M., & McCulloch, W. S. (1952). The limiting information capacity of a neuronal link. *Bull. Math. Biophys.*, *14*, 127–135.
- Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*, 1503–1506.
- McDonnell, M. D., & Stocks, N. G. (2008). Maximally informative stimuli and tuning curves for sigmoidal rate-coding neurons and populations. *Physical Review Letters*, *101*, 058103.
- Miura, K., Okada, M., & Amari, S. (2006). Estimating spiking irregularities under changing environments. *Neural Computation*, *18*, 2359–2386.
- Nikitin, A. P., Stocks, N. G., Morse, R. P., & McDonnell, M. D. (2008). *Neural population coding is optimized by discrete tuning curves*. Manuscript submitted for publication.
- Pawlas, Z., Klevanov, L. B., & Prokop, M. (2008). Parameters of spike trains observed in a short time window. *Neural Computation*, *20*, 1325–1343.
- Rapoport, A., & Horvath, W. J. (1960). The theoretical channel capacity of a single neuron as determined by various coding systems. *Information and Control*, *3*, 335–350.
- Rullen, R. V., & Thorpe, S. J. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, *13*, 1255–1283.
- Shamai (Shitz), S. (1990). Capacity of a pulse amplitude modulated direct detection photon channel. *IEE Proceedings*, *137*, 424–430.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.
- Shinomoto, S., Miyazaki, Y., Tamura, H., & Fujita, I. (2005). Regional and laminar difference in in vivo firing patterns of primate cortical neurons. *J. Neurophysiol*, *94*, 567–575.
- Shinomoto, S., Shima, K., & Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Computation*, *15*, 2823–2842.
- Smith, J. G. (1971). The information capacity of amplitude- and variance-constrained scalar gaussian channels. *Information and Control*, *18*, 203–219.

- Stein, R. B. (1967). The information capacity of nerve cells using a frequency code. *Biophysical Journal*, 797–826.
- Tchamkerten, A. (2004). On the discreteness of capacity-achieving distributions. *IEEE Transactions on Information Theory*, 50, 2773–2778.
- Verdú, S. (1999). *Poisson communication theory*. Invited talk in the International Technion Communication Day in honor of Israel Bar-David. Available online at <http://www.princeton.edu/~verdu>.

---

Received May 27, 2008; accepted September 17, 2008.