Research Memorandum No. 1076 10/24/2008

Capacity of a Single Neuron Channel

Ikeda Shiro (The Institute of Statistical Mathematics)
Manton Jonathan (The University of Melbourne)

The Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku,
Tokyo, 106-8569, Japan

Capacity of a Single Neuron Channel

Shiro Ikeda shiro@ism.ac.jp
The Institute of Statistical Mathematics, Tokyo 106-8569, Japan

Jonathan H. Manton j.manton@ieee.org
The University of Melbourne, Victoria 3010, Australia

Abstract

The information transfer through a single neuron is a fundamental information processing in the brain and computing the information channel capacity is important to understand the information processing in the brain. The problem is difficult since the capacity depends on various issues, such as coding, characteristics of the communication channel and optimisation over input distributions. In this letter, two different models are considered. The temporal coding model of a neuron as a communication channel assumes the output is τ where τ is a gamma-distributed random variable corresponding to the inter-spike 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 the 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 when a communication channel is corrupted with noise, the rate at which 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 paper, it is referred to 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 also biologically (Borst & Theunissen, 1999). The computation of the capacity is difficult since it depends on multiple factors, that is, 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. The temporal coding uses the inter-spike intervals (ISIs) to code information while the rate coding uses the number of spikes in a fixed interval. In this letter, both of them are considered.

The channel model is deeply related to the noise properties of ISIs. It has been reported in Baker and Lemon (2000) that the statistical properties of ISIs recorded from primary motor cortex and supplementary motor area (SMA) of monkeys are similar to the gamma distribution. In Shinomoto et al. (2003) and Shinomoto et al. (2005), the spike trains from multiple areas [1] are studied and a statistical index which describes the randomness of ISIs is proposed. The index is deeply related to the gamma distribution (Shinomoto et al., 2003; Ikeda, 2005). In this letter, ISIs are modelled with

---

[1] Data from pre-SMA, SMA, rostral cingulate motor area (CMAr), 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. (2003).

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 which 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 et al., 2001; Gursoy et al., 2002; Gursoy et al., 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 only have to consider 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 per sec. This is 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 the paper with some discussions. Most of the mathematical proofs are summarised in 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 which are the shape parameter κ and the scale parameter θ. From some studies, κ of individual neuron appears to be constant (the value of κ may depends on the type of neuron), while θ changes dynamically over time.

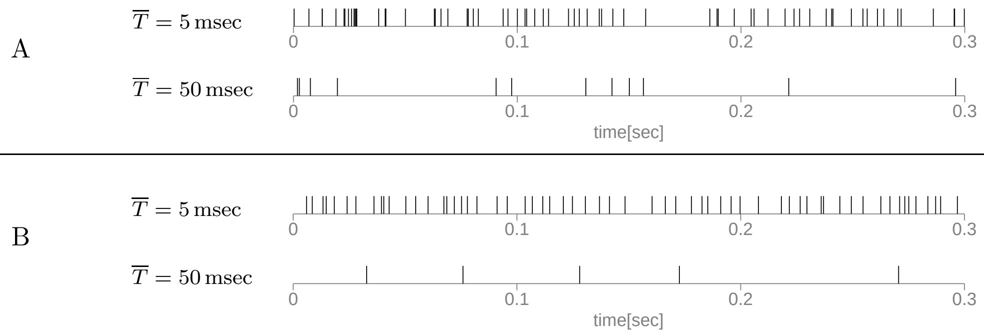


Figure 1: Simulated spike trains. ISIs follow a gamma distribution, where the shape parameter κ is 0.75 for A and 4.5 for B. The expected values of ISI is 5 msec in the upper trains and 50 msec in the lower trains for both A and B.

Figure 1 shows simulated spike trains with two different shape parameter κ's. It is 0.75 in Fig. 1.A and 4.5 in Fig. 1.B. When κ is small, spike trains become more irregular. In Ikeda (2005) and Miura et al. (2006), they studied the estimation methods of κ from spike trains. Estimation of κ is regarded as the semiparametric statistical estimation (Bickel et al., 1993).

In this letter, we focus not on the estimation but on the information processing of a single neuron. Each neuron is a communication channel since it is widely believed to be transmitting information with spikes. When a channel is noisy, only a limited amount of information can be transmitted through a channel, and this limit is the capacity (Shannon, 1948). Based on the gamma distribution model, the capacity of a neuron is investigated in the following.

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 can be defined as a stochastic model described as $p(y|x)$ and the mutual information is defined as

$$\begin{aligned} I(X;Y) &= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} p(y|x)\log\frac{p(y|x)}{p(y)}d\mu(y)dF(x), \\ \text{where} \quad p(y) &= \int_{x\in\mathcal{X}} p(y|x)dF(x). \end{aligned} \tag{1}$$

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

$$C = \sup_{F\in\mathcal{F}} I(F). \tag{2}$$

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 come back to a neuron model. First, we have to define what is 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 a ISI, which is 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 the shape parameter κ of each neuron is fixed and known. In Shinomoto et al. (2003), an statistical index L_V (local index), is defined as an index to characterise each neuron. For a $T \sim \Gamma(\kappa, \theta)$, $\overline{L_V} = 3/(2\kappa + 1)$ holds. From their investigation with biological data, it seems most of the cells' L_V are lying in an interval (0.3 to 1.2) and κ is thus assumed to be in an interval $\kappa \in [\kappa_m, \kappa_M]$ (κ_m and κ_M are set to 0.75 and 4.5, respectively in §4).

Under the assumption, the scale parameter θ is the only variable parameter which plays the role of input, that is, X in §.2.2. The density function of t is

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

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

$$p(t|\theta;\kappa) = \exp\Big[-\frac{1}{\theta}t + (\kappa-1)\log t - \log\Gamma(\kappa) - \kappa\log\theta\Big]. \tag{3}$$

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 entropy becomes

$$H(T\,|\,\theta\,;\kappa) = -\int_0^\infty p(t|\theta;\kappa)\log p(t|\theta;\kappa)dt = \kappa - (\kappa-1)\psi(\kappa) + \log\Gamma(\kappa) + \log\theta.$$

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

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

Thus, θ 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 Θ respectively, as far as no confusion occurs. Let us define $F(\theta)$ as the cumulative distribution function of θ and $\mathcal{F}$ as the set of all possible $F(\theta)$, that is

$$\mathcal{F} = \big\{F : \mathbb{R} \to [0,1] \mid F(\theta) = 0, (\forall\theta < a),\ F(\theta) = 1, (\forall\theta > b)\big\}. \tag{4}$$

Note that $F \in \mathcal{F}$ is right-continuous and non-decreasing on Θ and $\mathcal{F}$ includes continuous and discrete distributions[2] of θ.

Next, let us consider what is Y, that is, "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 and McCulloch (1952), for example). This is called "temporal coding" (Fig.3). 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" (Fig.3). In communication theory, "coding" is often used for "source coding," "error-control coding" and "cryptography coding." It seems the term "modulation" is more suitable to the above definition. However, we follow the standard usage of the neuroscience community. How to "encode" (or to modulate) the input θ to the neuron channel depends on which coding is used. For the temporal coding, θ is fixed during the interval t while θ is fixed during Δ for the rate coding, we will discuss this in §5.

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

Temporal coding

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

$$p(t; F, \kappa) = \int_a^b p(t|\theta; \kappa)dF(\theta) \tag{5}$$

where $p(t|\theta;\kappa)$ is defined in eq.(3). The existence of $p(t;F,\kappa)$ follows from the existence of $p(t|\theta;\kappa)$. The mutual information of T and θ is defined as

$$I_T(F) = \int_a^b i_T(\theta; F)dF(\theta), \quad \text{where} \quad i_T(\theta; F) = \int_0^\infty p(t|\theta;\kappa)\log\frac{p(t|\theta;\kappa)}{p(t;F,\kappa)}dt. \tag{6}$$

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

$$g(t;F,\kappa) = \int_a^b \frac{\exp[-t/\theta]}{\theta^\kappa}dF(\theta), \quad p(t;F,\kappa) = \frac{t^{\kappa-1}}{\Gamma(\kappa)}g(t;F,\kappa). \tag{7}$$

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

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

where

$$h_T(F;\kappa) = -\int_0^\infty p(t;F,\kappa)\log g(t;F,\kappa)dt, \quad h_{T|\theta}(F;\kappa) = \int_a^b \log\theta dF(\theta).$$

---

[2] In Stein (1967), the distribution of θ is assumed to be discrete, but we do not assume it.

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

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

The capacity C_T and the distribution which achieves C_T will be studied in the next section.

Rate coding

In rate coding, a time window is set and the spikes in the interval is counted. Let us denote the interval and the rate as Δ 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*

$$p(r|\theta;\kappa,\Delta) = P\big(r\kappa,\, \Delta/\theta\big) - P\big((r+1)\kappa,\, \Delta/\theta\big), \quad r \in \mathbb{Z}^*, \tag{8}$$

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

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

Proof. See §A. The same distribution is discussed in Pawlas et al. (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 θ is defined as

$$I_R(F) = \int_a^b i_R(\theta, F) dF(\theta), \quad \text{where} \quad i_R(\theta,F) = \sum_{r=0}^{\infty} p(r|\theta;\kappa,\Delta) \log \frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)}. \tag{9}$$

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

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

The capacity C_R and the distribution which achieves C_R will be studied in the next section.

3 Theoretical Studies

The cumulative distribution $F \in \mathcal{F}$ is a right-continuous non-decreasing function on a interval Θ. Thus, θ can be a discrete or continuous random variable over Θ. 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). The proofs of the discreteness in this letter 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 subsection. In the following, results of optimisation theory and probability theory will be used. Suppose X is a normed linear space. In optimisation 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* $x \in X$, $x_n^*(x) \to x^*(x)$. *In this case we write* $x_n^*(x) \overset{w^*}{\to} x^*(x)$ *(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 optimisation theory is 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*

$$C = \max_{x^* \in \mathcal{S}} J(x^*) \tag{10}$$

is achieved by a unique x^* *in* $\mathcal{S}$.

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

From the above discussion, $\mathcal{F}$ in eq.(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 eq.*(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 the Helly's compactness theorem (Doob (1994), section X). ☐

The Kuhn-Tucker (K-T) condition on the mutual information is used for the next step of the proof. Before showing the condition, let us 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* $J'_{F_0} : \mathcal{F} \to \mathbb{R}$ *such that*

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

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* $\mathcal{F}$.

And 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* $F \in \mathcal{F}$

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

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 showing these, 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 subsection, the capacity achieving distribution for temporal coding is shown to be a discrete distribution with a finite number of points. Let us start with the following lemma.

Lemma 2. *$I_T(F)$ in eq.(6) is a weak* continuous function on $F \in \mathcal{F}$ and strictly concave in $\mathcal{F}$.*

Proof. §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 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 eq.(6) is weakly differentiable in $\mathcal{F}$. The weak derivative at $F_0 \in \mathcal{F}$ has the form*

$$I'_{T,F_0}(F) = \int_a^b i_T(\theta; F_0)dF - I_T(F_0) - \kappa. \tag{11}$$

Proof. See §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*

$$\begin{aligned} i_T(\theta; F_0) &\leq I_T(F_0) + \kappa, \quad \forall\theta \in \Theta \\ i_T(\theta; F_0) &= I_T(F_0) + \kappa, \quad \forall\theta \in E_0. \end{aligned} \tag{12}$$

Proof. This is proved following the same steps in Smith (1971) (Corollary 1) with eq.(11). □

The main result of this subsection is summarised 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 $\mathrm{Re}\, z > 0$, 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 Θ 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 $\mathrm{Re}\, z > 0$. This region includes positive real line $\mathbb{R}^+$ and

$$-\int_0^\infty p(t|\theta; \kappa) \log g(t; F_0, \kappa)dt = \kappa \log\theta + I_T(F_0) + \kappa, \quad \theta \in \mathbb{R}^+ \tag{13}$$

is implied. The LHS of eq.(13) is bound as follows (see §B.1, eq.(26)).

$$\begin{aligned}\frac{1}{b}\int_0^\infty t\,p(t|\theta;\kappa)dt + \kappa\log a &\le -\int_0^\infty p(t|\theta;\kappa)\log g(t;F_0,\kappa)dt \\ &\le \frac{1}{a}\int_0^\infty t\,p(t|\theta;\kappa)dt + \kappa\log b.\end{aligned} \tag{14}$$

Since the expectation of T w.r.t $p(t|\theta;\kappa)$ is $\kappa\theta$, eq.(14) shows the LHS of eq.(13) grows linearly with θ. Since the RHS increases only with $\log\theta$, eq.(13) 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. In Shamai (Shitz) (1990), it has been 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 eq.(9) is a weak* continuous function on $F \in \mathcal{F}$ and strictly concave in $\mathcal{F}$.*

Proof. §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 §7.2 of Shamai (Shitz) (1990), which is an application of Carleman's theorem (Akhiezer, 1965). ☐

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

Lemma 5. *$I_R(F)$ in eq.(9) is weakly differentiable in $\mathcal{F}$. The weak derivative at $F_0 \in \mathcal{F}$ has the form*

$$I'_{R,F_0}(F) = \int_a^b i_R(\theta;F_0)dF - I_R(F_0). \tag{15}$$

Proof. The proof is identical to the proof of lemma 3 in §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*

$$\begin{aligned} i_R(\theta;F_0) &\le I_R(F_0), \quad \forall\theta\in\Theta \\ i_R(\theta;F_0) &= I_R(F_0), \quad \forall\theta\in E_0.\end{aligned} \tag{16}$$

Proof. This is proved following the same steps in Smith (1971) (Corollary 1) with eq.(15). ☐

Finally, the following theorem proves 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

$$\begin{aligned} i_R(z;F) &= \sum_{r=0}^{\infty} p(r|z;\kappa,\Delta)\log\frac{p(r|z;\kappa,\Delta)}{p(r;F,\kappa,\Delta)}, \\ p(r|z;\kappa,\Delta) &= P(r\kappa,\Delta/z) - P((r+1)\kappa,\Delta/z).\end{aligned}$$

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

If E_0 in corollary 2 has infinite points, since Θ is bounded and closed, E_0 has a limit point and hence, from eq.(16), the identity theorem implies $i_R(z; F_0) = I_R(F_0)$ for the region $\operatorname{Re} z > 0$. This region includes positive real line $\mathbb{R}^+$ and

$$\sum_{r=0}^{\infty} p(r|\theta; \kappa, \Delta) \log \frac{p(r|\theta; \kappa, \Delta)}{p(r; F_0, \kappa, \Delta)} = I_R(F_0), \quad \theta \in \mathbb{R}^+ \tag{17}$$

is implied. The proof (see §C.2) is completed by deriving a contradiction for eq.(17). 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 of related works (Smith, 1971; Shamai (Shitz), 1990; Abou-Faycal et al., 2001; Gursoy et al., 2005). In this section, the capacity and the capacity achieving distribution is computed numerically for temporal and rate coding.

4.1 Common Steps of Numerical Experiments

The computation of the capacity and the capacity achieving distribution of the neuron channel is difficult since the closed form expression of $i_T(\theta; F)$ in eq.(6) and $i_R(\theta; F)$ in eq.(9) are not provided for a general discrete $F(\theta)$: $i_T(\theta; F)$ requires an integration and $i_R(\theta; F)$ requires a summation of an infinite series. 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 other related works use similar methods (Smith, 1971; Abou-Faycal et al., 2001; Gursoy et al., 2005).

1. Initialise the number of the points N as 2.
2. Set the position and probability of each point as θ_j and π_j, ($j = \{1, \cdots, N\}$), respectively, where

$$a(\kappa) \leq \theta_1 < \cdots, < \theta_N \leq b(\kappa), \qquad \sum_{j=1}^{N} \pi_j = 1, \quad \pi_j > 0, \quad j \in \{1, \cdots, N\}.$$

3. Starting from some initial values, maximise 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 (eqs.(12) or (16)) to see if it is the capacity achieving distribution.
5. If K-T condition is satisfied, the capacity and the capacity achieving distribution is obtained, otherwise increase N by 1 and go to step 2.

The range of θ 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 will be discussed in §5.

The capacity and the capacity achieving distribution for temporal and rate coding are computed for multiple values of κ. As described in §2.3, a statistical index L_V has been proposed which characterises spike trains (Shinomoto et al., 2003). Its expectation is related to κ as $\overline{L_V} = 3/(2\kappa+1)$. In the following numerical studies, we vary κ from 0.75 to 4.5 (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 their papers.

4.2 Temporal Coding

Figure 2.A shows the computed capacity for each κ. The capacity C_T (bit[3] per channel use) increases monotonically as κ increases. This is natural since as κ increases ISIs become more regular and more information can be sent. The capacity becomes larger than 1 bit when κ becomes 3.85.

The capacity achieving distributions are shown in Figs. 2.C and 2.D. For each κ, the distribution has only 2 or 3 points. Moreover, two of them are both ends of the range $\Theta(\kappa)$ (a_0/κ and b_0/κ). If κ is smaller than 2.10, there are only 2 points. When it is equal to 2.10, the number of points becomes 3. The position of the third point is very stable for different κ's. The probability of each point is shown in Fig. 2.D. The probabilities of both ends tend to be similar, while the probability of the third point increases gradually as κ 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 distribution is computed, the following C'_T (bit per sec) is defined

$$C'_T = C_T/\overline{\tau}, \quad \text{where} \quad \overline{\tau} = \kappa \sum_{j=1}^{N} \pi_j \theta_j. \tag{18}$$

Note that $\overline{\tau}$ is around 25 msec for all κ in the experiments. The information rate is shown as the function of κ in Fig.2.B. Further discussions will be provided in §5.

4.3 Rate Coding

In rate coding, the time window Δ must be defined. Since the average time for sending a symbol with temporal coding is around 25 msec, Δ is set to 25 msec in the numerical experiment.

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

The capacity achieving distributions are shown in Figs. 4.C and 4.D. For each κ, the distribution has 2 to 4 discrete points and two of them are both ends of the range $\Theta(\kappa)$ (a_0/κ and b_0/κ). If κ is smaller than 1.25, there are only 2 points. Above 1.25, there are 3 points and it becomes 4 when κ becomes 4.0. The probability of each point is shown in Fig. 2.D. The probabilities of both ends tend to be similar, while the probability of the third point increases gradually as κ increases. When the number of mass points is 4, two middle points have similar probability.

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

5 Discussion and Conclusion

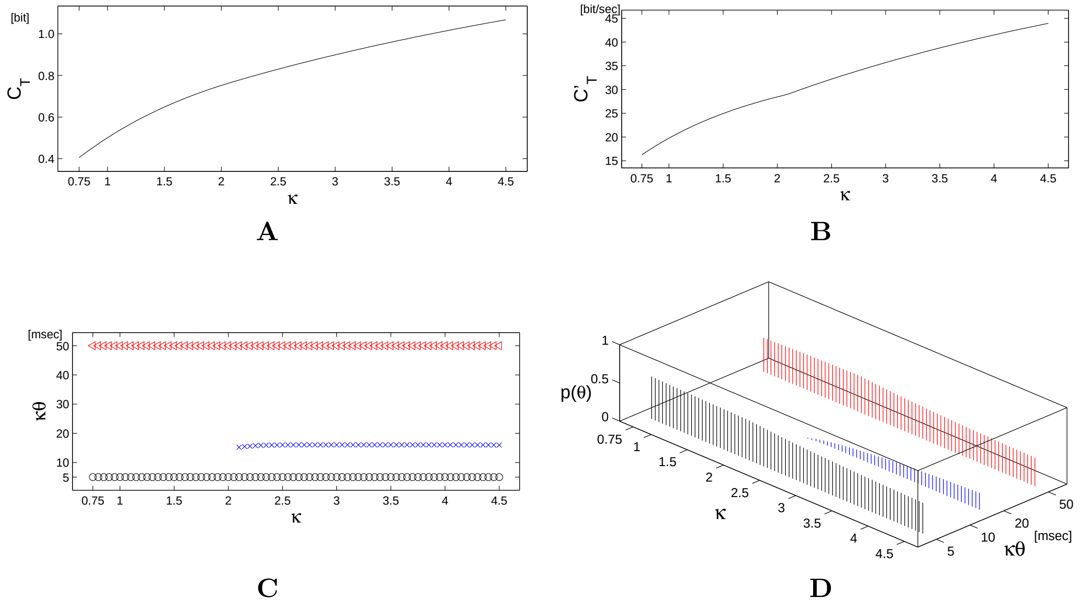
We have proved the channel capacities of a single neuron with temporal and rate coding are achieved with discrete distributions. Numerical studies show that the number of mass points are from 2 to 4 depending on coding and κ. The capacity of a single neuron evaluated in this letter is lower than what has been reported in MacKay and McCulloch (1952); Rapoport and Horvath (1960) (1000 $\sim$ 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 the channel capacity depends on various factors, each factor is discussed separately in the rest of this section.

Encoding: input distribution of θ

First, we discuss the input θ. 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 θ has been set to [5 msec, 50 msec]

---

[3] we used bit instead of nat by dividing capacity defined in eq.(2) by $\log 2$.



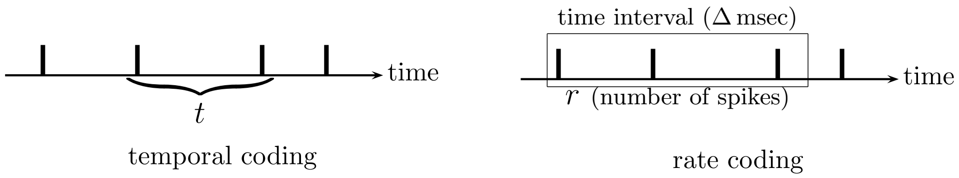
A

B

C

D

Figure 2: Numerical results of temporal coding. **A**: capacity C_T (bit per channel use) for each κ. **B**: information rate C'_T (bit per sec) for each κ (see eq.(18)). **C** and **D** show the capacity achieving distribution computed for each κ. **C**: the probability mass points. For every κ, two points are on the edges of $\Theta(\kappa)$ (a_0/κ and b_0/κ shown as ○ and ◁, respectively). The third point × appears as κ becomes 2.10. **D**: probability of each point shown as the height. The axis for $\kappa\theta$ is logarithmically scaled for visual clarity.



temporal coding

rate coding

Figure 3: Two types of coding: “temporal coding” and “rate coding.”

throughout the letter. The firing rate of each neuron 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 "dynamic range" and Δ, therefore, the capacity result of this paper may not be appropriate for some neurons.

In the range $\Theta(\kappa)$, the distribution of θ has been assumed to be memoryless, that is, θ can be different for every channel use. Scale parameter θ must be changed every 5 msec at most in temporal coding and 25 msec in rate coding. Biologically speaking, θ 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 θ, especially for temporal coding. When θ 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 synchronised with the receiver. One idea is to have a common "clock" and fix θ in a interval. This situation turns out to be rate coding. Another idea is to fix θ for a time proportional to the expected ISI, $\kappa\theta$. In this case, the receiver may miss some spikes. In either cases, 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 et al., 2008). There are multiple works 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) (its extension to real κ 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 points. Although this is good in the sense that neuron can transfer information maximumly with discrete numbers of "firing modes," this does not imply neurons are only using 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.

Noisy channel model

Characteristics of neurons strongly depend on their types. In MacKay and McCulloch (1952), a neuron is assumed to be able to fire within a fixed time precision. They have concluded that the each spike can carry up to 9 bits of information and $1000 \sim 3000$ bits per second could be transferred theoretically. Compared to some biological studies summarised in (Borst & Theunissen, 1999), this observation might be optimistic. We modelled 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 κ between 0.75 to 4.5, which has been indicated in Shinomoto et al. (2003), however in Baker and Lemon (2000), κ is set to 16, which is much larger than our choice. As κ 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 $\kappa = 16$ will 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 κ because of numerical precisions. This may be solved in a 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 them is $\Gamma(m\kappa, \theta/m)$. Since the channel capacity C_T and C_R increases as κ 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.

Decoding

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

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

while the optimal decoder of the rate coding compute the following posterior probability when r is observed

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

The discrete distributions $\varpi_{T,i}$ and $\varpi_{T,i}$ are the posterior distributions of the input θ conditioned on the observations. This is the "soft decoding" and is natural from mathematical viewpoint, however, it may not be a plausible computation for a neuron. Another problem is that the neuron must know the value of κ for soft decoding.

Another natural decoding is the "hard decoding," that is, depending on t or r, only a single θ is considered as the decoding result. The Bayes optimal hard decoding is to choose the θ_i which maximises the posterior distribution. In the case of single neuron information channels, they are defined as

$$\begin{aligned} \hat{\theta}_T &= \underset{\theta_{T,i}}{\operatorname{argmax}}\ \varpi_{T,i} = \underset{\theta_{T,i}}{\operatorname{argmax}}\ \pi_{T,i}\, p(t|\theta_{T,i}\,;\kappa), \\ \hat{\theta}_R &= \underset{\theta_{R,i}}{\operatorname{argmax}}\ \varpi_{R,i} = \underset{\theta_{R,i}}{\operatorname{argmax}}\ \pi_{R,i}\, p(r|\theta_{R,i}\,;\kappa,\Delta), \end{aligned}$$

for temporal and rate coding, respectively. Each decoders 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 Fig. 5.A. In rate coding, r is a nonnegative integer and decisions for integers are shown in Fig. 5.B.

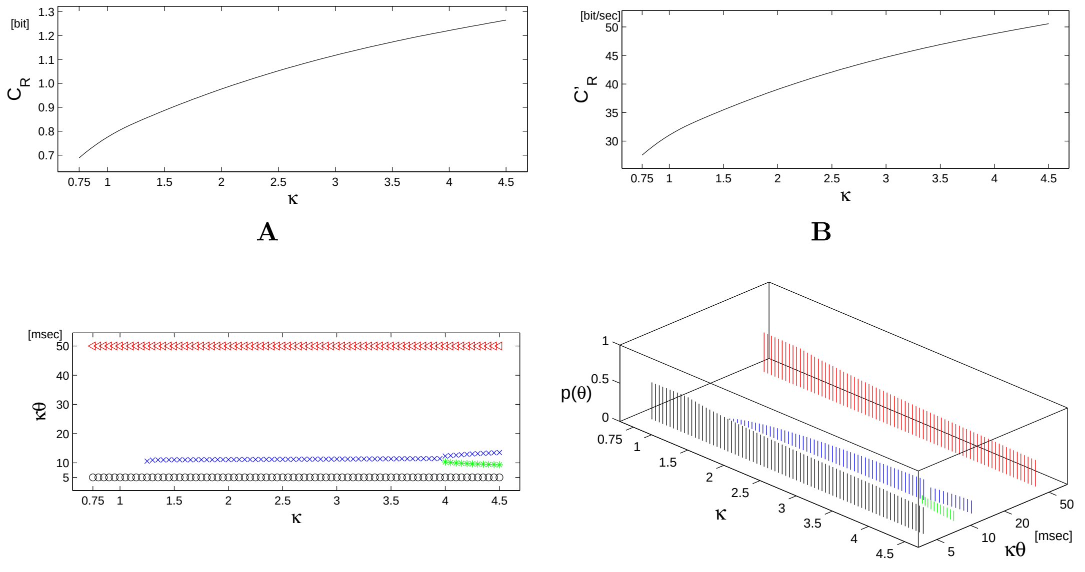
Note that the boundary in Fig. 5.A is stable between $\kappa = 0.75$ to 2.6 and even if the capacity achieving distribution has 3 states for $\kappa \geq 2.10$ (Fig.2), the third point does not appear in decisions until $\kappa > 2.6$. Similar results are observed for rate coding. Although the number of points is more than 3 if $\kappa > 1.20$ (Fig.4), the decision becomes 3 points only when $\kappa > 1.55$ and even if the number of points is 4 for $\kappa \geq 4$, it does not appear as the hard decision. Decision boundaries are not sensitive to small changes of κ.

When a hard decoding is employed, both of 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 Fig.6. It shows that the transferred information is degraded from the optimal soft decoder, however the lost information is not very large.

Related works

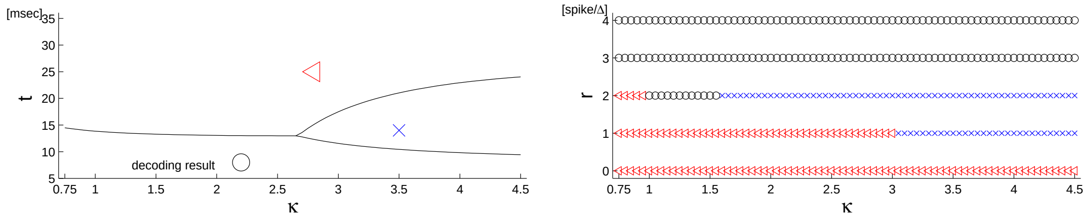
Stein (1967) has discussed the channel capacity of the rate coding where a gamma distribution with a fixed κ was the ISI model. The input was assumed to be discrete distribution of the scale parameter θ on an interval, which happened to be optimal, and the capacity was computed numerically in a similar 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 κ not only for rate coding but also for temporal coding.



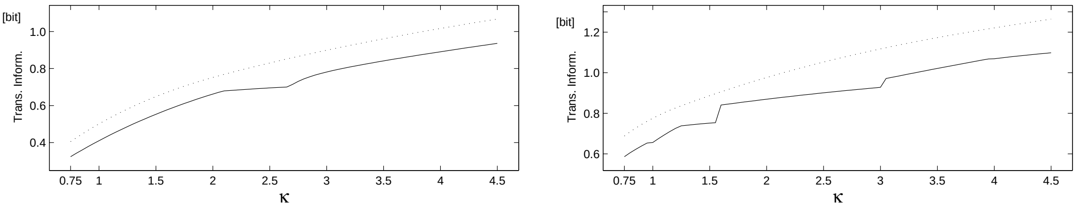
A **B** **C** **D**

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



A **B**

Figure 5: Hard decoding. **A**: hard decoding boundaries for temporal coding. Decision depends on received ISI and κ. $\circ$, $\triangleleft$ and $\times$ corresponds to θ's in Fig.2. **B**: hard decoding for rate coding. Decision depends on received rate and κ. $\circ$, $\triangleleft$ and $\times$ corresponds to θ's in Fig.4.



A **B**

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. Dashed line shows the capacity in Fig.2.A. **B**: transferred information with the capacity achieving distribution and the optimal hard decoder for rate coding. Dashed line shows the capacity in Fig.4.A

Conclusion

The channel capacity and the capacity achieving distribution are obtained for a single neuron information channel. ISIs are modelled with a gamma distribution and two types of coding, temporal and rate, are considered. As the results, capacity achieving distributions are proved to be discrete distributions with finite number of points. Numerical studies show the number of the points are relatively small for moderate choice of κ. It should also be noted that neurons may not use an efficient error-control codes which requires 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 nor 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 gives implications. If the input is a continuous distribution, the transferred information is lower than the capacity, and if the hard decoding is employed, the transferred information is lower than the capacity.

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

A Proof of lemma 1

It is proved by induction.

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

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

where $\Delta_\theta = \Delta/\theta$. Assuming eq.(8) is true for a $m \in \mathbb{Z}^*$, $p(m+1|\theta;\kappa,\Delta)$ is written as follows

$$\begin{aligned} p(m+1|\theta;\kappa,\Delta) &= \int_0^{\Delta}\Big(\frac{t^{\kappa-1}}{\theta^{\kappa}}\Big)\frac{\exp[-t/\theta]}{\Gamma(\kappa)}p\big(m|\theta;\kappa,(\Delta-t)\big)\,dt \\ =\frac{1}{\Gamma(\kappa)}&\int_0^{\Delta_\theta} u^{\kappa-1}e^{-u}\left[P\big(m\kappa,\,(\Delta_\theta-u)\big) - P\big((m+1)\kappa,\,(\Delta_\theta-u)\big)\right]du. \end{aligned} \tag{19}$$

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

$$\frac{1}{\Gamma(\kappa)}\int_0^{\Delta_\theta} u^{\kappa-1}e^{-u}P\big(m\kappa,\,(\Delta_\theta - u)\big)du = P((m+1)\kappa,\Delta_\theta). \tag{20}$$

It is easily checked eq.(20) for $m = 0$. For $m \in \mathbb{Z}^+$ ($\mathbb{Z}^+$ denotes the set of positive integers), it is justified as follows

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

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

$$\begin{aligned} P(\alpha,x) &= e^{-x}\sum_{i=0}^{\infty}\frac{x^{\alpha+i}}{\Gamma(\alpha+i+1)}, \qquad \alpha,x>0, \\ B(\beta,\gamma) &= \int_0^1 t^{\beta-1}(1-t)^{\gamma-1}dt = \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)}, \quad \beta,\gamma>0. \end{aligned} \tag{21}$$

Note that eq.(21) follows from the following equation

$$P(\alpha,x) = P(\alpha+n,x) + e^{-x}\sum_{i=0}^{n-1}\frac{x^{\alpha+i}}{\Gamma(\alpha+i+1)}, \qquad \lim_{n\to\infty}P(\alpha+n,x) = 0, \quad \alpha,x>0. \tag{22}$$

Since eq.(20) holds for $m \in \mathbb{Z}^*$, eq.(19) becomes

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

Equation (8) holds for every $r \in \mathbb{Z}^*$.

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,

$$F_n \overset{w^*}{\to} F \implies I_T(F_n) \to I_T(F), \tag{23}$$

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

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

$h_{T|\theta}(F_n;\kappa) \to h_{T|\theta}(F;\kappa)$ 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

$$\begin{aligned}\lim_n\ h_T(F_n;\kappa) &= -\lim_n \int_0^\infty p(t;F_n,\kappa)\log g(t;F_n,\kappa)dt \\ &= -\int_0^\infty \lim_n\ p(t;F_n,\kappa)\log g(t;F_n,\kappa)dt \qquad (24)\\ &= -\int_0^\infty p(t;F,\kappa)\log g(t;F,\kappa)dt = h_T(F;\kappa) \qquad (25)\end{aligned}$$

The interchange of integral and limit in eq.(24) is justified as follows. From eqs.(5) and (7), $p(t;F,\kappa)$ and $g(t;F,\kappa)$ are bounded as follows.

$$\begin{gathered}\frac{t^{\kappa-1}}{b^\kappa}\frac{\exp[-t/a]}{\Gamma(\kappa)} < p(t;F_n,\kappa) < \frac{t^{\kappa-1}}{a^\kappa}\frac{\exp[-t/b]}{\Gamma(\kappa)} \\ -\frac{t}{a} - \kappa\log b < \log g(t;F_n,\kappa) < -\frac{t}{b} - \kappa\log a.\end{gathered} \tag{26}$$

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

$$\big|p(t;F_n,\kappa)\log g(t;F_n,\kappa)\big| < A_1\, t^{\kappa-1}e^{-t/b} + A_2\, t^\kappa e^{-t/b}. \tag{27}$$

RHS of eq.(27) is integrable as

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

Since eq.(27) is bounded from above with an integrable function, eq.(24) 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 continuous function on F, hence $p(t;F_n,\kappa)\log g(t;F_n,\kappa)$ is also continuous for every $F_n\in\mathcal{F}$. These arguments justify eq.(25) and eq.(23) is justified.

B.2 Proof of lemma 3

Let us define F_η and rewrite $i_T(\theta;F)$ in eq.(6) as follows

$$F_\eta = (1-\eta)F_0 + \eta F, \quad i_T(\theta;F) = -\kappa\log\theta - \int_0^\infty p(t|\theta;\kappa)\log g(t;F,\kappa)dt.$$

Then

$$\begin{aligned}I_T(F_\eta) - I_T(F_0) &= \int_a^b i_T(\theta;F_\eta)dF_\eta - \int_a^b i_T(\theta;F_0)dF_0 \\ &= \eta\left[\int_a^b i_T(\theta;F_\eta)dF - \int_a^b i_T(\theta;F_\eta)dF_0\right] \qquad (28)\\ &\qquad + \int_a^b \big[i_T(\theta;F_\eta) - i_T(\theta;F_0)\big]dF_0. \qquad (29)\end{aligned}$$

The weak derivative of $I_T(F)$ at F_0 is defined as $I'_{T,F_0}(F) = \lim_{\eta \downarrow 0} \big(I_T(F_\eta) - I_T(F_0)\big)/\eta$. By dividing the term in eq.(28) with η 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 eq.(29) 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.

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 of proposition 3. The expectation of R is

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

Since $P(\alpha, x)$ is a strictly decreasing function of α 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

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

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

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

□

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

$$F_n \overset{w^*}{\to} F \implies I_R(F_n) \to I_R(F), \tag{30}$$

From the definitions of $I_R(F)$ and $i_R(\theta, F)$ in eq.(9),

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

Since $i_R(\theta, F)$ is a positive continuous function of θ, 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 a, the following inequality holds from eq.(22).

$$\begin{gathered}
P(r\kappa, \Delta_\theta) - P(r\kappa + \lfloor\kappa\rfloor, \Delta_\theta) \leq p(r|\theta; \kappa, \Delta) \leq P(r\kappa, \Delta_\theta) - P(r\kappa + \lceil\kappa\rceil, \Delta_\theta) \\
e^{-\Delta_\theta} \sum_{i=0}^{\lfloor\kappa\rfloor-1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)} \leq p(r|\theta; \kappa, \Delta) \leq e^{-\Delta_\theta} \sum_{i=0}^{\lceil\kappa\rceil-1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}.
\end{gathered} \tag{31}$$

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

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

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

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

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

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

For $\kappa < 1$

When $\kappa < 1$, the following relation holds from eq.(22).

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

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

$$\begin{aligned}
P(r\kappa, \Delta_\theta) - P(r\kappa + 1, \Delta_\theta) - &\frac{e^{-\Delta_\theta} \Delta_\theta^{(r+1)\kappa}}{\Gamma((r+1)\kappa+1)} \\
&< p(r|\theta; \kappa, \Delta) < P(r\kappa, \Delta_\theta) - P(r\kappa + 1, \Delta_\theta) \\
\frac{e^{-\Delta_\theta} \Delta_\theta^{r\kappa}}{\Gamma(r\kappa+1)} \Big(1 - \Delta_\theta^\kappa \frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+\kappa+1)}\Big) &< p(r|\theta; \kappa, \Delta) < \frac{e^{-\Delta_\theta} \Delta_\theta^{r\kappa}}{\Gamma(r\kappa+1)}.
\end{aligned} \tag{32}$$

Above equation gives the following bound of $p(r; F, \kappa, \Delta)$.

$$p(r; F, \kappa, \Delta) = \int_a^b p(r|\theta; \kappa, \Delta) dF(\theta) > \frac{e^{-\Delta_M} \Delta_m^{r\kappa}}{\Gamma(r\kappa+1)} \Big(1 - \Delta_M^\kappa \frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+\kappa+1)}\Big).$$

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,

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

With the result of proposition 3,

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

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

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

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 $x\to\infty$ $(x\in\mathbb{R}^+)$, the following equation holds*

$$\lim_{x\to\infty}\frac{\sum_{r=1}^{\infty}P(rm,x)}{x}=\frac{1}{m},\quad m\in\mathbb{Z}^+. \tag{33}$$

Proof of proposition 4. From proposition 3 in §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 (21)

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

It is easily checked that

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

Thus, the following linear differential equation is derived.

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

Solving the differential equation, the general solution gives the following form of $S_m(x)$

$$S_m(x)=\frac{x}{m}+\sum_{k=1}^{m-1}C_k e^{(-1+\alpha_k)x}-\frac{m-1}{m^2},\quad \alpha_k=\exp\Big[\frac{2\pi k\sqrt{-1}}{m}\Big]. \tag{34}$$

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

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

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

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

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 Δ_θ, which proves the corollary. ☐

Let us prove theorem 3.

For $\kappa \geq 1$

From eq.(31), $p(r; F, \kappa, \Delta)$ 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

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

where D is the following lower bound.

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

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

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

Since $\overline{R}_{\kappa, \Delta_\theta}$ grows with Δ_θ 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 eq.(32), $p(r; F, \kappa, \Delta)$ is bounded from above as follows

$$p(r; F, \kappa, \Delta) < \frac{e^{-\Delta_m} \Delta_M^{r\kappa}}{\Gamma(r\kappa + 1)} \tag{35}$$

Let us denote r as

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

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

$$p(r|\theta; \kappa, \Delta) = p(r'K + b|\theta; \kappa, \Delta) = q(r'|\theta; \kappa, \Delta) q(b|r', \theta; \kappa, \Delta),$$
$$q(r'|\theta; \kappa, \Delta) = \sum_{b=0}^{K-1} p(r'K + b|\theta; \kappa, \Delta), \quad q(b|r', \theta; \kappa, \Delta) = \frac{p(r'K + b|\theta; \kappa, \Delta)}{q(r'|\theta; \kappa, \Delta)}.$$

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,

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

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

$$q(r'|\theta; \kappa, \Delta) = P(r'K\kappa, \Delta_\theta) - P((r'+1)K\kappa, \Delta_\theta) \geq e^{-\Delta_\theta} \frac{\Delta_\theta^{r'K\kappa}}{\Gamma(r'K\kappa + 1)}. \tag{36}$$

With eqs.(35) and (36),

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

where E is the following lower bound.

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

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

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

Since $\sum_{r'=0}^{\infty} r'\, q(r'|\theta;\kappa,\Delta)$ is equivalent to $\overline{R}_{K\kappa,\Delta_\theta}$, proposition 4 shows that it grows proportional to Δ_θ as $\theta \downarrow 0$. Thus, $i_R(\theta,F)$ is lower bounded with a term which 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.

Acknowledgment

This work was supported in part by the 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.* Oliver & Boyd. translated to English by N. Kemmer.

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*, vol. 143 of *Graduate texts in mathematics*. 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* (Technical Report). Princeton University Technical Report.

Gursoy, M. C., Poor, V., & Verdú, S. (2005). The noncoherent Rician fading channel–part I: Structure of the capacity-achieving input. *IEEE transaction 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.* John Wiley & Sons, Inc.

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.

Miura, K., Okada, M., & Amari, S. (2006). Estimating spiking irregularities using changing environment. *Neural Computation*, *18*, 2359–2386.

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. *The Bell System Technical Journal*, *27*, 379–423 and 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*.

Verdú, S. (1999). Poisson communication theory. Invited talk in the International Technion Communication Day in honor of Israel Bar-David. `http://www.princeton.edu/~verdu`.