

Spiking Neuron Channel

Shiro Ikeda

Department of Mathematical Analysis and Statistical Inference
The Institute of Statistical Mathematics
Tokyo, 106-8569, Japan
Email: shiro@ism.ac.jp

Jonathan H. Manton

Department of Electrical and Electronic Engineering
The University of Melbourne
Victoria, 3010, Australia
Email: jmanton@unimelb.edu.au

Abstract—The information transfer through a single neuron is a fundamental information processing in the brain. This paper studies the information-theoretic capacity of a single neuron by treating the neuron as a communication channel. 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. We prove that for both models, the capacity achieving distribution has only a finite number of probability mass points. This allows us to compute numerically the capacity of a neuron. Our capacity results are in a plausible range based on biological evidence to date.

I. INTRODUCTION

It is widely believed that neurons send information to other neurons in the form of spike trains. Although the precise timing of the spikes is important for the transfer of information, it appears that spike patterns are noisy [1]. Information theory shows that when a communication channel is corrupted by 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 [2]. When a single neuron is modeled as a communication channel, computing the capacity is one of the fundamental problems in neuroscience.

Shortly after the seminal paper of Shannon [2], the capacity of a neuron channel is studied [3]. They assumed a simple noise model and concluded that each spike could carry up to 9 bits of information and that the capacity could be in the vicinity of 1000 to 3000 bps. The capacity of a neuron channel has also been investigated biologically. According to the experimental studies summarized in [4], a single neuron can send at a rate of at most several hundred bps.

The theoretical capacity of a neuron depends on how the neuron is model-led as a communication channel. In this paper, the gamma distribution is employed to describe the stochastic nature of the inter-spike intervals (ISIs), as is done in the neuroscience literature [5], [6]. This differs from that in [3].

To obtain a communication channel though, the coding must also be described. This has long been a subject for discussion. The two major ideas are temporal coding and rate coding. This paper computes the theoretical channel capacity with respect to each of these codings.

As a by-product of determining the capacity, the input distribution which achieves the channel capacity is derived.

Interestingly, for both coding schemes, it is proved to be a discrete distribution with only a finite number of probability mass points. Numerical computations then show that the number of mass points is small, and from this, the capacity of a neuron is computed numerically. The computed capacities are consistent with biologically obtained results.

II. SINGLE NEURON CHANNEL

A. Distribution of ISIs

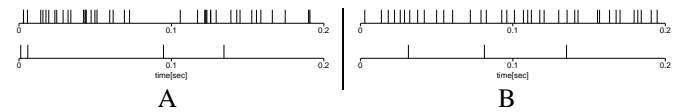


Fig. 1. Simulated spike trains with gamma distributions. The shape parameter κ is 0.75 and 4.5 in A and B, and the expected values of ISI is 5 and 50 msec in upper and lower train, respectively.

The gamma distribution is known to be a suitable distribution to describe the stochastic nature of ISIs [5], [7]. Let T denote an ISI. When it follows a gamma distribution, it is denoted as $T \sim \Gamma(\kappa, \theta)$. The parameter $\kappa > 0$ is the shape parameter and $\theta > 0$ is the scale parameter. Some studies have reported κ of an individual neuron is stable (the value may depends on the type of neuron), while θ changes dynamically over time. Fig. 1 shows simulated spike trains with two different shape parameter κ 's (0.75 and 4.5). When κ is small, spike trains become more irregular.

In [7], a statistical index L_V (local index), is defined to characterize each neuron. From their investigation with biological data, most of the cells' L_V 's are lying between 0.3 and 1.2. If T follows a gamma distribution, $\overline{L_V} = 3/(2\kappa + 1)$ holds and the corresponding interval of κ is [0.75, 4.5].

B. Communication Channel, Capacity, and Single Neuron

Each neuron is a communication channel since it is believed to be transmitting information with spikes. When a channel is noisy, only a limited information can be transmitted through a channel and this limit is the capacity. We focus on the information processing of a single neuron.

Let X be the input to a noisy channel and Y be the output. We assume $X \in \mathcal{X} \subseteq \mathbb{R}$ is a one-dimensional stochastic variable and let $F(X)$ be a cumulative distribution function of X . Communication channel is described as a stochastic model

$p(y|x)$. Let \mathcal{F} be the set of F , in which we are interested and the channel capacity C is defined as

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

$$I(F) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} p(y|x) \ln \frac{p(y|x)}{p(y)} d\mu(y) dF(x),$$

where $p(y) = \int_{x \in \mathcal{X}} p(y|x) dF(x)$ and $\mu(y)$ denotes the measure of $y \in \mathcal{Y}$. Since the channel $p(y|x)$ is not adjustable, the mutual information is a function of $F(x)$ and we denote it as $I(F)$. What are X and Y of a neuron channel? We assume the shape parameter κ of each neuron is fixed, and the scale parameter θ is the only variable parameter. This plays the role of X . An ISI, T , follows a gamma distribution as

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

where θ is a stochastic variable and κ is a parameter. The expectation of T is $\bar{T} = \kappa\theta$.

Before considering what is Y , let us consider the family of all the possible distributions of input θ . One natural assumption is 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), that is,

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

Thus, θ is bounded in $\Theta(\kappa) = \{\theta \mid a(\kappa) \leq \theta \leq b(\kappa)\}$, where $a(\kappa) = a_0/\kappa$ and $b(\kappa) = b_0/\kappa$. In the following, $a(\kappa)$, $b(\kappa)$ and $\Theta(\kappa)$ are denoted as a , b and Θ respectively. Let us define $F(\theta)$ as the cumulative distribution function of θ and \mathcal{F} as

$$\mathcal{F} = \{F : \mathbb{R} \rightarrow [0, 1] \mid F(\theta) = 0, (\theta < a), F(\theta) = 1, (\theta \geq b)\}.$$

F is right-continuous and non-decreasing on Θ and \mathcal{F} includes continuous and discrete distributions.

Now, what is “the output of the channel,” Y , of a neuron communication channel. There are mainly two different ideas in neuroscience. One idea is that Y is the ISI, T , itself. This is called “temporal coding” (Fig.2). The other is that Y is the rate, which is the number of spikes in a fixed time interval. This is called “rate coding” (Fig.2). The mutual information and the capacity depend on coding. The capacity of each coding is formally defined in the following.

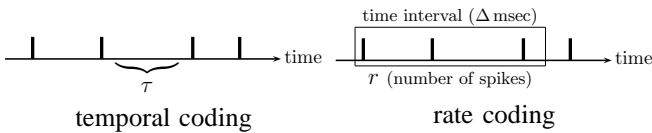


Fig. 2. Two types of coding: “temporal coding” and “rate coding.”

Temporal coding

Received information is T for temporal coding. The mutual information of T and θ is

$$I_T(F) = \int_a^b i_T(\theta; F) dF(\theta), \quad (1)$$

$$\text{where } i_T(\theta; F) = \int_0^\infty p(\tau|\theta; \kappa) \ln \frac{p(\tau|\theta; \kappa)}{p(\tau; F, \kappa)} d\tau,$$

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

The capacity per channel use or equivalently per spike is

$$C_T = \sup_{F \in \mathcal{F}} I_T(F).$$

Rate coding

In rate coding, a time window is set and spikes in an interval Δ is counted. Let us denote the rate as R and the distribution of R as $p(r|\theta; \kappa, \Delta)$, which is defined as follows.

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

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

here \mathbb{Z}^* denotes non-negative integers and $P(\alpha, x)$ is the regularized incomplete gamma function

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

When $\kappa = 1$, a gamma distribution is an exponential distribution and R follows a Poisson distribution. The mutual information of R and θ is defined as

$$I_R(F) = \int_a^b i_R(\theta, F) dF(\theta), \quad (2)$$

$$\text{where } i_R(\theta, F) = \sum_{r=0}^{\infty} p(r|\theta; \kappa, \Delta) \ln \frac{p(r|\theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)},$$

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

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

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

III. THEORETICAL STUDIES

A. Steps to Prove the Discreteness of the Capacity Achieving Distribution

In this section, we prove the capacity achieving distribution of a single neuron channel is a discrete distribution with finite mass points for temporal and rate coding. The capacity achieving distributions have been proved to be discrete for some channels [8], [9], [10], [11] and we follow the same steps of them. First, the common steps of the proofs are shown.

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 $x_n^*(x) \xrightarrow{w^*} x^*(x)$. (See [13], 5.10).*

If X is a real normed linear space of all bounded continuous functions on \mathbb{R} , X^* includes the set of all probability

measures, and “weak convergence” of probability measures is “weak* convergence” on X^* . 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 following maximum is achieved by a unique x^* in S . (See [8], [9], and [13], 5.10).*

$$C = \max_{x^* \in S} J(x^*).$$

From the above discussion, \mathcal{F} is a subset of X^* . It is clear that \mathcal{F} is convex. The following proposition states \mathcal{F} is compact.

Proposition 1. *\mathcal{F} is compact in the Lévy metric topology. (See [11], proposition 1).*

Thus, from theorem 1, if $I_T(F)$ (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 Kuhn-Tucker (K-T) condition on the mutual information is used for the next step of the proof. Before showing the K-T 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 $J'_{F_0} : \mathcal{F} \rightarrow \mathbb{R}$ 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} .

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 $F \in \mathcal{F}$ (See [11], proposition 1)*

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

If $I_T(F)$ (or $I_R(F)$) is weakly differentiable, the K-T condition is derived 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 points of increase. 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.

If they are true, the K-T condition is derived and the discreteness and the finiteness will be checked.

B. 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. The following lemma 2 and theorem 1 imply the capacity for temporal coding is achieved by a unique distribution in \mathcal{F} . The details of proofs will be found in [12].

Lemma 2. *$I_T(F)$ in eq.(1) is a weak* continuous function on $F \in \mathcal{F}$ and strictly concave 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.(1) is weakly differentiable in \mathcal{F} . The weak derivative at $F_0 \in \mathcal{F}$ is*

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

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*

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

(This is proved following the same steps in [11], corollary 1.)

The main result of this subsection 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 $\text{Re } z > 0$. Let us denote the function as $i_T(z; F_0)$.

$$\begin{aligned} i_T(z; F_0) &= -\kappa \ln z - \int_0^\infty p(\tau|z; \kappa) \ln g(\tau; F_0, \kappa) d\tau, \\ \text{where } g(\tau; F_0, \kappa) &= \int_a^b \frac{\exp[-\tau/\theta]}{\theta^\kappa} dF_0(\theta). \end{aligned}$$

If E_0 in corollary 1 has infinite points, E_0 has a limit point. Hence, the identity theorem implies $i_T(z; F_0) = I_T(F_0)$ for the region $\text{Re } z > 0$. This region includes the positive real line and

$$-\int_0^\infty p(\tau|\theta; \kappa) \ln g(\tau; F_0, \kappa) d\tau = \kappa \ln \theta + I_T(F_0), \quad (3)$$

is implied for $\theta > 0$. However, it is not difficult to show the LHS of eq.(3) is bounded as follows.

$$-\int_0^\infty p(\tau|\theta; \kappa) \ln g(\tau; F_0, \kappa) d\tau \geq \frac{\kappa\theta}{b} + \kappa \ln a.$$

The LHS of eq.(3) grows at least with $\kappa\theta/b + \text{const}$ while the RHS is $\kappa \ln \theta + \text{const}'$. This cannot hold for $\theta \in \mathbb{R}^+$. This is a contradiction and the optimal distribution has a finite number of mass points. ■

C. Discreteness of the Capacity Achieving Distribution for Rate Coding

The capacity achieving distribution of a Poisson channel under peak and average power constraints has been proved to be discrete with a finite point of supports [10]. This directly proves the case $\kappa = 1$ of rate coding. For $\kappa \neq 1$ further study is needed. The following lemma 4 and theorem 1 imply the capacity for rate coding is achieved by a unique distribution in \mathcal{F} . The details of proofs will be found in [12].

Lemma 4. $I_R(F)$ in eq.(2) is a weak* continuous function on $F \in \mathcal{F}$ and strictly concave in \mathcal{F} . (The strict concavity follows the proof in §7.2 of [10].)

Lemma 5. $I_R(F)$ in eq.(2) is weakly differentiable in \mathcal{F} . The weak derivative at $F_0 \in \mathcal{F}$ is

$$I'_{R,F_0}(F) = \int_a^b i_R(\theta; F_0) dF - I_R(F_0). \quad (4)$$

(The proof is identical to the proof of lemma 3.)

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

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

(This is proved following the same steps in [11], corollary 1.)

The following theorem proves that the capacity achieving distribution is a discrete distribution with a finite number of mass points.

Theorem 3. Under the constraint $\theta \in \Theta$, the channel capacity of a single neuron channel with rate coding is achieved by a discrete distribution with a finite number of mass points.

Outline of proof: The extension of $i_R(\theta; F)$ to the complex plain z is defined as

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

Since $P(\alpha, z)$ and $\ln z$ 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, E_0 has a limit point. Hence 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 and the following is implied

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

The proof is completed by deriving a contradiction. ■

IV. NUMERICAL STUDIES

A. Common Steps

Although the capacity achieving distribution of each coding is proved to be discrete with a finite number of mass points, position and probability of each point are not provided. In this

section, the capacity and the capacity achieving distribution is computed numerically for each coding.

The strategy to compute the capacity and the capacity achieving distributions for temporal and rate coding is as follows. Other related works use similar methods [8], [9], [11].

- 1) Initialize the number of the points N as 2.
- 2) Set the position and probability of each point as θ_j and π_j , ($j = \{1, \dots, N\}$), respectively.
 $a \leq \theta_1 < \dots < \theta_N \leq b$, $\sum_{j=1}^N \pi_j = 1$, $\pi_j > 0$, $j \in \{1, \dots, N\}$.
- 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) Check the corresponding K-T condition to see if it is the capacity achieving distribution.
- 5) If K-T condition is satisfied, exit, otherwise increase N by 1 and go to step 2.

The computation of each step is difficult since $i_T(\theta; F)$ in eq.(1) or $i_R(\theta; F)$ in eq.(2) are not provided in closed forms. We evaluated $i_T(\theta; F)$ with the Gauss-Laguerre quadrature and $i_R(\theta; F)$ with truncated series. The range of θ is set to $5/\kappa \leq \theta \leq 50/\kappa$ which makes the expected firing rate from 5 msec to 50 msec. The capacity and the capacity achieving distribution for each coding are computed for every 0.05 of $\kappa \in [0.75, 4.5]$.

B. Numerical Results

Figure 3 (left) shows the computed capacity for each κ . The information C_T increases monotonically as κ increases. This is natural since ISIs become more regular as κ increases 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 Fig. 3 (right). 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. The probabilities of both ends tend to be similar, while the probability of the third point increases gradually as κ increases.

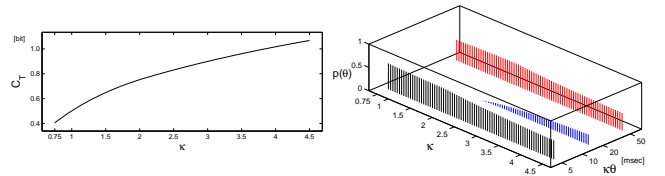


Fig. 3. Capacity C_T (bit per channel use) as a function of κ (left). The capacity achieving distribution of each κ (right): since the range of θ scales with κ , $\kappa\theta$ is taken as an axis. Probability of each point is shown as the height. The axis for $\kappa\theta$ is log-scaled for visual clarity.

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. Fig. 4 (left) 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 Fig. 4 (right). 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 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.

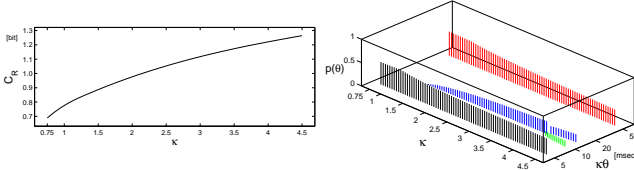


Fig. 4. Capacity C_R (bit per channel use) as a function of κ (left). The capacity achieving distribution of each κ (right): Probability of each point is shown as the height. The axis for $\kappa\theta$ is log-scaled for visual clarity.

C_T and C_R are the maximum information transferred per a channel use. It is also important to show the information rate. In the case of temporal coding, we define the rate C'_T bps as $C'_T = C_T/\bar{\tau}$, where $\bar{\tau} = \kappa \sum_{j=1}^N \pi_j \theta_j$ and show it in Fig.5 (left). In the case of rate coding, the maximum information transferred per time is easily computed since Δ is fixed, let us define $C'_R = C_R/\Delta$ bps, which is shown in Fig.5 (right).

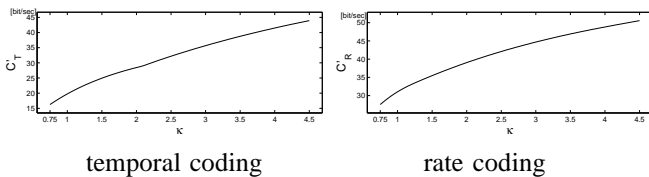


Fig. 5. Information rate of temporal coding (left) and rate coding (right).

V. DISCUSSION

The transmission of information through a single neuron is perhaps the most basic processing of information carried out by the brain. This paper assumed the output of a neuron is a gamma-distributed spike train. Furthermore, the information transmission was modeled by a communication channel whose output is either the time between successive spikes (ISI) or the number of spikes in a unit of time (rate), corresponding to temporal coding and rate coding respectively. The numerical studies show that the channel capacity of a “typical” neuron is less than 100 bps, which is consistent with the biological measurements reported in [4] and is lower than what was reported in [3] where temporal coding is considered. For the rate coding, Stein numerically computed the capacity [6]. His result is almost identical to our numerical results. However, Stein assumed that the channel capacity achieving distribution is a discrete distribution. We believe this paper gives the

the first proof for the discreteness of the channel capacity distribution for every $\kappa > 0$.

Interestingly, this paper proved that the capacity achieving distributions are discrete and have only a finite number of probability mass points. This has several implications. It allows for the capacity to be computed numerically. Furthermore, it may prove to be helpful for measuring the capacity of a neuron [4], since it implies that only a few modes of inputs are necessary in order to measure the capacity accurately.

Even if our models of the neuron are accurate, the resulting capacity calculations would not be biologically attainable; there is no biological evidence to suggest the input to a neuron follows a discrete distribution, nor is it likely the output can be measured exactly. Nevertheless, we hope our results serve as a general guide for how much information can be obtained from a single recording. This may help neurophysiological experiments where the results of many trials are accumulated.

In terms of further work, we remark that there are many tunable parameters, such as κ , Δ , a_0 , and b_0 . Their admissible values are likely to depend on the type of neuron under consideration. For example, we chose to set κ to be between 0.75 to 4.5, but [5] suggested that it may be as high as 15 for some cells. Another issue is that we have only considered independent sources. This should be extended to correlated sources.

ACKNOWLEDGMENT

This work was partially supported by the Grant-in-Aid for Scientific Research No. 18079013, MEXT, Japan.

REFERENCES

- [1] Z. F. Mainen and T. J. Sejnowski, “Reliability of spike timing in neocortical neurons,” *Science*, vol. 268, pp. 1503–1506, 1995.
- [2] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
- [3] D. M. MacKay and W. S. McCulloch, “The limiting information capacity of a neuronal link,” *Bull. Math. Biophys.*, vol. 14, pp. 127–135, 1952.
- [4] A. Borst and F. E. Theunissen, “Information theory and neural coding,” *Nature Neuroscience*, vol. 2, no. 11, pp. 947–957, 1999.
- [5] S. N. Baker and R. N. Lemon, “Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels,” *J. Neurophysiol.*, vol. 84, pp. 1770–1780, 2000.
- [6] R. B. Stein, “The information capacity of nerve cells using a frequency code,” *Biophysical Journal*, vol. 797–826, 1967.
- [7] S. Shinomoto, K. Shima, and J. Tanji, “Differences in spiking patterns among cortical neurons,” *Neural Computation*, vol. 15, no. 12, pp. 2823–2842, 2003.
- [8] I. C. Abou-Faycal, M. D. Trott, and S. Shamaï (Shitz), “The capacity of discrete-time memoryless Rayleigh-fading channels,” *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.
- [9] M. C. Gursoy, H. V. Poor, and S. Verdú, “The capacity of the noncoherent Rician fading channel,” Princeton Univ. Tech. Report, 2002.
- [10] S. Shamaï (Shitz), “Capacity of a pulse amplitude modulated direct detection photon channel,” *IEE Proc.*, vol. 137, no. 6, pp. 424–430, 1990.
- [11] J. G. Smith, “The information capacity of amplitude- and variance-constrained scalar Gaussian channels,” *Information and Control*, vol. 18, pp. 203–219, 1971.
- [12] S. Ikeda and J. H. Manton, “Capacity of a single spiking neuron channel,” *Neural Computation*, vol.21, no. 6, 2009.
- [13] D. G. Luenberger, *Optimization by Vector Space Method*. John Wiley & Sons, Inc., 1969.