

Energy Efficient Neurons with Generalized Inverse Gaussian Conditional and Marginal Hitting Times

Jie Xing

Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, Virginia 22903
Email: jx7g@virginia.edu

Toby Berger

Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, Virginia 22903
Email: tb6n@virginia.edu

Abstract—Neuronal information processing is energetically costly. Energy supply restrictions on information processing have caused brains to evolve to compute and communicate information with remarkable efficiency. Indeed, energy minimization subject to functional constraints is a unifying principle. Toward better comprehension of neuronal information processing and communication from an information-energy standpoint, we consider a neuron model with a generalized inverse Gaussian (GIG) conditional density. This GIG model arises from a Lévy diffusion process that is more general than that of a Wiener process with drift. We show that, when the GIG neuron operates so as to maximize bits per Joule (bpJ), the output interspike interval (ISI) distribution is a related GIG marginal distribution. Also, we specify how to obtain the associated input distribution $f_{\Lambda}(\lambda)$ numerically. By generalizing from the Gamma and inverse Gaussian (IG) families to the GIG family, the derived results contain both the homogeneous Poisson and Wiener processes as special cases. The results allow us to readily compute the tradeoff between information rate (bits/second) and average power (Joules/second) in the GIG class, reminiscent of Shannon's celebrated formula for such curves for the additive Gaussian family.

I. PROBLEM STATEMENT AND PRELIMINARIES

A. Introduction

Neuronal information processing is energetically costly. The approximately 86 billion neurons composing the human brain, despite comprising only 2% of the body weight, consume on average 20% of the energy provided to the whole body [1][2]. Energy supply restrictions on information processing have caused brains to evolve to compute and communicate information with remarkable efficiency. Indeed, energy minimization subject to functional constraints is a unifying principle [3]. Since neurons are expressly designed to exchange information with one another, understanding information processing and energy expenditure at the nodal level of the network is fundamental. We need to understand how an adequate supply of adenosine triphosphate (ATP), the molecular unit of currency for energy transfer, is ensured for each neuron's information processing. Furthermore, an increasing fraction of neuroscientists subscribe to the view that each neuron's design should maximize the ratio of the rate at which it conveys information to the rate at which it expends energy.

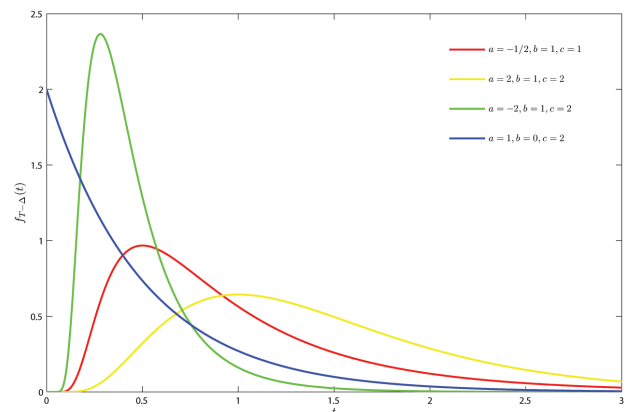


Figure 1. GIG densities for four choices of parameters (a, b, c) in Eq. (11).

Information theory has often been applied in neuroscientific data analysis and biological systems modeling [4][5][6]. However, energy-efficient neural codes have only been studied for less than twenty years [7][8][9]. Evidence supporting energy efficiency has been reported for ion channel [3], action potential [10], synapse [11], photoreceptor [12], retina [13][14], grey matter [15], white matter [16] and cortex [17]. Laughlin and Sejnowski discussed communication in cortical networks from an energy-efficiency point of view [18].

Toward better comprehension of neuronal information processing and communication from an information-energy standpoint, we earlier proposed and studied mathematical models of single neurons as engines of computation and communication based both on homogeneous Poisson processes [19][20] and on the Wiener process [21].

In this paper the goal is to extend the analysis to a model supported by a more general Lévy diffusion process. This results in a fixed threshold hitting time, the probability density function (pdf) of which belongs to the three-parameter generalized inverse Gaussian (GIG) family. The GIG class subsumes many key two-parameter first passage time distribution families including the Gamma and IG distributions. By virtue of the infinitely divisible property of the GIG family, the associated GIG neuron channel model represents the random excitation and inhibition intensity at the axon initial segment in a manner that appeals both

to neuroscientists and to mathematical statisticians, thereby promising to deepen our understanding of how information processing is performed at the ionic level, at the neuronal level, and perhaps eventually at the network level.

B. GIG Hitting Times

The work of Barndorff-Nielsen et al. [22] [23] describes a diffusion that, for a given value λ of a positive drift rate random variable (r.v.) Λ , creates a random first passage time (hitting time) to a fixed threshold level that possesses the GIG conditional pdf with parameters α, β and γ

$$f_{\tilde{T}|\Lambda}(t|\lambda) = C_\lambda t^{\alpha-1} \exp\left(-\gamma\lambda t - \frac{\beta}{\lambda t}\right), \quad (1)$$

where the normalizing constant's reciprocal is

$$C_\lambda^{-1} = 2\left(\frac{\beta}{\gamma\lambda^2}\right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma}) \quad (2)$$

$$= 2\left(\frac{\beta}{\gamma}\right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma})\lambda^{-\alpha} := C_\alpha^{-1}\lambda^{-\alpha}, \quad (3)$$

α is real, β is real and nonnegative, γ is real and positive, $K_\alpha(\cdot)$ is a modified Bessel function of the second kind of order α and $\tilde{T} := T - \Delta$, where Δ is the duration of the modeled neuron's refractory period.

C. Modeling Neural Inputs and Outputs

We first introduce a mathematical framework for how a single neuron stochastically processes and communicates information. Let us call the cortical neuron being modeled "neuron j ", or simply " j ". The output of j is a temporal sequence of effectively identically shaped narrow voltage spikes, also known as action potentials (AP's). The durations of the interspike intervals (ISI's) between j 's AP's constitute a random sequence which we denote by $\{T_k, k = 1, 2, \dots\}$.

Neuron j 's external excitation and inhibition comprises a collection of neural spike trains generated by the afferent presynaptic neurons, called j 's afferent cohort. j 's afferent synapses are connections between these neurons and j . When a spike arrives at the presynaptic terminal of a member of j 's afferent cohort, the depolarization it provides opens the calcium channels in the presynaptic bouton, producing a calcium influx. Calcium elevations in the cytoplasm permit vesicles containing neurotransmitters to fuse quickly with the presynaptic membrane and to release an amount of neurotransmitters based on the size of the vesicle. Those neurotransmitters diffuse within the synaptic cleft. Although some escape, many bind to the postsynaptic receptors. That, in turn, causes the opening of the ion channels, eventually resulting in a net gain of positive or negative voltage across the postsynaptic membrane, either via an excitatory postsynaptic potential (EPSP) or an inhibitory postsynaptic potential (IPSP), respectively. Both EPSP's and IPSP's combine to produce the postsynaptic potential (PSP) at the axon hillock. Once the PSP exceeds a triggering threshold, j generates another output spike that propagates along its axon to each member of a family of neurons, referred to as j 's efferent cohort. A typical neuron in primate sensory cortex has afferent and efferent cohorts, each of which is composed of circa 10,000 neurons.

As in [20], we allow the weights of the synapses to differ from one another. Now we also admit inhibitory synapses. Furthermore, we model the excitation intensity not presynaptically as in [19] but postsynaptically as the PSP at the axon hillock generated by the arrival thereto and the leakage therefrom of both positively and negatively charged ions. We continue to assume that each of j 's refractory periods has the same duration, Δ . This theoretical extension embraces the plasticity of a neuron's synaptic weights, widely considered essential to learning and memory, thereby increasing the chances that the model has practical significance [20][25].

We model the PSP as a random measure with a continuous-time instantaneous rate function, $\Lambda(t)$, defined by

$$\Lambda(t) := \lim_{dt \rightarrow 0} \frac{\pm Q \cdot \Pr[\text{arrival/departure of ion in } (t, t + dt)]}{dt}, \quad (4)$$

where Q is the magnitude of the charge of the ion and "+" applies if the ion is either positively charged and is arriving or negatively charged and is departing; otherwise "-" applies. As in [19] [20] [21], we take a time average operation over the rate function $\Lambda(t)$ and obtain

$$\Lambda_k = \frac{1}{T_k - \Delta} \int_{S_{k-1} + \Delta}^{S_k} \Lambda(u) du, \quad (5)$$

where T_k is the k th ISI duration of j and $S_k = T_1 + \dots + T_k$.

Henceforth, we suppress the ISI index k and just write T and Λ . Thus, when $\Lambda = \lambda$, the PSP is being assumed to build up according to a diffusion process with a time-averaged intensity λ which must be positive for the threshold to be reached.

D. Information Rate

Here we are interested in the Shannon mutual information, $I(\Lambda; T)$. Although this has been defined for a single pair of r.v.'s Λ and T , it has been shown that it is a good first-order approximation to the long term information in bits per spike, namely

$$I := \lim_{N \rightarrow \infty} \frac{1}{N} I(\Lambda_1, \dots, \Lambda_N; T_1, \dots, T_N), \quad (6)$$

lacking only an information decrement that addresses correlation among successive Λ_i 's, which is approximated to first order in [19]. Since \tilde{T} is a one-to-one function of T , we have $I(\Lambda; T) = I(\Lambda; \tilde{T})$, which in turn is defined as

$$I(\Lambda; \tilde{T}) = \mathbb{E} \left[\log \left(\frac{f_{\tilde{T}|\Lambda}(\tilde{T}|\Lambda)}{f_{\tilde{T}}(\tilde{T})} \right) \right], \quad (7)$$

where the expectation is taken with respect to the joint pdf (jpdf) of Λ and \tilde{T} . Henceforth, we contract notation from \tilde{T} to T . Toward determining $I(\Lambda; T)$, we proceed to analyze $f_{T|\Lambda}(t|\lambda)$ and $f_T(t)$ in the GIG case.

II. WHAT IS BEING OPTIMIZED INSIDE THE NEURON?

Neuron j has the ability to slowly adapt parameters in $f_{T|\Lambda}(t|\lambda)$ by changing synaptic weights via the size and number of presynaptic active zones, the number of vesicles and postsynaptic receptors, among many other properties. In the case of the GIG channel of Eq. (1) above, it is able to change the three parameters α, β and γ in the

conditional pdf $f_{T|\Lambda}(t|\lambda)$. It is reasonable to postulate that, if the input distribution $f_\Lambda(\lambda)$ changes slowly over time, j will adjust these parameters in an attempt to preserve its bits per joule (bpJ) optimization. From the information theory point of view, it is the joint distribution of Λ and T that determines $I(\Lambda; T)$. Moreover, $f_{\Lambda, T}(\cdot, \cdot)$ also determines the average energy that j expends, which we model in terms of functions of Λ and T . Neuron j has a modicum of control over the marginal distribution of Λ , which we measure at the ion level at the axon hillock. For example, j can do this by the aforementioned synaptic plasticity including Long-term potentiation (LTP) and Long-term depression (LTD). However, the bulk of the variations of $f_\Lambda(\cdot)$ is effected by phenomena external to j . We consider that j 's internal modifications to $f_\Lambda(\cdot)$ are in the spirit of a one-to-one transformation of the afferent excitation/inhibition intensity statistics into a form compatible with bpJ optimization in conjunction with a GIG $f_{T|\Lambda}(\cdot|\cdot)$ with appropriately modified values of α, β, γ .

III. BITS PER JOULE OPTIMALITY CONDITION

For any choice of $f_\Lambda(\lambda)$, the corresponding pdf of T is given by the integral equation

$$f_T(t) = \int_0^\infty d\lambda f_{T|\Lambda}(t|\lambda) f_\Lambda(\lambda), \quad \forall t \geq 0. \quad (8)$$

From Appendix A of [19], we know that $f_{\Lambda, T}(\cdot, \cdot) = f_\Lambda(\cdot) f_{T|\Lambda}(\cdot|\cdot)$ is bpJ-maximizing if and only if the $f_T(\cdot)$ that it generates via Eq. (8) satisfies

$$\int_0^\infty dt f_{T|\Lambda}(t|\lambda) \left[\log \left(\frac{f_{T|\Lambda}(t|\lambda)}{f_T(t)} \right) - \sum_{i=1}^6 g_i(\lambda, t) \right] = 0, \quad \forall \lambda \geq 0, \quad (9)$$

where $g_i, 1 \leq i \leq 6$, are the energy terms itemized as below:

- $g_1(\lambda, t) = A$, a positive constant. An example of this is the energy j must expend in order to propagate the spike it generates at the end of the ISI to all its target neurons. This and other fixed costs per ISI sum to A .
- $g_2(\lambda, t) = Bt$, where B is a positive constant. The metabolic energy the neuron expends keeping itself healthy and nourished during an ISI would be an example. This grows linearly with the duration of the ISI no matter whether j is processing relatively intensely during it or not.
- $g_3(\lambda, t) = C\lambda t$, C a positive constant. This term addresses the energy j devotes to generating contributions to the PSP accumulation by processing all the spikes afferent to it during the ISI.¹
- $g_4(\lambda, t) = \frac{L}{t}$, L being a positive constant. g_4 charges a high energy cost for small values, t , of the r.v. T . That such a small- t penalty must exist is apparent from the definition of Δ as the least amount of time after one AP that another can be produced in response to natural afferent excitation. Although an experimenter

¹In order for $C\lambda t$ not to be an oversimplification, there needs to be, regardless of the value of λ , enough separate contributions to PSP during an ISI that the coefficient of variation of the number of them and that of the weights of the synapses that generate them both are smaller than, say, 0.1. Primary cortical neurons, among others, conform to these conditions.

can evoke AP's separated by less than Δ by injecting a large current into the axon hillock, this requires supplying power far exceeding what the neuron can expend on its own. It stands to reason that the closer to $t = \Delta$ the neuron produces an AP in response to natural stimulation, the more energy it takes to do so.

- $g_5(\lambda, t) = \frac{G}{\lambda t}$, G being a positive constant. Although $\mathbb{E}[T|\Lambda = \lambda]$ is of the form $Const./\lambda$, so that $\mathbb{E}[\Lambda T] = Const.$, it is still true that no matter how small (large) any λ is, it will occasionally produce a t that also is small (large). g_5 associates an energy expenditure that "corrects" for departures of ΛT from $\mathbb{E}[\Lambda T]$.
- $g_6(\lambda, t) = -D \log t$. If $D > 0$, then this logarithmic term charges an energy penalty whenever $t < 1$ and "rewards" t -values smaller than the time unit, and conversely, if $D < 0$. Note that this logarithmic term is dominated by $g_2(\lambda, t) = Bt$ when t is large and by $g_4(\lambda, t) = L/t$ when t is small, provided D is not many times larger than B or L , respectively. Hence g_6 serves as a fine tuning adjustment to g_2 and/or g_4 . It is widely felt that most neurons are finely tuned.

Thus, equation (9) becomes the following:

$$\int_0^\infty dt f_{T|\Lambda}(t|\lambda) \left[\log \left(\frac{f_{T|\Lambda}(t|\lambda)}{f_T(t)} \right) - \left(A + Bt + C\lambda t + \frac{G}{\lambda t} + \frac{L}{t} \right) + D \log t \right] = 0, \quad \forall \lambda \geq 0, \quad (10)$$

where the conditional pdf $f_{T|\Lambda}(t|\lambda)$ follows Eq. (1).

IV. MARGINAL OUTPUT ISI DISTRIBUTION $f_T(t)$ IS A GIG DISTRIBUTION

Temporarily assume that when bpJ is maximized by virtue of satisfaction of Eq. (10), the marginal ISI pdf $f_T(t)$ follows a GIG distribution

$$f_T(t) = C_a t^{a-1} \exp \left(-ct - \frac{b}{t} \right), \quad (11)$$

where

$$C_a^{-1} = 2 \left(\frac{b}{c} \right)^{\frac{a}{2}} K_a(2\sqrt{bc}). \quad (12)$$

After substituting (1) and (11) into (10), it follows that

$$\begin{aligned} & \int_0^\infty dt f_{T|\Lambda}(t|\lambda) \left[\log C_\lambda - A + (\alpha - 1 + D) \log t \right. \\ & \quad \left. - Bt - (\gamma + C)\lambda t - \frac{(\beta + G)/\lambda + L}{t} \right] \\ &= \int_0^\infty dt f_{T|\Lambda}(t|\lambda) \left[\log C_a + (a - 1) \log t - ct - \frac{b}{t} \right], \\ & \quad \forall \lambda \geq 0. \end{aligned} \quad (13)$$

Next, note the following expectation equalities:

$$\begin{aligned} \mathbb{E}[T|\lambda] &= \int_0^\infty dt t f_{T|\Lambda}(t|\lambda) \\ &= \int_0^\infty dt C_\lambda t^\alpha \exp \left(-\gamma\lambda t - \frac{\beta}{\lambda t} \right) \\ &= \frac{2 \left(\frac{\beta}{\gamma\lambda^2} \right)^{\frac{\alpha+1}{2}} K_{\alpha+1}(2\sqrt{\beta\gamma})}{2 \left(\frac{\beta}{\gamma\lambda^2} \right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma})} \\ &= \sqrt{\frac{\beta}{\gamma}} \frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \frac{1}{\lambda}; \end{aligned} \quad (14)$$

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{T} \middle| \lambda\right] &= \int_0^\infty dt \frac{1}{t} f_{T|\Lambda}(t|\lambda) \\
&= \int_0^\infty dt C_\lambda t^{\alpha-2} \exp\left(-\gamma\lambda t - \frac{\beta}{\lambda t}\right) \\
&= \frac{2\left(\frac{\beta}{\gamma\lambda^2}\right)^{\frac{\alpha-1}{2}} K_{\alpha-1}(2\sqrt{\beta\gamma})}{2\left(\frac{\beta}{\gamma\lambda^2}\right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma})} \\
&= \sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \lambda; \tag{15}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\log T|\lambda] &= \mathbb{E}[\log \lambda T|\lambda] - \log \lambda \\
&= \int_0^\infty dt C_\lambda t^{\alpha-1} \exp\left(-\gamma\lambda t - \frac{\beta}{\lambda t}\right) \log \lambda t - \log \lambda \\
&\stackrel{a}{=} \frac{1}{\lambda^\alpha} \int_0^\infty dx C_\lambda x^{\alpha-1} \exp\left(-\gamma x - \frac{\beta}{x}\right) \log x - \log \lambda \\
&= \frac{1}{2\left(\frac{\beta}{\gamma}\right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma})} \\
&\quad \times \int_0^\infty dx x^{\alpha-1} \exp\left(-\gamma x - \frac{\beta}{x}\right) \log x - \log \lambda \\
&\stackrel{b}{=} \frac{2\left(\frac{\beta}{\gamma}\right)^{\frac{\alpha-1}{2}} K_{\alpha-1}(2\sqrt{\beta\gamma}) - 2\left(\frac{\beta}{\gamma}\right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma})}{2\left(\frac{\beta}{\gamma}\right)^{\frac{\alpha}{2}} K_\alpha(2\sqrt{\beta\gamma})} - \log \lambda \\
&= \sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} - 1 - \log \lambda. \tag{16}
\end{aligned}$$

Equality (a) is due to the replacement of λt with x . Our derivation of Equality (b) is too lengthy to be included here.

Rewriting Eq. (13) as

$$\begin{aligned}
\log C_\lambda - \log C_a - A + (\alpha + D - a)\mathbb{E}[\log T|\lambda] - (B - c \\
+ (\gamma + C)\lambda)\mathbb{E}[T|\lambda] - \left(L - b + \frac{\beta + G}{\lambda}\right)\mathbb{E}\left[\frac{1}{T} \middle| \lambda\right] &= 0, \\
\forall \lambda \geq 0, \tag{17}
\end{aligned}$$

and substituting Eqs. (14), (15) and (16) into Eq. (17) yields

$$\begin{aligned}
\log C_\lambda - \log C_a - A + (\alpha + D - a)\left(\sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \right. \\
\left. - 1 - \log \lambda\right) - (B - c + (\gamma + C)\lambda)\left(\sqrt{\frac{\beta}{\gamma}} \frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \right. \\
\left. \times \frac{1}{\lambda}\right) - \left(L - b + \frac{\beta + G}{\lambda}\right)\left(\sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \lambda\right) &= 0, \\
\forall \lambda \geq 0. \tag{18}
\end{aligned}$$

Since

$$\begin{aligned}
&\log C_\lambda - \log C_a - A \\
&= \left[\frac{a}{2} \log \frac{b}{c} - \frac{\alpha}{2} \log \frac{\beta}{\gamma} + \log \frac{K_a(2\sqrt{bc})}{K_\alpha(2\sqrt{\beta\gamma})} - A\right] + \alpha \log \lambda \\
&:= C_{const} + \alpha \log \lambda, \tag{19}
\end{aligned}$$

Eq. (18) becomes

$$\begin{aligned}
C_{const} + \alpha \log \lambda + (\alpha + D - a)\left(\sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} - 1 \right. \\
\left. - \log \lambda\right) - (B - c + (\gamma + C)\lambda) \cdot \left(\sqrt{\frac{\beta}{\gamma}} \frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \right. \\
\left. \times \frac{1}{\lambda}\right) - \left(L - b + \frac{\beta + G}{\lambda}\right)\left(\sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} \lambda\right) &= 0, \\
\forall \lambda \geq 0. \tag{20}
\end{aligned}$$

By letting $a = D$, $b = L$ and $c = B$, all λ 's cancel on both sides of Eq. (20), yielding

$$\begin{aligned}
C_{const} - \alpha + \alpha \sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} - (\gamma + C) \sqrt{\frac{\beta}{\gamma}} \\
\times \frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} - (\beta + G) \sqrt{\frac{\gamma}{\beta}} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})} &= 0, \\
\forall \lambda \geq 0. \tag{21}
\end{aligned}$$

Multiplying both sides of Eq. (21) by $K_\alpha(2\sqrt{\beta\gamma})$ produces

$$\begin{aligned}
(C_{const} - \alpha)K_\alpha(2\sqrt{\beta\gamma}) + (\alpha - \beta - G) \sqrt{\frac{\gamma}{\beta}} K_{\alpha-1}(2\sqrt{\beta\gamma}) \\
- (\gamma + C) \sqrt{\frac{\beta}{\gamma}} K_{\alpha+1}(2\sqrt{\beta\gamma}) &= 0, \forall \lambda \geq 0. \tag{22}
\end{aligned}$$

Due to the equality that [26]

$$K_\alpha(z) = -\frac{z}{2\alpha}(K_{\alpha-1}(z) - K_{\alpha+1}(z)), \tag{23}$$

we have

$$K_\alpha(2\sqrt{\beta\gamma}) = -\frac{\sqrt{\beta\gamma}}{\alpha}(K_{\alpha-1}(2\sqrt{\beta\gamma}) - K_{\alpha+1}(2\sqrt{\beta\gamma})). \tag{24}$$

Furthermore, combining Eq. (24) and Eq. (22), we can obtain two equations in the coefficients in front of the Bessel functions $K_{\alpha-1}(2\sqrt{\beta\gamma})$ and $K_{\alpha+1}(2\sqrt{\beta\gamma})$, respectively, namely

$$\begin{cases} -\frac{\sqrt{\beta\gamma}}{\alpha}(C_{const} - \alpha) + (\alpha - \beta - G) \sqrt{\frac{\gamma}{\beta}} = 0 \\ \frac{\sqrt{\beta\gamma}}{\alpha}(C_{const} - \alpha) - (\gamma + C) \sqrt{\frac{\beta}{\gamma}} = 0. \end{cases} \tag{25}$$

Next, after canceling mutual terms in both equations of Eq. (25) and equating the rest, it follows

$$(\alpha - \beta - G) \sqrt{\frac{\gamma}{\beta}} = (\gamma + C) \sqrt{\frac{\beta}{\gamma}}. \tag{26}$$

Finally, we derive an equation describing the relationship between the parameters featured in the conditional GIG distribution (α , β and γ) and the rest of the Lagrangian coefficients appearing in front of the energy terms as:

$$G + \frac{\beta}{\gamma} C = \alpha - 2\beta. \tag{27}$$

Therefore, our temporary assumption that the bpJ-maximizing marginal output ISI distribution is GIG is validated because the bpJ-maximizing $f_T(t)$ is the GIG pdf

$$f_T(t) = C_a t^{a-1} \exp\left(-ct - \frac{b}{t}\right), \tag{28}$$

which is indeed a GIG pdf with C_a defined as in Eq. (12) and with $a = D$, $b = L$ and $c = B$.

V. FINDING THE OPTIMUM $f_\Lambda(\lambda)$

Eq. (8) is repeated below as Eq. (29)

$$f_T(t) = \int_0^\infty d\lambda f_{T|\Lambda}(t|\lambda) f_\Lambda(\lambda), \quad \forall t \geq 0. \quad (29)$$

From the previous sections we have

$$f_{T|\Lambda}(t|\lambda) = \lambda \cdot C_\alpha (\lambda t)^{\alpha-1} \exp\left(-\gamma \lambda t - \frac{\beta}{\lambda t}\right); \quad (30)$$

$$f_T(t) = C_a t^{D-1} \exp\left(-Bt - \frac{L}{t}\right), \quad (31)$$

where C_α and C_a are defined as in Eq. (3) and Eq. (12).

In order to obtain $f_\Lambda(\lambda)$, we first need a lemma.

Lemma 1. *Channel 1 and Channel 2 are equivalent in the sense of probabilistically generating the same output r.v. T given the identical input distribution $f_\Lambda(\lambda)$. (The proof of Lemma 1 is omitted due to space limitation.)*

Channel 1: The r.v.'s Λ and T are related by

$$T \stackrel{d}{=} \frac{1}{\Lambda} \cdot U, \quad (32)$$

where U is independent of Λ and has the pdf

$$f_U(u) = C_\alpha u^{\alpha-1} \exp\left(-\gamma u - \frac{\beta}{u}\right), \quad 0 < u < \infty, \quad (33)$$

which is the special case of the conditional pdf $f_{T|\Lambda}(t|\lambda)$ in which $\lambda = 1$ and u plays the role of t . Eq. (32) is equality in distribution, which implies that T and $\frac{U}{\Lambda}$ are identically distributed r.v.'s.

Channel 2: The property characterizing this channel is the conditional pdf

$$f_{T|\Lambda}(t|\lambda) = \lambda \cdot C_\alpha (\lambda t)^{\alpha-1} \exp\left(-\gamma \lambda t - \frac{\beta}{\lambda t}\right), \quad (34)$$

where C_α is defined as in Eq. (3).

According to Lemma 1, by defining $W = \log U$, $V = \log T$ and $Z = -\log \Lambda$, we obtain

$$V = W + Z. \quad (35)$$

Therefore, the characteristic function of Z can be written as:

$$\psi_Z(z) = \frac{C_a}{C_\alpha} \left(\frac{b}{c}\right)^{a-jz} \left(\frac{\gamma}{\beta}\right)^{\alpha-jz} \frac{K_{a-jz}(2\sqrt{bc})}{K_{\alpha-jz}(2\sqrt{\beta\gamma})}, \quad (36)$$

which can be Fourier inverted through numerical calculations allowing us to eventually obtain $f_\Lambda(\lambda)$ numerically.

VI. CONCLUSION

We have shown that, when neuron j employs a GIG conditional distribution as the channel density function and bpJ-maximizing is achieved, the output ISI distribution is a related GIG marginal distribution. This has allowed us to compute the tradeoff between information rate and average power in the GIG class in a markedly simpler way requiring only a single one dimensional integral instead of several multidimensional integrals. Also, we show how to

numerically obtain the associated input distribution $f_\Lambda(\lambda)$. By generalizing from the Gamma and IG families to the GIG family, the derived results contain [19] [20] [21] as special cases in which the three parameters are set to specific values.

ACKNOWLEDGMENT

This research was supported in part by the US National Science Foundation Grant CCF-1162449.

REFERENCES

- [1] J. M. Kinney, H. N. Tucker, Clintec International Inc, *Energy Metabolism: Tissue Determinants and Cellular Corollaries*, 1992.
- [2] L. C. Aiello, P. Wheeler, *The expensive-tissue hypothesis—the brain and the digestive-system in human and primate evolution*, *Curr Anthropol* 36:199-221, 1995.
- [3] A. Hasenstaub, S. Otte, E. Callaway and T. J. Sejnowski, *Metabolic cost as a unifying principle governing neuronal biophysics*, *Proc Natl Acad Sci USA* 107: 12329-12334, 2010.
- [4] D. M. MacKay, W. S. McCulloch, *The limiting information capacity of a neuronal link*, *Bull Math Biophysics*, vol. 14, 127-135, 1952.
- [5] W. S. McCulloch, *An upper bound on the informational capacity of a synapse*, In *Proceedings of the 1952 ACM national meeting*, Pittsburgh, Pennsylvania.
- [6] W. Bialek, F. R. Rieke, R. R. de Ruyter van Steveninck and D. Warland, *Reading a neural code*, *Science*, 252, 1854-7, 1991.
- [7] W. B. Levy, R. A. Baxter, *Energy efficient neural codes*, *Neural Comput* 8:531-543, 1996.
- [8] W. B. Levy, R. A. Baxter, *Energy-efficient neuronal computation via quantal synaptic failures*, *J Neurosci* 22:4746-4755, 2001.
- [9] V. Balasubramanian, D. Kimber, M. J. Berry, *Metabolically efficient information processing*, *Neural Comput* 13:799-815, 2001.
- [10] B. Sengupta, M. Stemmler, S. B. Laughlin and J. E. Niven, *Action potential energy efficiency varies among neuron types in vertebrates and invertebrates*, *PLoS Comput. Biol.* 6, e1000840, 2010.
- [11] J. J. Harris, R. Jolivet, D. Attwell, *Synaptic energy use and supply*, *Neuron* 75, 762-777, 2012.
- [12] J. E. Niven, I. C. Anderson, S. B. Laughlin, *Fly photoreceptors demonstrate energy information trade-offs in neural coding*, *PLoS Biol* 5:e116, 2007.
- [13] V. Balasubramanian, M. J. Berry, *A test of metabolically efficient coding in the retina*, *Network* 13:531-552, 2002.
- [14] S. B. Laughlin, R. R. de Ruyter van Steveninck, J. C. Anderson, *The metabolic cost of neural information*, *Nat Neurosci* 1:36-41, 1998.
- [15] D. Attwell, S. B. Laughlin, *An energy budget for signaling in the grey matter of the brain*, *J Cereb Blood Flow Metab* 21:1133-1145, 2001.
- [16] J. J. Harris, D. Attwell, *The energetics of central nervous system white matter*, *J. Neurosci.* 2, 356-371, 2012.
- [17] B. D. Willmore, J. A. Mazer, J. L. Gallant, *Sparse coding in striate and extrastriate visual cortex*, *J Neurophysiol*, 105:2907-19, 2011.
- [18] S. B. Laughlin, T. J. Sejnowski, *Communication in neuronal networks*, *Science* 301:1870, 2003.
- [19] T. Berger, W. B. Levy, *A mathematical theory of energy efficient neural computation and communication*, *IEEE Trans. IT*, vol. 56, No. 2, pp. 852-874, February 2010.
- [20] J. Xing, T. Berger, T. J. Sejnowski, *A Berger-Levy energy efficient neuron model with unequal synaptic weights*, *IEEE International Symposium on Information Theory Proceedings*, 2964 - 2968, 2012.
- [21] T. Berger, W. B. Levy, J. Xing, J. *Energy efficient neurons with generalized inverse Gaussian interspike interval durations*, 2011 49th Annual Allerton Conference on Communication, Control, and Computing, 1737-1742, 2011.
- [22] O. E. Barndorff-Nielsen, P. Blaesild and C. Halgreen, *First hitting time models for the generalized inverse Gaussian distribution*, *Stochastic Process and Their Applications*, vol. 7, pp. 49-54, 1978.
- [23] O. E. Barndorff-Nielsen and P. Blaesild, *Hyperbolic distributions and ramifications: contributions to the theory and application*, *Statistical Distributions in Scientific Work*. (C. Tallie et al., Eds.) 4 Reidel, Dordrecht, pp. 19-44, 1981.
- [24] G. Mitchison, *Neuronal branching patterns and the economy of cortical wiring*, *Proc Biol Sci* 245:151-158, 1991.
- [25] B. Barbour, N. Brunel, V. Hakim, J. Nadal, *What can we learn from synaptic weight distributions?*, *Trends Neurosci* 30(12): 622-629, 2007.
- [26] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 7th edition, 2007.