The Karush-Kuhn-Tucker conditions for a generalized inverse Gaussian neural model

Mustafa Sungkar
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, Virginia 22904
Email: mhs5mq@virginia.edu

Toby Berger
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, Virginia 22904
Email: tb6n@virginia.edu

William B Levy
Department of Neuroscience and Department of Psychology
University of Virginia
Charlottesville, Virginia 22908
Email: wbl@virginia.edu

***Abstract*—Our view is that the purpose of a neuron is to send information to its targets about the state of the neuron's input. The neuron's input signals contribute to the neuron's excitation called the postsynaptic potential (PSP). Once the PSP hits a threshold, the neuron fires an electrical pulse called the action potential (AP). We subscribe to the time coding viewpoint, i.e., information is encoded in the time interval between two subsequent AP's called the interpulse interval (IPI). We use Shannons mutual information as a measure of neural information between the input intensity and the duration of the output IPI's. Due to a neuron's remarkable energy-efficiency, we assume that the neuron maximizes the information between its input and output for a given mean energy budget.**

The PSP buildup of the neuron can be modeled by the generalized inverse Gaussian (GIG) diffusion, which is a generalization of the Wiener process with constant drift. The first hitting time (FHT) distribution of the GIG diffusion is known as the GIG distribution. A GIG neuron is one whose FHT of the threshold has a GIG distribution. The information-energy relationship of such a neural model has been much studied by our research group. However, for certain values of the parameters in the GIG distribution and energy terms, although the resulting marginal distribution of the IPI duration is another GIG distribution, the information-maximizing input distribution in a Lagrangian optimization setting results in an input distribution that is negative at certain points. This implies that some portions of the information-energy curve act as an upper bound to the maximum information. This paper begins to address that problem. The Karush-Kuhn-Tucker (KKT) conditions are used to determine the point of minimum average energy and the slope of the information-energy curve at that point. The implication is that a tighter upper bound of the information-energy curve and the point of maximum bits per joule can be obtained.

***Index Terms*—Neurons, Postsynaptic potential (PSP), generalized inverse Gaussian (GIG) diffusion, Karush-Kuhn-Tucker conditions, Shannon capacity curve, mutual information.**

I. Introduction

It is widely believed that the function of a neuron is to send to its targets information about the state of its input. The neuron can receive inputs from sensory cells or other neurons, whereas its targets could be muscle cells or other neurons. Our neuron of interest is a pyramidal neuron in the primary sensory cortex. We refer to this neuron as neuron j, or just j when that is unambiguous. Neuron j receives input from ca. ten thousand other neurons and sends this information to roughly the same number of targets. The set of neuron's whose output reach j is called j's afferent cohort.

We use Shannon's mutual information [1] as a measure of neural information. Shannon's mutual information is an intuitive measure of "information" and has many theories surrounding it. Other measures of neural information is possible, but we believe that Shannon's mutual information is appropriate for our purposes. Furthermore, Shannon's mutual information has been used and studied in the context of neuroscience [2]–[6].

Neurons are remarkably energy efficient and are widely believed to minimize the amount of energy needed to function [7]–[12]. Thus, we assume that j minimizes the energy needed to send a certain number of bits of information. However, this is also equivalent to maximizing the amount of bits sent given a fixed amount of energy. We adopt the latter point of view, which is more consistent with traditional information theory. Often, information theorists are concerned with finding the constrained capacity of a channel, which is the maximum mutual information that can be conveyed given a restriction on the channel use [13].

A. Neural Function

Neuron j is composed of three parts: the dendrite, the soma, and the axon. The dendrite acts as the input line and carries the input to the soma. In the soma, the input is integrated into the neuron's excitation called the postsynaptic potential (PSP). The axon acts as the output line, carrying j's output to all of its targets. Neuron j's dendrite forms a connection called a synapse with the axon of each neuron in j's afferent cohort. The synapse can be either excitatory or inhibitory. Excitatory synapses increase the PSP while the inhibitory synapses decrease the PSP or cancel excitatory signals. The PSP builds up over time and once it reaches a threshold, j emits a pulse called an action potential (AP) along the axon to its targets. An AP at a synapse elicits a pulse in the target dendrite, which becomes an input to the neuron. We refer the readers to [14] for a more comprehensive introduction to basic neural function. We subscribe to the timing code viewpoint: the output information is encoded in the time duration between adjacent AP's. We call this time duration the interpulse interval

(IPI). We model the average input intensity and duration of the kth IPI as the random variables Λ_k and T_k, respectively. We will assume that each IPI is independent of any other[1]. We also assume that the behavior of the neuron does not change in the observed time interval. Hence, we drop the k subscript and refer to the corresponding random variables as Λ and T in an arbitrary IPI.

II. The GIG Neuron Model

Since the neurons in j's afferent cohort are numerous and each contribute a small amount to the PSP, a time-homogeneous stochastic diffusion can be used to model the behavior of the PSP as it approaches the threshold. A time-homogeneous stochastic diffusion Y_t can be described a stochastic differential equation:

$$\mathrm{d}Y_t = d(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t, \tag{1}$$

where $\mu(y)$ is called the drift and $\sigma(y)$ is called the infinitesimal variance. The random process $\{W_t\}$ is the standard Wiener process. The first of the time-homogeneous stochastic diffusion model was proposed in [16], where the diffusion was the Wiener process with constant drift. Others include the Ornstein-Uhlenbeck, the Feller, and processes based on biophysical models of the neuron [17]–[28]. However, we adopt the GIG diffusion model as our neural model [29].

The PSP of the GIG neuron is modeled as a GIG diffusion given by a constant infinitesimal variance σ^2 and a drift given by [29], [30]

$$d(y) = \sigma^2 \frac{2\alpha - 1}{2(\theta - y)} + \sigma \frac{\sqrt{2\gamma} K_{\alpha-1}(\frac{\theta-y}{\sigma}\sqrt{2\gamma})}{K_\alpha(\frac{\theta-y}{\sigma}\sqrt{2\gamma})}, \tag{2}$$

where $K_\alpha(\cdot)$ is the modified Bessel function of the second kind of order α. The parameters α controls the attraction to the threshold, θ is the threshold, and γ controls a constant drift component [31]. The parameters satisfy $\alpha \leq 0$ and $\gamma > 0$.

We choose the GIG model because it models the upswing of the PSP as it approaches the threshold elicited by the sodium channels in the neural membrane. Furthermore, the first hitting time (FHT) distribution of the GIG diffusion is known, allowing us to evaluate the model analytically.

The FHT of the GIG diffusion is known as the generalized inverse Gaussian (GIG) distribution [30]. For our neural model, given a fixed input intensity $\Lambda = \lambda$, the FHT is described by the following GIG distribution

$$Q(t|\lambda) = \frac{\lambda^\alpha t^{\alpha-1} \exp(-\beta/\lambda t - \gamma\lambda t)}{2(\beta/\gamma)^{\alpha/2} K_\alpha(2\sqrt{\beta\gamma})}, \quad \lambda, t > 0. \tag{3}$$

The parameter

$$\beta = \frac{\theta^2}{2\sigma^2}. \tag{4}$$

The neuron can serve as our "channel", where the input intensity Λ is converted into an output IPI duration T. In fact, this relationship is described by a multiplicative noise channel given by

$$T = \frac{U}{\Lambda}, \quad U \perp\!\!\!\perp \Lambda, \tag{5}$$

where U is distributed as a GIG distribution.

A. Energy Costs

There are energy costs for the input $\Lambda = \lambda$ and output $T = t$ given by the following components:

1) $g_1(\lambda, t) = A$, where $A > 0$. This term is associated with fixed energy costs.
2) $g_2(\lambda, t) = Bt$, where $B > 0$. This term is associated with energy costs that vary linearly in time.
3) $g_3(\lambda, t) = L/t$, where $L > 0$. This term increases the energy of two adjacent AP's that are close in time.
4) $g_4(\lambda, t) = -D\log(t)$, where $D \in \mathbb{R}$. This term is associated with maintaining neural "clocks".
5) $g_5(\lambda, t) = G\lambda t$, where $G > 0$. This is the cost of processing incoming spikes.

The total energy cost $g(\lambda, t)$ is given by the sum of its components

$$g(\lambda, t) = A + Bt + L/t - D\log(t) + G\lambda t. \tag{6}$$

III. Maximizing Energy-Constrained Information Neuron

As mentioned, the neuron is remarkably energy-efficient and we assume it maximizes its the mutual information between Λ and T given an average energy budget. In essence, the optimal neuron is found by solving this problem

$$\begin{aligned} \underset{\{p(\lambda)\}}{\text{maximize}} \quad & I(\Lambda; T) \\ \text{subject to} \quad & \mathbb{E}[g(\Lambda, T)] = \mathcal{E} \\ & \int_0^\infty p(\lambda)\mathrm{d}\lambda = 1 \\ & p(\lambda) \geq 0, \ \lambda \geq 0. \end{aligned} \tag{7}$$

We have solved a relaxed version of the problem where the inequality constraint $p(\lambda) \geq 0$ is removed [32]. The solution for the input distribution can be calculated by the integral

$$p(\lambda) = \frac{1}{2\pi\lambda} \frac{K_\alpha(2\sqrt{\beta\gamma})}{K_a(2\sqrt{bc})} \times \int_{-\infty}^{\infty} \left(\lambda\sqrt{\frac{\gamma L}{\beta B}}\right)^{jx} \frac{K_{a+jx}(2\sqrt{bc})}{K_{\alpha+jx}(2\sqrt{\beta\gamma})}\mathrm{d}x. \tag{8}$$

However, it turns out that the resulting $p(\lambda)$ can fall below 0 for certain parameter sets. Thus the relaxation of (7) yielded a non-viable solution.

Thus, we have not actually found the mutual information, but rather an upper bound to it. This resulted in an information energy relationship given by

$$I(\Lambda; T) \leq s\mathcal{E}_s + \mu_s, \tag{9}$$

where s is the derivative of $I(\Lambda; T)$ with respect to $\mathcal{E}$ and μ is a constant for a given s. We have subscripted $\mathcal{E}$ and μ with

[1]It has been shown that the IPI's are not independent. A short IPI is more likely to be followed by a long IPI. However, it has been shown that the information decrement from the correlation of the IPI is small [15].

s because they vary with s. In terms of the parameters and s, the information can be written as [33]

$$I(\Lambda;T) \leq \log \frac{K_a(2\sqrt{bc})}{K_\alpha(2\sqrt{\beta\gamma})} + \frac{\mathrm{d}}{\mathrm{d}x}\left[\log\left(\frac{K_{\alpha x}(2x\sqrt{\beta\gamma})}{K_{ax}(2x\sqrt{bc})}\right)\right]_{x=1} \quad (10)$$

where $a = sD$, $b = sL$, and $c = sB$.

A. The Marginal of Λ for $\gamma = 0$

Here, we show that for $\gamma = 0$, the resulting $p(\lambda)$ fails to meet the inequality constraint. When $\gamma = 0$, the constant drift component is non-existent and only the attraction to the threshold propels the diffusion upward. The conditional distribution in this case is given by the inverse gamma distribution

$$Q(t|\lambda) = \frac{\beta^{-\alpha}\lambda^\alpha t^{\alpha-1}\exp(-\frac{\beta}{\lambda t})}{\Gamma(-\alpha)}, \quad (11)$$

where the limit of K_α is used. The optimality condition can be obtained via calculus of variations:

$$I(\Lambda = \lambda;T) = s\mathbb{E}[g(\Lambda,T)|\Lambda=\lambda] + \mu + \nu(\lambda, \quad (12)$$

where

$$I(\Lambda=\lambda) = \mathbb{E}[\log(Q(\Lambda|T)/q(t))] \quad (13)$$

is the information conditioned on $\Lambda = \lambda$. The distribution $q(t)$ is the marginal of T and is given by $q(t) = \int_0^\infty Q(t|\lambda)p(\lambda)\mathrm{d}\lambda$. It turns out that the marginal of T is the same regardless of $Q(t|\lambda)$ and is given by the GIG distribution

$$q(t) = \frac{t^{a-1}\exp(-c/t - bt)}{2(b/c)^{a/2}K_a(2\sqrt{bc})}. \quad (14)$$

We make the substitution $\omega = (\beta/\lambda)^{-1}$ and $\tau = t^{-1}$. Let $r(\omega) = \omega^{-2}p(\beta/\omega)$, i.e., the pdf of Λ^{-1}. We can write the following equation:

$$\int_0^\infty \frac{\omega^{-\alpha}\tau^{-\alpha+1}\exp(-\omega\tau)}{\Gamma(-\alpha)} r(\omega)\mathrm{d}\omega = q(\tau^{-1}). \quad (15)$$

Note that the left hand side is in a form of a Laplace transform. Hence, we can move all terms of τ to the right-hand side, take the inverse transform and get [34]

$$r(\omega) = C_3^{-1}\omega^\alpha(\omega-b)^{\frac{a-\alpha-1}{2}}J_{a-\alpha-1}(2\sqrt{c(\omega-b)}),\ \omega \geq b, \quad (16)$$

where J_a is the Bessel function of the first kind of order a and

$$C_3 = \frac{2b^{a/2}c^{(-\alpha-1)/2}K_a(2\sqrt{bc})}{\Gamma(\alpha)}. \quad (17)$$

However, we see that $r(\omega) < 0$ for some values of ω because J_a falls below zero. Hence, $r(\omega)$ is not a pdf, which implies that $p(\lambda)$ is not a pdf. The relaxed version of (7) leads to an answer we cannot use. Therefore, we cannot remove the inequality constraint when solving (7) in general.

IV. Point of Least Energy

In this section, we solve for the point on the information-energy curve with $I(\Lambda;T) = 0$. We call this the point of least energy, where this is the amount of energy j must use in a given IPI without even trying to convey any information. First we recast (7) as a minimization problem: we seek the input distribution that minimizes energy for fixed mutual information. The problem can be written as

$$\begin{aligned} \underset{\{p(\lambda)\}}{\text{minimize}} \quad & \mathbb{E}[g(\Lambda,T)] \\ \text{subject to} \quad & I(\Lambda;T) = \mathcal{I} \\ & \int_0^\infty p(\lambda)\mathrm{d}\lambda = 1 \\ & p(\lambda) \geq 0,\ \lambda \geq 0. \end{aligned} \quad (18)$$

We choose the value of $\mathcal{I}$ to be 0. The information $I(\Lambda;T) = 0$ if $Q(t|\lambda) = p(t)$ for all values of λ such that $p(\lambda) > 0$. This is true if $p(\lambda) = \delta(\lambda - \lambda^*)$ for some λ^*, where δ is the Dirac-delta function. Furthermore, we rewrite the energy function in terms of the input. Define $\bar{g}(\lambda)$ as

$$\bar{g}(\lambda) = \mathbb{E}[g(\Lambda;T)|\Lambda=\lambda]. \quad (19)$$

For (6), this is

$$\bar{g}(\lambda) = \bar{A} + \bar{L}\lambda + \bar{B}/\lambda + \bar{D}\log(\lambda), \quad (20)$$

where $\bar{A} = A + k_{\text{lin}}G - k_{\log}D$, $\bar{B} = k_{\text{lin}}B$, $\bar{L} = k_{\text{inv}}L$, and $\bar{D} = D$. The constants k_{lin}, k_{inv}, and $k_{\log}$ are defined as

$$k_{\text{lin}} = \mathbb{E}[\Lambda T|\Lambda] = \sqrt{\frac{\beta}{\gamma}}\frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})}, \quad (21)$$

$$k_{\text{inv}} = \mathbb{E}\left[\frac{1}{\Lambda T}\middle|\Lambda\right] = \sqrt{\frac{\gamma}{\beta}}\frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_\alpha(2\sqrt{\beta\gamma})}, \quad (22)$$

and

$$k_{\log} = \mathbb{E}[\log\Lambda T|\Lambda] = \frac{1}{2}\log\frac{\beta}{\gamma} + \frac{\frac{\partial}{\partial\alpha}[K_\alpha(2\sqrt{\beta\gamma})]}{K_\alpha(2\sqrt{\beta\gamma})}, \quad (23)$$

respectively.

Now the minimization problem is simply to find the value of λ^* that minimizes $\bar{g}(\lambda)$. This is can be obtained by obtaining the critical point of $\bar{g}$ via its derivative and setting it equal to zero:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\bar{g}(\lambda) = \bar{L} - \frac{\bar{B}}{\lambda^2} + \frac{\bar{D}}{\lambda} = 0 \quad (24)$$

The quadratic formula states that there are two solutions to this equation, but since $\lambda \geq 0$, we only choose the one that is non-negative, which is

$$\lambda^* = \frac{-\bar{D} + \sqrt{\bar{D}^2 + 4\bar{B}\bar{L}}}{2\bar{L}}. \quad (25)$$

We also observe that the second derivative of $\bar{g}$ at λ^* is

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\bar{g}|_{\lambda=\lambda^*} = \frac{2\bar{B}}{(\lambda^*)^3} - \frac{\bar{D}}{(\lambda^*)^2} = \frac{\bar{B}}{(\lambda^*)^3} + \frac{\bar{L}}{\lambda^*}, \quad (26)$$

which is always positive for $\lambda^* > 0$. Hence, the critical point is a minimum.

The marginal of T is then given by

$$q(t) = Q(t|\lambda^*) = \frac{(\lambda^*)^\alpha t^{\alpha-1}\exp(-\beta/\lambda^* t - \gamma\lambda^* t)}{2(\beta/\gamma)^{\alpha/2}K_\alpha(2\sqrt{\beta\gamma})}, t > 0, \tag{27}$$

which is another GIG distribution. The minimum energy $\mathcal{E}^*$ is given by $\bar{g}(\lambda^*)$, or equivalently, $\mathbb{E}[g(\Lambda, T)|\Lambda = \lambda^*]$.

V. Initial Slope of the Information-Energy Curve

Of interest is also the slope of the information-energy curve at $\mathcal{E} = \mathcal{E}^*$. The slope can be obtained from the Karush-Kuhn-Tucker conditions. The KKT conditions are the conditions that optimization problems with inequality and equality constraints must satisfy. For (7), the KKT conditions are

$$\begin{aligned} &I(\Lambda = \lambda; T) = s\bar{g}(\lambda) + \mu + \nu(\lambda) \\ &\nu(\lambda) \leq 0 \\ &\nu(\lambda) f_\Lambda(\lambda) = 0, \end{aligned} \tag{28}$$

and including the original equality constraints in (7). The value of s is still the rate of change of information with respect to $\mathcal{E}$. From section IV, the marginals of Λ and T for $\mathcal{E} = \mathcal{E}^*$ are known. We can use this to determine the value of s, μ, and $\nu(\lambda)$ for $\mathcal{E} = \mathcal{E}^*$. Now we substitute in the expression for $I(\Lambda = \lambda; T)$ and $\bar{g}(\lambda)$,

$$\begin{aligned} \mathbb{E}\left[\log\left(\frac{\Lambda^\alpha \exp(-\frac{\beta}{\Lambda T} - \gamma\Lambda T)}{(\lambda^*)^\alpha \exp(-\frac{\beta}{\lambda^* T} - \gamma\lambda^* T)}\right)\middle|\Lambda = \lambda\right] = \\ s\left(\bar{A} + \bar{L}\lambda + \frac{\bar{B}}{\lambda} + \bar{D}\log(\lambda)\right) + \mu + \nu(\lambda) \end{aligned} \tag{29}$$

Carrying out all of the expectations yields

$$\begin{aligned} \alpha\log(\lambda) + \frac{\beta k_{\text{inv}}}{\lambda^*}\lambda + \gamma k_{\text{lin}}\lambda^*\lambda^{-1} + Y = \\ s\bar{D}\log(\lambda) + s\bar{L}\lambda + s\bar{B}\lambda^{-1} + s\bar{A} + \mu + \nu(\lambda), \end{aligned} \tag{30}$$

where $Y = -\beta k_{\text{inv}} - \gamma k_{\text{lin}} - \alpha\log(\lambda^*)$. The left-hand and right-hand sides must equal. Therefore, the functions ν and μ must make them equal. Hence, we let ν be

$$\nu(\lambda) = -s\hat{L}\lambda - s\frac{\hat{B}}{\lambda} - s\hat{D}\log(\lambda) + k, \tag{31}$$

where k is chosen so that $\nu(\lambda) \leq 0$. Thus, $\hat{B}$, $\hat{L}$, and $\hat{D}$ compensate for $\bar{B}$, $\bar{L}$, and $\bar{D}$ so that the λ terms hold. Hence, we can create a system of equations so that the λ terms match up

$$\begin{aligned} &s(\bar{D} - \hat{D}) = \alpha, \\ &s(\bar{L} - \hat{L}) = \frac{\beta k_{\text{inv}}}{\lambda^*}, \\ &s(\bar{B} - \hat{B}) = \gamma k_{\text{lin}}\lambda^*, \\ &-s\hat{L} + s\frac{\hat{B}}{(\lambda^*)^2} - s\frac{\hat{D}}{\lambda^*} = 0. \end{aligned} \tag{32}$$

The last equation is necessary in order for the critical point of $\nu(\lambda)$ occurs at λ^*. The value of $\nu(\lambda)$ must be 0 for $\lambda = \lambda^*$ and negative everywhere else in order to satisfy the KKT conditions. Thus, the maximum must occur at λ^*. For the critical point to be a maximum, it is sufficient to have

$$\begin{aligned} &\hat{B} > 0, \hat{L} > 0, \text{ if } \hat{D} = 0, \\ &\hat{B} \geq 0, \hat{L} > 0, \text{ if } \hat{D} < 0, \\ &\hat{B} > 0, \hat{L} \geq 0, \text{ if } \hat{D} > 0. \end{aligned} \tag{33}$$

It turns out that only three out of the four equations in (32) are independent. Hence, we fix s and solve the system using only the first three equations, which yields

$$\begin{aligned} \hat{D} &= \bar{D} - \frac{\alpha}{s}; \\ \hat{L} &= \bar{L} - \frac{\beta k_{\text{inv}}}{\lambda^* s}; \\ \hat{B} &= \bar{B} - \frac{\gamma k_{\text{lin}}\lambda^*}{s}, \end{aligned} \tag{34}$$

Due to (33), it must be that

$$s \geq \max\left(\frac{\beta k_{\text{inv}}}{\lambda^*\bar{L}}, \frac{\gamma k_{\text{lin}}\lambda^*}{\bar{B}}\right). \tag{35}$$

The slope of the tangent at $\mathcal{E} = \mathcal{E}_0$ is the smallest allowable value of s. We call this value s_0 and its value is

$$s_0 = \max\left(\frac{\beta k_{\text{inv}}}{\lambda^*\bar{L}}, \frac{\gamma k_{\text{lin}}\lambda^*}{\bar{B}}\right). \tag{36}$$

To determine the value of k, recall that the maximum of ν must be 0. Since this maximum occurs at $\lambda = \lambda^*$, the value of k must be

$$k = s\hat{L}\lambda^* + s\frac{\hat{B}}{\lambda^*} + s\hat{D}\log(\lambda^*). \tag{37}$$

The value of μ must make the rest of the terms in (30) equal. Therefore, its value is

$$\mu = Y - s\bar{A} - k. \tag{38}$$

We have forced the KKT conditions to be true for all λ and thus have found the initial slope of the information-energy curve.

VI. Discussion

Two examples of the upper bound of the information-energy curve with the point of least energy and its initial slope are shown in Figure 1. The point of least energy should occur below the upper bound. However, the gap between this point and the upper bound depends on the parameter set. We empirically found that using smaller values of α, β, and γ leads to a larger gap. However, the exact relationship between the gap and the parameter set is an open question.

Using the point of least energy and the initial slope, we can improve the upper bound. Since the information-energy curve is convex and non-decreasing [35], $I(\Lambda; T)$ and $\mathcal{E}_s$ is a decreasing function of the slope s. In other words, as energy increases, information increases and the slope of the curve decreases. Hence, the information-energy curve cannot exceed the line of slope $s0$ at the point of least energy (Figure 1). So this is an improved upper bound on the information-energy curve. Furthermore, if the line intersects the upper bound at

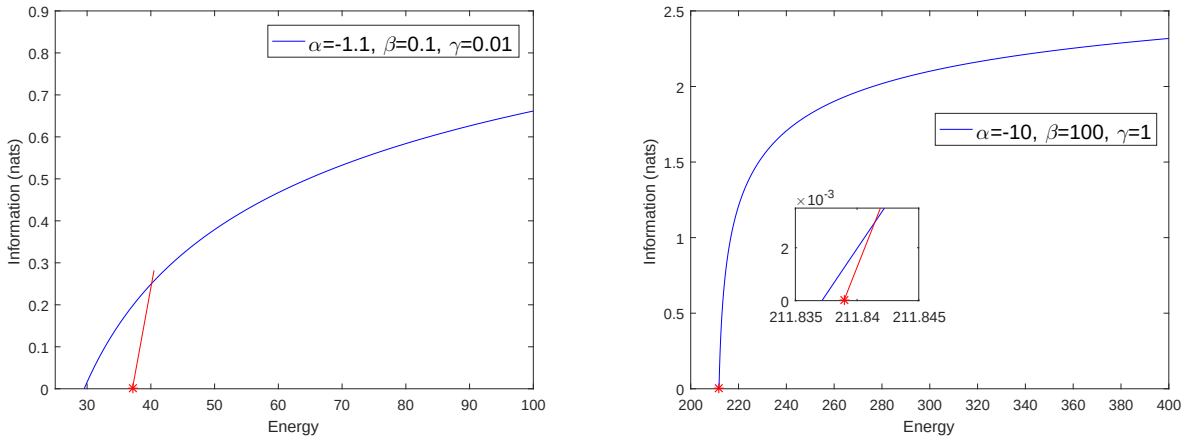

Fig. 1. We show the upper bound for two information-energy curves (blue). In each plot, we show its respective point of least energy (red asterisk), and a line with the initial slope of the actual information-enrgy curve (red). For the plot on the left, $A = 1$, $B = 10$, $L = 5$, $G = 5$, and $D = 5$. For the plot on the right, $A = 50$, $B = 25$, $L = 10$, $G = 20$, and $D = 5$. The inset in the right plot is the information-energy curve with smaller range for the axes. The gap between the point of least energy and its upper bound is 0.197 nats for the plot on the left and 0.00122 nats for the plot on the right

some point after the maximum bits per Joule point, then we also have an improved upper bound on the maximum bits per Joule. Whether this will actually occur remains to be proven.

In all of the examples we examined, the initial slope $s0$ was greater than the slope of the upper bound at the corresponding energy $\mathcal{E}^*$. If we can show that the slope of the actual curve is greater than the slope of the upper bound, then we can bound the actual information-energy curve to within the gap at the point of least energy. This is in essence a lower bound on the information-energy curve. However, whether this fact is true or not is an open question.

VII. Conclusion

Previously, we found an upper bound of the information-energy curve of the GIG neuron. Here, we found one point on the actual curve. This point happens to be the point of least energy. We also found the initial slope of the information-energy curve. Using the point of least energy and the initial slope we have an improved upper bound. A possible next step is to determine whether the slope of the actual curve is greater than the slope of the upper bound at every energy value. If this is true, then we have found a lower bound on the information energy curve.

Acknowledgements

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

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[2] D. M. MacKay and W. S. McCulloch, "The limiting information capacity of a neuronal link," *The Bulletin of Mathematical Biophysics*, vol. 14, no. 2, pp. 127–135, 1952. [Online]. Available: http://dx.doi.org/10.1007/BF02477711

[3] W. S. McCulloch, "An upper bound on the informational capacity of a synapse," in *Proceedings of the 1952 ACM National Meeting (Pittsburgh)*, ser. ACM '52. New York, NY, USA: ACM, 1952, pp. 113–117. [Online]. Available: http://doi.acm.org/10.1145/609784.609799

[4] W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck, and D. Warland, "Reading a neural code," *Science*, vol. 252, no. 5014, pp. pp. 1854–1857, 1991. [Online]. Available: http://www.jstor.org/stable/2875885

[5] L. Kostal, "Information capacity in the weak-signal approximation," *Phys. Rev. E*, vol. 82, p. 026115, Aug 2010. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.82.026115

[6] P. Sturges, "Avances de la neurociencia y sus implicaciones para la ciencia de la informacin. (spanish)." *El Profesional de la Informacin*, vol. 24, no. 2, pp. 168 – 175, 2015, abstract. [Online]. Available: http://www.proxy.its.virginia.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=zbh&AN=101662224&site=ehost-live&scope=site

[7] W. B. Levy and R. A. Baxter, "Energy efficient neural codes," *Neural Computation*, vol. 8, no. 3, pp. 531–543, April 1996.

[8] ——, "Energy-efficient neuronal computation via quantal synaptic failures," *The Journal of Neuroscience*, vol. 22, no. 1, pp. 4746–4755, June 2002.

[9] J. E. Niven and S. B. Laughlin, "Energy limitation as a selective pressure on the evolution of sensory systems," *Journal of Experimental Biology*, vol. 211, no. 11, pp. 1792–1804, jun 2008. [Online]. Available: http://dx.doi.org/10.1242/jeb.017574

[10] A. Hasenstaub, S. Otte, E. Callaway, and T. J. Sejnowski, "Metabolic cost as a unifying principle governing neuronal biophysics," *Proceedings of the National Academy of Sciences*, vol. 107, no. 27, pp. 12 329–12 334, jun 2010. [Online]. Available: http://dx.doi.org/10.1073/pnas.0914886107

[11] J. E. Niven, M. Vahasoyrinki, and M. Juusola, "Shaker k+-channels are predicted to reduce the metabolic cost of neural information in drosophila photoreceptors," *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, no. Suppl_1, pp. S58–S61, aug 2003. [Online]. Available: http://dx.doi.org/10.1098/rsbl.2003.0010

[12] V. Balasubramanian, D. Kimber, and M. J. Berry, "Metabolically efficient information processing," *Neural Computation*, vol. 13, no. 4, pp. 799–815, April 2001.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, N.J.: John Wiley and Sons, Inc., 2006.

[14] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of Neural Science*, 4th ed. McGraw-Hill, 2000.

[15] T. Berger and W. B. Levy, "A mathematical theory of energy efficient neural computation and communication," *IEEE Transactions on Information Theory*, vol. 56, no. 2, pp. 852–874, 2010.

[16] G. L. Gerstein and B. Mandelbrot, "Random walk models for the spike activity of a single neuron," *Biophysical journal*, vol. 4, no. 1 Pt 1, p. 41, 1964.

[17] P. I. M. Johannesma, *Neural Networks: Proceedings of the -School on Neural Networks - June 1967 in Ravello*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1968, ch. Diffusion Models for the

Stochastic Activity of Neurons, pp. 116–144. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-87596-0_11

[18] R. M. Capocelli and L. M. Ricciardi, "Diffusion approximation and first passage time problem for a model neuron," *Kybernetik*, vol. 8, no. 6, pp. 214–223, 1971. [Online]. Available: http://dx.doi.org/10.1007/BF00288750

[19] L. M. Ricciardi and L. Sacerdote, "The ornstein-uhlenbeck process as a model for neuronal activity," *Biological Cybernetics*, vol. 35, no. 1, pp. 1–9, 1979. [Online]. Available: http://dx.doi.org/10.1007/BF01845839

[20] H. C. Tuckwell and D. K. Cope, "Accuracy of neuronal inerspike times calculated from a diffusion approximation," *Journal of Theoretical Biology*, vol. 83, no. 3, pp. 377–387, 1980.

[21] B. Lindner and A. Longtin, "Effect of an exponentially decaying threshold on the firing statistics of a stochastic integrate-and-fire neuron," *Journal of Theoretical Biology*, vol. 232, no. 4, pp. 505 – 521, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022519304004187

[22] P. Lansky and J. P. Rospars, "Ornstein-uhlenbeck model neuron revisited," *Biological Cybernetics*, vol. 72, no. 5, pp. 397–406, 1995. [Online]. Available: http://dx.doi.org/10.1007/BF00201415

[23] F. B. Hanson and H. C. Tuckwell, "Diffusion approximations for neuronal activity including synaptic reversal potentials," *Journal of Theoretical Neurobiology*, vol. 2, pp. 127–153, 1983.

[24] P. Lansky and V. Lanska, "Diffusion approximation of the neuronal model with synaptic reversal potentials," *Biological Cybernetics*, vol. 56, no. 1, pp. 19–26, 1987. [Online]. Available: http://dx.doi.org/10.1007/BF00333064

[25] E. Bibbona, P. Lansky, and R. Sirovich, "Estimating input parameters from intracellular recordings in the feller neuronal model," *Physical Review E*, vol. 81, p. 031916, 2010.

[26] R. Rodriguez and H. C. Tuckwell, "A dynamical system for the approximate moments of nonlinear stochastic models of spiking neurons and networks," *Mathematical and computer modelling*, vol. 31, no. 4, pp. 175–180, 2000.

[27] H. C. Tuckwell, R. Rodriguez, and F. Y. Wan, "Determination of firing times for the stochastic fitzhugh-nagumo neuronal model," *Neural computation*, vol. 15, no. 1, pp. 143–159, 2003.

[28] H. C. Tuckwell and F. Y. Wan, "Time to first spike in stochastic hodgkinhuxley systems," *Physica A: Statistical Mechanics and its Applications*, vol. 351, no. 24, pp. 427 – 438, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437104015353

[29] M. Sungkar, "Energy-efficient cortical neuron models with gig likelihood functions," Master's thesis, University of Virginia, 2015.

[30] O. Barndorff-Nielsen, P. Blaesild, and C. Halgreen, "First hitting time models for the generalized inverse gaussian distribution," *Stochastic Processes and their Applications*, vol. 7, no. 1, pp. 49 – 54, 1978.

[31] M. Sungkar, T. Berger, and W. B. Levy, "Mutual information and parameter estimation in the generalized inverse gaussian diffusion model of cortical neurons," *IEEE Transactions on Molecular, Biological and Multi-scale Communciations*, In Review.

[32] J. Xing, T. Berger, M. Sungkar, and W. B. Levy, "Energy efficient neurons with generalized inverse gaussian conditional and marginal hitting times," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4390–4398, 2015.

[33] M. Sungkar and T. Berger, "Parameter estimation in the Barndorff-Nielsen diffusion model of cortical neurons," in *International Symposium on Information Theory*, 2016.

[34] F. Oberhettinger and L. Badii, *Tables of Laplace Transforms*. New York: Springer-Verlag, 1973.

[35] A. A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.