

Energy-Efficient Cortical Neuron Models with GIG Likelihood Functions

A Thesis

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Master of Science (Electrical Engineering)

by

Mustafa Sungkar

April 2015

Abstract

Neurons are remarkably efficient. The human brain consumes 20 percent of the rest metabolic rate (RMR), but expends energy at a rate of just 25 watts. This efficiency can be explained by natural selection having forced organisms to develop more efficient neural structures, especially during trying times. Hence we have reasons to believe that neurons' energy usage is highly optimized. We hypothesize that a neuron optimizes the mutual information between its inputs and output, given a fixed energy budget.

We propose the generalized inverse Gaussian (GIG) distribution as the pdf of the random IPI given an input excitation intensity. A strong reason is the GIG distribution is the hitting time of the Barndorff-Nielsen (BN) diffusion, which exhibits attraction towards a threshold. Biological data reveal that the rate of PSP buildup increases as the PSP approaches the threshold, hence can be modeled by the BN diffusion.

Using the GIG model, the optimal input and output marginal distributions are obtained. In a given IPI, let the input intensity be Λ and output IPI be T . The energy costs are the sum of the following: a constant term and terms proportional to T , T^{-1} , $\log(T)$, and ΛT . The source of the energy terms are discussed. Under these assumptions, the marginal output distribution is also GIG with parameters related to the energy costs and the conditional GIG. The input distribution is determined as the inverse Fourier transform of an expression involving modified Bessel functions of the second kind; a procedure for numerically obtaining the distribution is described. The information per IPI is plotted against average energy expended. The result is a concave curve of bits vs. energy analogous to the familiar curves of channel capacity vs. constrained input power in classical information theory where information increases with energy but with diminishing return. A point of interest on the curve is the point with maximum information per energy cost.

This neural model can be viewed as a channel with multiplicative noise that is independent of its input. Accordingly, possible connections between the neuron model and fading channels in communications are discussed. Also, the optimization condition can be generalized to other channels and is discussed.

Acknowledgements

I would like to thank the NSF for the partial support of this research through grant CCF-1162449.

I would like to express my gratitude to my adviser, Professor Toby Berger, for introducing me to the world of information theory, with all of its elegance and beauty. I hope that his love and passion for the subject will inspire me on my journey.

I would like to thank Professor William Levy for invaluable discussions on the wonders of neurons. I would also like to thank Professor Wilson for being my Master's committee chair and his feedback on the thesis.

A special thank you goes to Jie Xing, for showing me the ropes and the discussions on research.

Last but certainly not least, my deepest gratitude and appreciation goes to my parents, who have supported me all my life through its ups and downs. Much love to them.

Contents

Abstract	i
Acknowledgements	ii
List of Figures	v
Glossary of Terms	vii
1 Introduction	1
1.1 Energy Efficiency in Neurons	1
1.2 Information Theory in Neuroscience	1
1.3 Motivation	2
2 Background	3
2.1 Parts of a Neuron	3
2.2 Action Potential	4
2.2.1 Action Potential Generation Mechanism	5
2.2.2 Jitter	6
3 Problem Formulation	7
3.1 Defining the Variables	7
3.2 Information Rate	8
3.3 The Problem Statement	8
3.4 Modeling the PSP buildup	9
3.5 Modeling the Energy Cost	11
4 Results	13
4.1 Optimality Condition	13
4.2 The $f_T(t)$ Marginal	13
4.3 The $f_\Lambda(\lambda)$ Marginal	15
4.4 Information-Energy Function	18
5 Discussion	21
5.1 Relationship to Fading	22

6 Conclusion	23
A Optimal Condition Derivation	24
Bibliography	27

List of Figures

2.1	The three parts of a neuron.	4
2.2	An AP measured and simulated with the Hodgkin-Huxley equations. The top plots are the simulated result. The bottom plots are the measured results. In the two plots on the bottom, the line above the measured the AP is the curvature of the window screen of the measuring apparatus used by Hodgkin and Huxley [1].	5
2.3	The IPI is the interval between two AP's. The reference point is the point of highest slope in the AP. The AP's were simulated using the equations in [1].	5
3.1	Three Barndorff-Nielsen diffusions with threshold $\theta = 100$ and a fixed variance. All parameters are kept the same except for α . The lower the value of α , the stronger the attraction towards the threshold. The black curve is also a Wiener process with a drift rate of 10 and a variance of 10.	10
3.2	Three Barndorff-Nielsen diffusions with threshold $\theta = 100$ and a fixed variance. The blue curve is also a Wiener process with a drift rate of 7.5 and variance of 1.	10
3.3	Three GIG pdfs corresponding with the diffusions in figure 3.2	11
4.1	The marginal densities for Λ at different values of μ_1 for $\alpha = -10$, $\beta = 250$, $\gamma = 0.75$, $A = 100$, $B = 25$, $C = 23$, $D = 3$ and $L = 20$. The values are (A) $\mu_1 = 0.001$, (B) $\mu_1 = 0.01$, and (C) $\mu_1 = 0.1$. The values of μ_1 depends on the the energy J . The energies are (A) $J = 710$, (B) $J = 484$, and (C) $J = 450$. The unit for energy is arbitrary. We have removed points near $\lambda = 0$ since the error at those points are amplified.	17
4.2	The marginal densities for Λ at different values of μ_1 for $\alpha = -10$, $\beta = 166.67$, $\gamma = 1$, $A = 100$, $B = 25$, $C = 23$, $D = 3$ and $L = 20$. The values are (A) $\mu_1 = 0.001$, (B) $\mu_1 = 0.01$, and (C) $\mu_1 = 0.1$. The values of μ_1 depends on the the energy J . The energies are (A) $J = 618$, (B) $J = 392$, and (C) $J = 358$. The unit for energy is arbitrary. We have removed points near $\lambda = 0$ since the error at those points are amplified.	18
4.3	The curve for $f_\Lambda(\lambda)$ dips below 0 near $\lambda = 10$. The parameters used were $\alpha = -10$, $\beta = 250$, $\gamma = 0.75$, $B = 250$, $C = 23$, $D = 3$, $L = 10$, and $\mu_1 = 1$	18
4.4	Information-energy curves for $A = 100$, $B = 25$, $C = 23$, $D = 3$, and $L = 20$. The energy unit is arbitrary. Note that as J becomes smaller, the curve over predicts the actual amount of information because $f_\Lambda(\lambda) < 0$ by a significant amount (see section 4.3).	19

-
- 4.5 The point of maximum bits per energy unit is indicated by the black star. This point is where $J = 648$ and $I = 3.39$. The maximum bits per energy is $\mu_1 = 0.0052$. The black line is the line through the origin that is tangent to the information-energy curve. The energy unit is arbitrary. The parameters are $\alpha = -75$, $\beta = 5000$, $\gamma = 10$, $A = 100$, $B = 25$, $C = 23$, $D = 3$, and $L = 20$ 20
- 5.1 The neuron as an independent multiplicative noise channel. Here Λ is the intensity of the input of the neuron, T is the output of the neuron and U is the “noise”. 22

Glossary of Terms

rest metabolic rate (RMR)	The rate of energy use in an organism when it is not under stress
action potential (AP)	The pulse generated by a neuron as a form of communication
glial cells	Cells that perform “house-keeping” tasks for neurons, such as directing growth and removing dead cells
dendrite	The input device of the neuron
soma	The cell body of the neuron
axon	The output line of the neuron
synapse	The connection formed when by an axon and a dendrite. This is where communication between neurons occur
interpulse interval (IPI)	The time interval between two consecutive spikes
absolute refractory period	The time period following the firing of an AP; the neuron cannot fire during this period
relative refractory period	The time period following the absolute refractory period; the neuron can fire during this period, but with a relatively high energy cost
post-synaptic potential (PSP)	The voltage of the neuron. Also called the excitation
jitter	The error in spike timing due to noise

Chapter 1

Introduction

The study of neurons is a challenging task. It requires expertise from diverse areas such as biology, chemistry, medicine, psychology, electrical engineering, physics, among others. Our understanding of a single neuron is far from complete, and yet there are on the order of 10^{11} neurons in the human brain, forming a complex network. This network is the basis of thoughts, feelings, and consciousness. One could be overwhelmed by the intricacies of neural networks. This thesis will focus on the single neuron, hoping to acquire a better understanding of the complex biological thinking machine through its building blocks.

1.1 Energy Efficiency in Neurons

Neurons are energetically demanding. The human brain uses up to 20 percent of a person's rest metabolic rate (RMR)[2, 3]. Yet, the human brain only consumes energy at a rate of only 25 watts. In comparison, modern computers use energy at a rate in the order of 100 watts. This efficiency arises from natural selection. Energy can be a scarce resource for which organisms must compete. Hence, natural selection has favored organisms that are energy efficient. Evidence of energy minimization is observed in the sensory system, even to the single neuron level [4] and is found in ion channel kinetics in neurons [5, 6].

1.2 Information Theory in Neuroscience

Information theory has been applied to neuroscience since the early 1950s [7]. In [8], the neuron of a fly was analyzed, assuming temporal coding of neural spikes. The resulting

information rate was 64 bits per second. In [9], it was found that information increases as time was discretized into smaller intervals, even into the sub-millisecond level. In [10], the trade-off between energy cost and information rate of neural codes in photoreceptors are investigated for flies *in vivo*.

These analyses focus on neurons and cells with a small number of inputs. Experiments measuring the thousands of inputs of cortical neurons have not been performed due to the numerous necessary measurements. Studies that include cortical neurons with a large number of inputs were investigated with neural models instead. Energy efficient neural codes for population and frequency codes were investigated in [11] and expanded to include synaptic failures in [12]. Energy-efficiency of receiving information was investigated in [13] using a discrete set of input and output. Using a simplified model of the neuron, energy efficiency was investigated using a continuous-time code in [14] and expanded in [15]. The generalized inverse Gaussian distribution was first used to model a neuron in [16]. For a history of the application of information theory in neuroscience, see [17].

1.3 Motivation

This thesis continues the work in [16] and [18]. The thesis proposes the Barndorff-Nielsen diffusion as a model for a neuron's PSP buildup that takes into account the effect of the fast Na^+ channels by allowing the drift rate to increase with amplitude, which has not yet been done. The trade-off of energy and information for a neuron is then explored. The goal is to more accurately characterize the efficiency of the neuron and to better understand the information-energy trade-off of a neuron.

Chapter 2

Background

Neurons¹ are the computational units of the nervous system, including the brain. The function of neurons is to convey information to other neurons or muscles regarding the neurons' inputs by sending a pulse called the action potential (AP). Such inputs may originate from sensory cells or other neurons. Along with neurons, glial cells can also be found in the brain. Glial cells perform “housekeeping” tasks, such as directing the growth of neurons' axons, keeping neurons insulated from other neurons, removing dead cells, etc. Glial cells do not do any signaling, i.e. all signaling is performed strictly by the neurons.

2.1 Parts of a Neuron

Neurons are typically composed of three parts: the dendrite, the soma, and the axon (figure 2.1). The dendrite acts as the input device to the neuron, branching out and connecting to other neurons' axons. The soma is the cell body, where cellular components such as the nucleus is found. This is where excitation occurs. The axon acts as the output line, extending far from the neuron and connecting to other neurons' dendrites. The axon is modeled as a lossy transmission line. However, along the axon are essentially repeaters to minimize distortion of the signal. The repeaters regenerate the signal passing through the axons at certain intervals. In a cortical neuron, there are ca. 10,000 synapses.

The connection between a neuron's axon and another neuron's dendrite form the synapse. This is where information is transmitted. The synapse can be classified into two types: electrical and chemical. In electrical synapses, the axon and dendrite form a connection that allows ions to flow through. Thus, such synapses use ion flow as a mean of communication. Bidirectional communication is possible here as the ions can flow in either

¹also called nerve cells

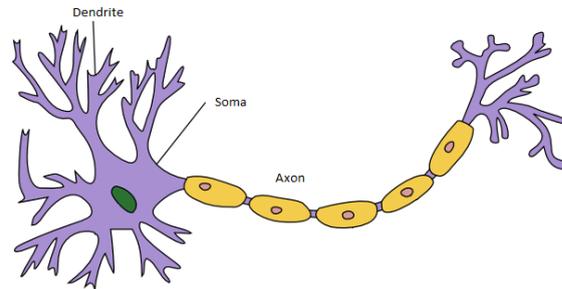


FIGURE 2.1: The three parts of a neuron.

direction. More commonly found are chemical synapses. The axon and dendrite are not physically in contact but form the synaptic cleft, a gap where neurotransmitters can diffuse across. Once the AP reaches the axon terminal, neurotransmitters are released. The neurotransmitters bind to ion channels at the dendrite. This opens the channels and allows ions to enter or exit the cell. This in turn alters the voltage of the soma, called the post-synaptic potential (PSP). A chemical synapse is either excitatory, where inputs to it increase the neuron's excitation, or inhibitory, where inputs to it decrease it. The excitatory synapses outnumber the inhibitory ones, thereby making the net excitation of a neuron positive. Communication here is unidirectional from the axon to the dendrite. Since electrical synapses are found only in specialized neurons, we are only interested in neurons with chemical synapses. Henceforth, "synapse" refer to chemical synapse.

2.2 Action Potential

The action potential (AP) is an all-or-nothing pulse. Figure 2.2 shows a recorded and simulated AP for the giant squid axon. For any particular neuron type, the AP's are the same in amplitude and shape every time. There is still some controversy regarding whether neurons use frequency or timing codes, i.e. whether neurons count the number of pulses in an interval or measure the time between two consecutive pulses. We take the latter view. Hence, neurons use time-continuous differential pulse position modulation (tcdppm) where the message is encoded in the duration between two consecutive pulses. This duration of time is the interpulse interval (IPI)² (figure 2.3). The pulse is generated in the initial segment of the axon and is propagated along the axon to synapses of the target neurons. The arrival of an AP either increases or decreases the PSP in the soma, depending on whether the synapse is excitatory or inhibitory. The PSP builds up until it reaches a "threshold", whereupon the neuron fires an AP. After an AP is fired, there is an interval called the absolute refractory period during which the neuron resets itself

²The term interspike interval (ISI) is also used. However, ISI can also refer to "intersymbol interference". Thus, we have adopted "IPI" to minimize confusion.

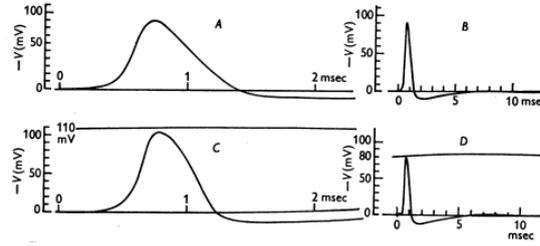


FIGURE 2.2: An AP measured and simulated with the Hodgkin-Huxley equations. The top plots are the simulated result. The bottom plots are the measured results. In the two plots on the bottom, the line above the measured the AP is the curvature of the window screen of the measuring apparatus used by Hodgkin and Huxley [1].

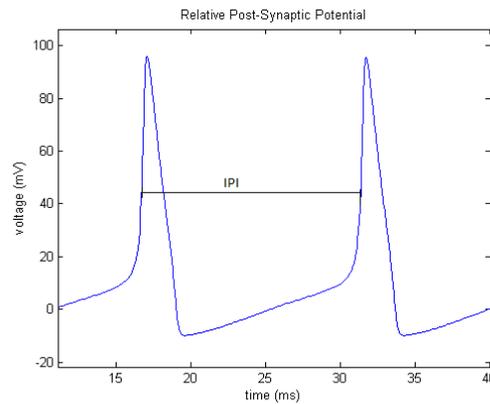


FIGURE 2.3: The IPI is the interval between two AP's. The reference point is the point of highest slope in the AP. The AP's were simulated using the equations in [1].

and cannot fire another AP. Following the absolute refractory period is a short relative refractory period, wherein firing an AP is possible, but at a relatively large energy cost.

2.2.1 Action Potential Generation Mechanism

The AP generation mechanism involve sodium (Na^+) and potassium (K^+) ion channels on the cell membrane of the neuron. The kinetics of the Na^+ and K^+ channels were studied and first modeled in [1]. Pumps on the cell membrane maintain a concentration difference of Na^+ and K^+ across the cell membrane. The Na^+ concentration is greater outside the cell, whereas the K^+ concentration is greater inside the cell. The ion channels have states: The Na^+ channels have a closed, an open, and an inactive state; whereas the K^+ channels have a closed and an open state. The probability of a state transition is a function of the PSP. As the PSP increases, the Na^+ ion channels open, creating an influx of Na^+ current. Among the Na^+ ion channels are fast Na^+ ion channels that open faster than the regular Na^+ ion channels. The fast Na^+ channels largely contribute a

sharp rise in the PSP. Past a certain voltage, the Na^+ channels enter the inactivated state, i.e. the channels are closed and cannot be reopened for a time. As the PSP increases, the K^+ ion channels begin to open, creating an efflux K^+ current to restore the PSP back to the steady-state level. This change in the PSP is propagated along the axon toward the target neurons. It is the propagated voltage that is known as the AP.

2.2.2 Jitter

It should be noted that these ion channels are subject to thermal noise, which distorts the precise timing of the APs. This distortion is referred to as jitter. This in turn distorts the IPI. To measure the IPI most accurately, take the point of the highest slope on both APs as a reference point. The higher the slope, the smaller the jitter. One can understand this by considering the AP's as rectangular pulses. Any amount of noise at the time of rise, which is the point of highest slope, will not affect the timing of the rise. Therefore the higher the slope of rise for the AP, the lower the jitter. The function of the fast Na^+ channels is to increase the maximum slope of the AP in order to reduce jitter [19]. For more background on neurons, the reader is referred to [20] and [21].

Chapter 3

Problem Formulation

3.1 Defining the Variables

We will take the viewpoint that neurons maximize mutual information for a given energy cost. Let \tilde{T}_k be the interpulse interval in interval k . Define Δ as the absolute refractory period and let $T_k = \tilde{T}_k - \Delta$. Let us define $\Lambda(t)$, the random net algebraic input intensity at time t , as [16]

$$\Lambda(t) = \lim_{\tau \downarrow 0} \frac{\pm |Q| * \text{P}[\text{arrival/departure of ion in } (t, t + \tau)]}{\tau}, \quad (3.1)$$

where $|Q|$ is the magnitude of the charge of the ion. The \pm indicates two possibilities: the sign is positive (+) when the charge is positive and arriving or the charge is negative and departing; otherwise, the sign is negative (-). $\Lambda(t)$ is a reflection of the environment of the organism and neural network. Though it is possible to find the mutual information between a random function, $\Lambda(t)$, and a set of random variables, $\{T_k\}$, it is far more manageable to find the mutual information between two random variables. We define Λ_k as

$$\Lambda_k = \frac{1}{T_k} \int_{S_{k-1}-\Delta}^{S_k} \Lambda(t) dt. \quad (3.2)$$

where $S_k = \tilde{T}_1 + \dots + \tilde{T}_k$. In other words Λ_k is the average excitation over the k th interval sans its absolute refractory period.

3.2 Information Rate

We wish to calculate the information rate between $\{\Lambda_k\}$ and $\{T_k\}$ for all intervals [22],

$$I := \lim_{n \rightarrow \infty} \frac{1}{n} I(\Lambda_1, \dots, \Lambda_n; T_1, \dots, T_n). \quad (3.3)$$

This is equivalent to the information rate between $\{\Lambda_k\}$ and $\{\tilde{T}_k\}$ because of a bijection between $\{T_k\}$ and $\{\tilde{T}_k\}$. We will assume that each IPI is mutually independent of the others¹. We will also assume the neuron is memoryless because it resets after firing an AP during the absolute refractory period. Finally, we assume the environment of the organism is stationary so that each (Λ_k, T_k) pairs are identically distributed. When unambiguous, we will drop the k index and refer to Λ_k and T_k as Λ and T , respectively. The information rate is then

$$I = I(\Lambda; T), \quad (3.4)$$

i.e. the mutual information between T and Λ in one interval.

3.3 The Problem Statement

We wish to maximize $I(\Lambda; T)$ given a fixed energy budget, J . There are energy costs associated with receiving Λ and producing T . Each cost function is denoted as $g_i(\lambda, t)$, for $i = 1, \dots, N$, where λ and t are realizations of Λ and T , respectively. We will assume a fixed conditional probability density function (pdf) for T given Λ , $f_{T|\Lambda}(t|\lambda)$, as determined by the biology and physics of the postsynaptic build up. We seek the marginal for Λ , i.e. $f_\Lambda(\lambda)$, that solves the following problem for fixed $f_{T|\Lambda}(t|\lambda)$,

$$\begin{aligned} & \underset{f_\Lambda(\lambda)}{\text{maximize}} && I(\Lambda; T) \\ & \text{subject to} && \mathbb{E} \left[\sum_{i=0}^N g_i(\Lambda, T) \right] = J \\ & && \int_0^\infty f_\Lambda(\lambda) d\lambda = 1 \\ & && f_\Lambda(\lambda) \geq 0, \lambda \geq 0, \end{aligned}$$

where the energy cost contributions are discussed in section 3.5. We will remove the constraint, $f_\Lambda(\lambda) \geq 0, \lambda \geq 0$, and check if the solution satisfies the constraint.

Since Λ is a product of the network, $f_\Lambda(\lambda)$ describes the behavior of the network. Even though we are studying a single neuron, we seek the behavior of the network that would

¹This is more accurate for large T_k . A short T_k is usually followed by a short T_{k+1} . But for optimality, a neuron will seek to minimize the correlation even among the lengths of successive short IPI's.

this neuron optimal through $f_\Lambda(\lambda)$. In a sense, the network is also considered in this study.

3.4 Modeling the PSP buildup

The PSP is often modeled as a diffusion building up towards a threshold. A studied model is the homogeneous Poisson counting process, as in [14], which for any λ , results in T being gamma distributed conditioned on $\Lambda = \lambda$. However, this model is too simplistic, ignoring the synaptic “weights”, PSP leakage, and inhibitory synapses. Another model is the Wiener process with drift, which for any λ , results in T being in the inverse Gaussian (IG) family given λ [23]. However, this model assumes a constant rate of buildup and lacks the increasing rate of PSP growth attributable to the opening of the sodium channels. Hence, we desire a more general diffusion process that displays this effect. These considerations motivate our adapting the the Barndorff-Nielsen diffusion for our neuron model.

A continuous-time diffusion is governed by the equation [24]

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t, \quad (3.5)$$

where Y_t is the position at time t , $\mu(x)$ is the drift rate at position x , $\sigma(x)$ is the standard deviation at position x , and W_t is the Wiener process. For a Barndorff-Nielsen diffusion [25],

$$\mu(x) = \sigma(x) \left(\frac{2\alpha - 1}{2\phi(x)} + \frac{\sqrt{2\gamma}K_{\alpha-1}(\phi(x)\sqrt{2\gamma})}{K_\alpha(\phi(x)\sqrt{2\gamma})} \right) + \frac{1}{4} \frac{d}{dx}[\sigma(x)^2] \quad (3.6)$$

where α and γ are parameters of the diffusion, $K_\nu(x)$ is the modified Bessel function of the second kind of degree ν , $\phi(x) = \int_x^\theta \frac{1}{\sigma(u)} du$ and θ is the threshold. It must be that $\alpha < 0$ and $\gamma > 0$. The standard deviation $\sigma(x)$ can be any function as long as $\phi(x) < \infty$ for $0 < x < \infty$. For the case where $\sigma(x) = \sigma$ is constant and $\alpha = -\frac{1}{2}$, the diffusion becomes a Wiener process with a constant drift of $2\gamma\sigma^2$ and variance σ^2 . For $\alpha < -\frac{1}{2}$, drift rate increases rapidly as the diffusion approaches threshold. Figures 3.1 and 3.2 shows a sample path of Barndorff-Nielsen diffusions.

In a neuron, Λ indicates the average incoming rate of the APs from the input neurons. Suppose we fix $\Lambda = \lambda$. If we let $\alpha = -\frac{1}{2}$, then 2γ is the drift rate and $\frac{1}{2\beta}$ is the variance. It is clear that if Λ were to increase k -fold in an IPI, then the drift rate becomes $2k\gamma$ and the variance becomes $\frac{k}{2\beta}$ because there are twice as many infinitesimal Gaussian random variables to add. If we extend this to the Barndorff-Nielsen diffusion by letting $\alpha < 0$, then this diffusion results in a pdf for T conditional on $\Lambda = \lambda$ that is a generalized inverse gaussian distribution $GIG(\alpha, \frac{\beta}{\lambda}, \gamma\lambda)$ where λ multiplies γ and divides β , i.e.

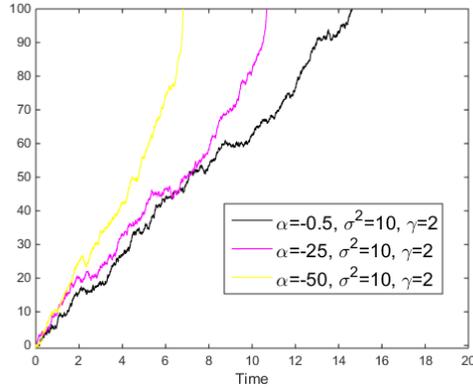


FIGURE 3.1: Three Barndorff-Nielsen diffusions with threshold $\theta = 100$ and a fixed variance. All parameters are kept the same except for α . The lower the value of α , the stronger the attraction towards the threshold. The black curve is also a Wiener process with a drift rate of 10 and a variance of 10.

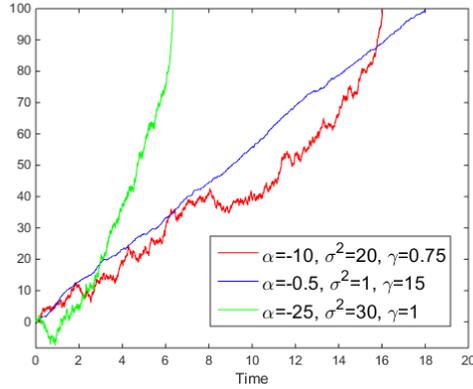


FIGURE 3.2: Three Barndorff-Nielsen diffusions with threshold $\theta = 100$ and a fixed variance. The blue curve is also a Wiener process with a drift rate of 7.5 and variance of 1.

$$f_{T|\Lambda}(t|\lambda) = C_{\underline{\alpha}}^{-1} \lambda^{\alpha} t^{\alpha-1} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right), \quad (3.7)$$

where α and γ correspond to the values in the Barndorff-Nielsen diffusion and $\beta = \frac{\phi(0)^2}{2}$; $C_{\underline{\alpha}}$ is the normalizing term, which has value

$$C_{\underline{\alpha}} = 2 \left(\frac{\beta}{\gamma}\right)^{\frac{\alpha}{2}} K_{\alpha}(2\sqrt{\beta\gamma}), \quad (3.8)$$

Figure 3.3 shows the GIG pdf for the diffusions in figure 3.2.

The GIG is selected as the conditional pdf because of several features. First, the GIG pdf has three parameters. The power of a model increases with the number of parameters except for beyond four parameters, where overfitting may occur [26]. Second, the GIG

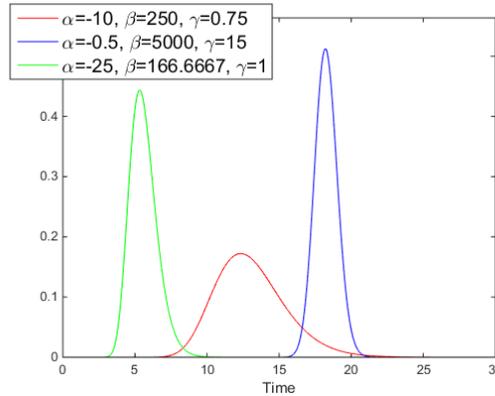


FIGURE 3.3: Three GIG pdfs corresponding with the diffusions in figure 3.2

pdf encompasses different subclasses of distributions, including the gamma and IG pdfs previously mentioned. The sodium channels involved with the PSP are more likely to open as the PSP increases, creating an attraction towards the so called threshold. A GIG pdf with $\alpha < -\frac{1}{2}$ indicates a crossing time of a diffusion with increasing rate. Thus the attractiveness of the GIG pdf.

3.5 Modeling the Energy Cost

We model the costs as the following functions:

- $g_1(\lambda, t) = A$, where $A > 0$. This is the fixed costs associated with processes that do not change, regardless of the input or output. An example is the cost to propagate an action potential (AP).
- $g_2(\lambda, t) = Bt$, where $B > 0$. This is the cost associated with processes that depend linearly with time, such as metabolic costs.
- $g_3(\lambda, t) = C\lambda t$, where $C > 0$. Since ΔT is approximately the number of AP's that arrive, this is the cost associated with processing the AP's that arrive at the synapse.
- $g_4(\lambda, t) = \frac{L}{t}$, where $L > 0$. This is the cost associated with the relative refractory period where the threshold increases right after the absolute refractory period.
- $g_5(\lambda, t) = -D \log(t)$ where $D > 0$. This is the cost associated with neural “clocks”. The accuracy of the timing is more crucial for short IPI's, so the neuron will expend more energy ensuring the accuracy of the timing of the AP for shorter IPI's. Note that $g_5(\lambda, t)$ changes sign at $t = 1$ so as to decrease energy for $D > 0$ and $t > 1$.

However this term is small compared to $g_2(\lambda, t)$ as long as $D < B$, so energy remains positive.

The total energy cost is the sum of the above functions.

Chapter 4

Results

4.1 Optimality Condition

The optimality conditions were determined in [14] to be,

$$\mathbb{E}\left[\log\left(\frac{f_{T|\Lambda}(t|\lambda)}{f_T(t)}\right)\middle|\Lambda = \lambda\right] = \mu_1\mathbb{E}\left[\sum_{i=1}^N g_i(\Lambda, T)\middle|\Lambda = \lambda\right] + \mu_0, \quad \forall \lambda \geq 0, \quad (4.1)$$

where $f_T(t) = \int_0^\infty f_{T|\Lambda}(t|\lambda)f_\Lambda(\lambda)d\lambda$, and μ_0 and μ_1 are constants that solve the problem. Note that μ_0 and μ_1 are invariant to λ . For an alternate derivation of this optimality condition, see appendix A. Since I is concave in $f_\Lambda(\lambda)$ for a fixed $f_{T|\Lambda}(t|\lambda)$ [22], the optimality condition is the condition for maximum I for fixed J .

4.2 The $f_T(t)$ Marginal

Making the substitutions in (4.1) for the conditional distribution and the cost functions defined above, we have

$$\begin{aligned} &(\alpha - 1)\mathbb{E}[\log(\Lambda T)|\Lambda = \lambda] + \log(\lambda) - \beta\mathbb{E}\left[\frac{1}{\Lambda T}\middle|\Lambda = \lambda\right] \\ &\quad - \gamma\mathbb{E}[\Lambda T|\Lambda = \lambda] - \log(C_\alpha) - \mathbb{E}[\log(f_T(T))|\Lambda = \lambda] = \\ &\quad \mu_1 A + \mu_1 B\mathbb{E}[T|\Lambda = \lambda] + \mu_1 C\mathbb{E}[\Lambda T|\Lambda = \lambda] \\ &\quad \quad + \mu_1 L\mathbb{E}\left[\frac{1}{T}\middle|\Lambda = \lambda\right] - \mu_1 D\mathbb{E}[\log(T)|\Lambda = \lambda] + \mu_0. \end{aligned} \quad (4.2)$$

The following expectations are evaluated:

$$\begin{aligned}\mathbb{E}[T|\Lambda = \lambda] &= C_{\underline{\alpha}}^{-1} \int_0^{\infty} \lambda^{\alpha} t^{\alpha} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right) dt \\ &= \sqrt{\frac{\beta}{\gamma} \frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})}} \frac{1}{\lambda};\end{aligned}\quad (4.3)$$

$$\begin{aligned}\mathbb{E}\left[\frac{1}{T} \middle| \Lambda = \lambda\right] &= C_{\underline{\alpha}}^{-1} \int_0^{\infty} \lambda^{\alpha} t^{\alpha-2} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right) dt \\ &= \sqrt{\frac{\gamma}{\beta} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})}} \lambda;\end{aligned}\quad (4.4)$$

$$\begin{aligned}\mathbb{E}[\log(T)|\Lambda = \lambda] &= C_{\underline{\alpha}}^{-1} \int_0^{\infty} \frac{\partial}{\partial \alpha} \left[\lambda^{\alpha} t^{\alpha-1} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right) \right] dt - \log(\lambda) \\ &= \frac{1}{2} \log\left(\frac{\beta}{\gamma}\right) + \frac{\frac{\partial}{\partial \alpha} K_{\alpha}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})} - \log(\lambda).\end{aligned}\quad (4.5)$$

Next, we define the following constants,

$$k_{lin} := \sqrt{\frac{\beta}{\gamma} \frac{K_{\alpha+1}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})}};\quad (4.6)$$

$$k_{inv} := \sqrt{\frac{\gamma}{\beta} \frac{K_{\alpha-1}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})}};\quad (4.7)$$

$$k_{log} := \frac{1}{2} \log\left(\frac{\beta}{\gamma}\right) + \frac{\frac{\partial}{\partial \alpha} K_{\alpha}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})}.\quad (4.8)$$

Note that $\mathbb{E}[\Lambda T|\Lambda = \lambda] = k_{lin}$, $\mathbb{E}[\frac{1}{\Lambda T}|\Lambda = \lambda] = k_{inv}$, and $\mathbb{E}[\log(\Lambda T)|\Lambda = \lambda] = k_{log}$ are constants independent on the value of λ .

Hence, (4.2) becomes

$$\begin{aligned}\mathbb{E}[\log(f_T(T))|\Lambda = \lambda] &= \\ &= (\mu_1 D - 1)(k_{log} - \log(\lambda)) - \mu_1 L k_{inv} \lambda - \mu_1 B \frac{k_{lin}}{\lambda} + G - \mu_1 \bar{A} - \mu_0,\end{aligned}\quad (4.9)$$

Where $G = \alpha k_{log} - \beta k_{inv} - \gamma k_{lin} - \log(C_{\underline{\alpha}})$ and $\bar{A} = A + C k_{lin}$ are constants determined by the parameters given in the problem. When taking the expectation of the marginal, we get $\log(\lambda)$, λ , and $\frac{1}{\lambda}$ terms. A GIG distribution satisfies these conditions. In fact, the marginal is $\text{GIG}(\mu_1 D, \mu_1 L, \mu_1 B)$, i.e.

$$f_T(t) = C_{\underline{A}}(\mu_1)^{-1} t^{\mu_1 D - 1} \exp\left(-\frac{\mu_1 L}{t} - \mu_1 B t\right),\quad (4.10)$$

where $C_{\underline{A}}(\mu_1) = 2\left(\frac{L}{B}\right)^{\frac{\mu_1 D}{2}} K_{\mu_1 D}(2\mu_1 \sqrt{LB})$. In order to satisfy (4.9), μ_0 has taken the value,

$$\mu_0 = G - \mu_1 \bar{A} + \log(C_{\underline{A}}(\mu_1)). \quad (4.11)$$

The value of μ_1 is determined from the energy constraint. Substituting the energy terms, the constraint is

$$J = A + B\mathbb{E}[T] + C\mathbb{E}[\Lambda T] + L\mathbb{E}\left[\frac{1}{T}\right] - D\mathbb{E}[\log(T)]. \quad (4.12)$$

Next, we define the following functions,

$$c_{in}(\mu_1) := \mathbb{E}[T] = \sqrt{\frac{L}{B}} \frac{K_{\mu_1 D+1}(2\mu_1 \sqrt{LB})}{K_{\mu_1 D}(2\mu_1 \sqrt{LB})}; \quad (4.13)$$

$$c_{inv}(\mu_1) := \mathbb{E}\left[\frac{1}{T}\right] = \sqrt{\frac{B}{L}} \frac{K_{\mu_1 D-1}(2\mu_1 \sqrt{LB})}{K_{\mu_1 D}(2\mu_1 \sqrt{LB})}; \quad (4.14)$$

$$c_{log}(\mu_1) := \mathbb{E}[\log(T)] = \frac{1}{2} \log\left(\frac{L}{B}\right) + \frac{\frac{\partial}{\partial \nu} K_{\nu}(2\mu_1 \sqrt{LB})|_{\nu=\mu_1 D}}{K_{\mu_1 D}(2\mu_1 \sqrt{LB})}. \quad (4.15)$$

Hence,

$$J = \bar{A} + Bc_{in}(\mu_1) + Lc_{inv}(\mu_1) - Dc_{log}(\mu_1). \quad (4.16)$$

Using $\frac{\partial K_{\nu}(x)}{\partial x} = -\frac{1}{2}(K_{\nu-1}(x) + K_{\nu+1}(x))$, we can get that

$$J = \bar{A} - \frac{d}{d\mu_1} \log(C_{\underline{A}}(\mu_1)). \quad (4.17)$$

This equation may be solved numerically for μ_1 or used in part of plotting information as a function of energy as described in section 4.4.

4.3 The $f_{\Lambda}(\lambda)$ Marginal

First, we define U such that,

$$U = \Lambda T. \quad (4.18)$$

We now derive the pdf of U . Let $\Lambda = \lambda$ be fixed. Then the conditional cdf of U given $\Lambda = \lambda$ is,

$$F_{U|\Lambda}(u|\lambda) = F_{T|\Lambda}\left(\frac{u}{\lambda} \mid \lambda\right). \quad (4.19)$$

Differentiating with respect to u yields the conditional pdf of U given $\Lambda = \lambda$. From 3.7,

$$\begin{aligned} f_{U|\Lambda}(u|\lambda) &= \frac{1}{\lambda} f_{T|\Lambda}\left(\frac{u}{\lambda}|\lambda\right) \\ &= C_{\underline{\alpha}}^{-1} u^{\alpha-1} \exp\left(-\frac{\beta}{u} - \gamma u\right). \end{aligned} \quad (4.20)$$

Therefore, U is distributed as $\text{GIG}(\alpha, \beta, \gamma)$. Furthermore, since $f_{U|\Lambda}(u|\lambda)$ is not a function of λ , U is independent of Λ . It should be noted that U is independent of Λ in this GIG case.

Defining $W = \log(U)$, $V = \log(T)$, and $Z = -\log(\Lambda)$, then,

$$V = W + Z. \quad (4.21)$$

W is independent of Z because of a bijection between W and U , and between Z and Λ . Hence the distribution for V is the convolution of the distributions of W and Z . The distribution of W is $f_W(w) = e^w f_U(e^w)$, i.e.

$$f_W(w) = C_{\underline{\alpha}}^{-1} e^{\alpha w} \exp(-\beta e^{-w} - \gamma e^w). \quad (4.22)$$

Similarly, $f_V(v) = e^v f_T(e^v)$. So,

$$f_V(v) = C_{\underline{A}}(\mu_1)^{-1} e^{\mu_1 D v} \exp(-\mu_1 L e^{-v} - \mu_1 B e^v). \quad (4.23)$$

The distribution of Z is the deconvolution of the two previous distributions. This can be done using characteristic functions (CF) of the respective random variables. Note that,

$$\Phi_V(x) = \Phi_W(x) \Phi_Z(x). \quad (4.24)$$

The CFs for W and V can be determined,

$$\begin{aligned} \Phi_W(x) &= E[\exp(jxW)] \\ &= C_{\underline{\alpha}}^{-1} \int_{-\infty}^{\infty} e^{(\alpha+jx)w} \exp(-\beta e^{-w} - \gamma e^w) dw \\ &= \left(\frac{\beta}{\gamma}\right)^{\frac{jx}{2}} \frac{K_{\alpha+jx}(2\sqrt{\beta\gamma})}{K_{\alpha}(2\sqrt{\beta\gamma})}. \end{aligned} \quad (4.25)$$

Similarly,

$$\Phi_V(x) = \left(\frac{L}{B}\right)^{\frac{jx}{2}} \frac{K_{\mu_1 D + jx}(2\mu_1 \sqrt{LB})}{K_{\mu_1 D}(2\mu_1 \sqrt{LB})}. \quad (4.26)$$

Rearranging (4.24) and substituting (4.22) and (4.23), we have the CF for Z ,

$$\begin{aligned}\Phi_Z(x) &= \frac{\Phi_V(x)}{\Phi_W(x)} \\ &= \left(\frac{\gamma L}{\beta B}\right)^{\frac{jx}{2}} \frac{K_{\mu_1 D + jx}(2\mu_1 \sqrt{LB}) K_\alpha(2\sqrt{\beta\gamma})}{K_{\alpha + jx}(2\sqrt{\beta\gamma}) K_{\mu_1 D}(2\mu_1 \sqrt{LB})}.\end{aligned}\quad (4.27)$$

Then the distribution of Z is the inverse transform,

$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\Phi_V(x)}{\Phi_W(x)} e^{-jxz} dx. \quad (4.28)$$

Since $f_\Lambda(\lambda) = \frac{1}{\lambda} f_Z(-\log(\lambda))$,

$$\begin{aligned}f_\Lambda(\lambda) &= \frac{1}{2\pi\lambda} \int_{-\infty}^{\infty} \frac{\Phi_V(x)}{\Phi_W(x)} \lambda^{jx} dx \\ &= \frac{1}{2\pi\lambda} \int_{-\infty}^{\infty} \left(\lambda \sqrt{\frac{\gamma L}{\beta B}}\right)^{jx} \frac{K_{\mu_1 D + jx}(2\mu_1 \sqrt{LB}) K_\alpha(2\sqrt{\beta\gamma})}{K_{\alpha + jx}(2\sqrt{\beta\gamma}) K_{\mu_1 D}(2\mu_1 \sqrt{LB})} dx.\end{aligned}\quad (4.29)$$

The values of $f_\Lambda(\lambda)$ can be calculated numerically.

Alternatively, the fast Fourier transform (FFT) algorithm can be used to acquire a numerical representation of $f_Z(z)$. Then the distribution for $f_\Lambda(\lambda)$ can be calculated. Note that to transform the pdf of Z to the pdf of Λ , we must divide by λ . Any error near $\lambda = 0$ will be greatly amplified. Sample distributions are shown in figure 4.1 and 4.2.

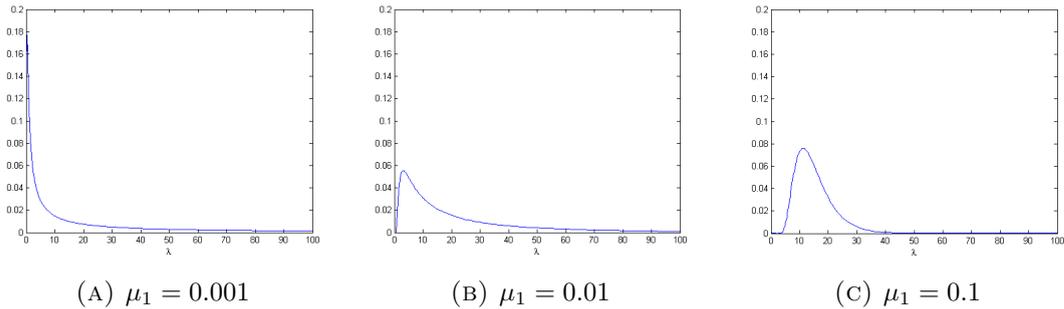


FIGURE 4.1: The marginal densities for Λ at different values of μ_1 for $\alpha = -10$, $\beta = 250$, $\gamma = 0.75$, $A = 100$, $B = 25$, $C = 23$, $D = 3$ and $L = 20$. The values are (A) $\mu_1 = 0.001$, (B) $\mu_1 = 0.01$, and (C) $\mu_1 = 0.1$. The values of μ_1 depends on the the energy J . The energies are (A) $J = 710$, (B) $J = 484$, and (C) $J = 450$. The unit for energy is arbitrary. We have removed points near $\lambda = 0$ since the error at those points are amplified.

The pdf's of Λ produced in figures 4.1 and 4.2 actually has a minimum of roughly -10^{-8} to -10^{-7} , which violates the non-negativity of probability distributions. The cause for this is not clear: it may be machine error from the FFT algorithm or an actual result. This can be determined by using a larger window for $\Phi_Z(x)$, which should produce a

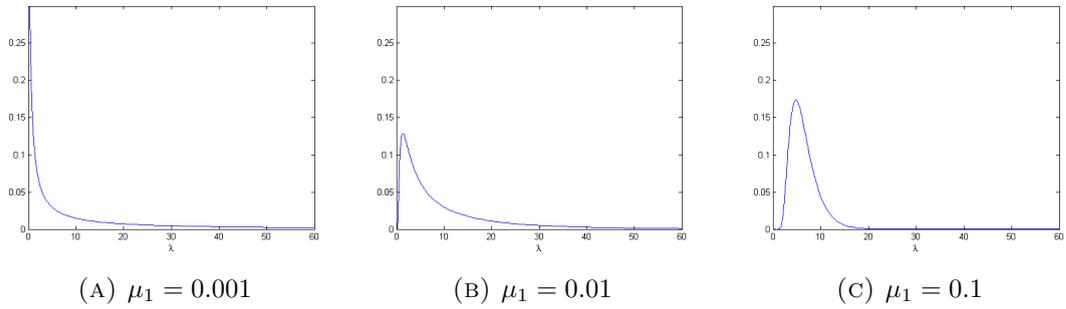


FIGURE 4.2: The marginal densities for Λ at different values of μ_1 for $\alpha = -10$, $\beta = 166.67$, $\gamma = 1$, $A = 100$, $B = 25$, $C = 23$, $D = 3$ and $L = 20$. The values are (A) $\mu_1 = 0.001$, (B) $\mu_1 = 0.01$, and (C) $\mu_1 = 0.1$. The values of μ_1 depends on the the energy J . The energies are (A) $J = 618$, (B) $J = 392$, and (C) $J = 358$. The unit for energy is arbitrary. We have removed points near $\lambda = 0$ since the error at those points are amplified.

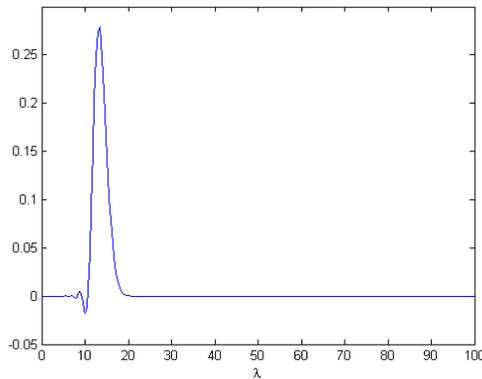


FIGURE 4.3: The curve for $f_\Lambda(\lambda)$ dips below 0 near $\lambda = 10$. The parameters used were $\alpha = -10$, $\beta = 250$, $\gamma = 0.75$, $B = 250$, $C = 23$, $D = 3$, $L = 10$, and $\mu_1 = 1$

more accurate plot of $f_Z(z)$ when the FFT algorithm is performed. This in result would produce a more accurate $f_\Lambda(\lambda)$. Nevertheless, the value is almost negligible, hence our approximation holds. However, as $\mu_1 D$ approaches α or the product $(\mu_1 L)(\mu_1 B)$ approaches $\beta\gamma$, the drop below zero becomes more noticeable (see figure 4.3). At this point, it is clear that the function that optimizes the information has negative values. Therefore, it is necessary for certain, perhaps all, parameters in the problem to include the restriction $f_\Lambda(\lambda) \geq 0$. This requires the Karush-Kuhn-Tucker (KKT) conditions for a function space in order to solve the problem.

4.4 Information-Energy Function

By taking the expectation of both sides of (4.1), we have the following result,

$$I = \mu_1 J + \mu_0. \quad (4.30)$$

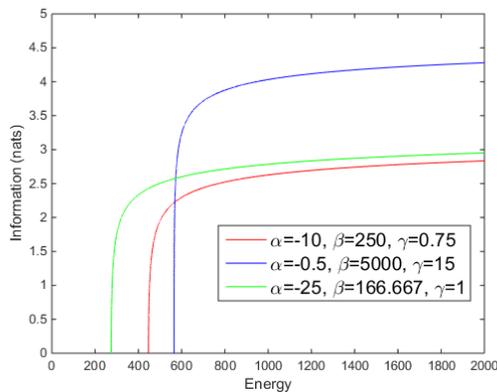


FIGURE 4.4: Information-energy curves for $A = 100$, $B = 25$, $C = 23$, $D = 3$, and $L = 20$. The energy unit is arbitrary. Note that as J becomes smaller, the curve over predicts the actual amount of information because $f_{\Lambda}(\lambda) < 0$ by a significant amount (see section 4.3).

The variables J and μ_0 were determined as functions of μ_1 in (4.16) and (4.11). Hence, I and J can be plotted parametrically by letting μ_1 be the parameter. Some sample curves are plotted in figure 4.4.

We can determine the slope of the information-energy curve,

$$\frac{dI}{dJ} = \frac{dI/d\mu_1}{dJ/d\mu_1} = \frac{J + \mu_1(dJ/d\mu_1) + d\mu_0/d\mu_1}{dJ/d\mu_1} = \mu_1, \quad (4.31)$$

using the fact that,

$$J + \frac{d\mu_0}{d\mu_1} = \bar{A} - \frac{d}{d\mu_1} \log(C_{\underline{A}}(\mu_1)) + \frac{d}{d\mu_1} (G - \mu_1 \bar{A} + \log(C_{\underline{A}}(\mu_1))) = 0. \quad (4.32)$$

The point of maximum bits per energy is where a line through the origin is tangent to the curve. We have determined that the slope of the information-energy curve is μ_1 . Hence, the point of maximum bits per energy is where $I = \mu_1 J$, i.e. $\mu_0 = 0$. Also, the maximum bits per energy is μ_1 at that point. This point can be determined numerically. Figure 4.5 illustrates this point for a sample curve.

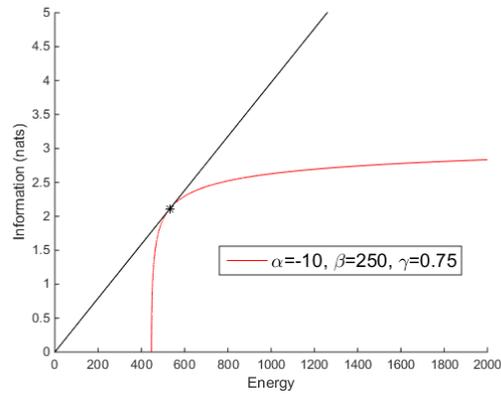


FIGURE 4.5: The point of maximum bits per energy unit is indicated by the black star. This point is where $J = 648$ and $I = 3.39$. The maximum bits per energy is $\mu_1 = 0.0052$. The black line is the line through the origin that is tangent to the information-energy curve. The energy unit is arbitrary. The parameters are $\alpha = -75$, $\beta = 5000$, $\gamma = 10$, $A = 100$, $B = 25$, $C = 23$, $D = 3$, and $L = 20$.

Chapter 5

Discussion

The derived pdf of Λ is a unimodal function. It appears that as μ_1 decreases, the mode of Λ moves to the left towards the vertical axis. Since a lower value of μ_1 means a higher value of J , the neuron prefers to receive a lower intensity of inputs at higher energies. This seems to suggest that lower values of Λ results in higher information. Nevertheless, lower values of Λ is costly. This is because a low value of Λ is correlated with a high value of T , which increases energy as it increases. Hence, for low Energy values, low values of Λ is less desirable and the network will assign them low probabilities. The network trades off high information from low values of Λ with the energy cost to produce them.

There are two ways to view the information-energy curves. First, the curves represent the amount of information for a given energy that an energy-optimized neuron can convey. Alternatively, the curves are the cost-constrained capacity of neurons, i.e. the maximum amount of information conveyed given the neuron has an energy cost constraint. Efficient neurons' operating points should lie somewhere near the curve. For a given energy, it is possible to increase information by changing $f_{T|\Lambda}(t|\lambda)$. However, $f_{T|\Lambda}(t|\lambda)$ is determined by the biology of the neuron. Changing the parameters of the neuron requires trade-offs beyond the scope of this paper, e.g. size, ion concentrations, temperature, etc.

The point of maximum bits per energy cost is near the large change in slope of the information-energy curve. A neuron at rest will operate near that point because that is where the neuron is the most efficient. In situations where more information is necessary, energy can be increased, though with diminishing returns. The higher the energy increase, the slower the increase in information. This result agrees with [10], where photoreceptors can increase the amount of bits conveyed by increasing energy, but with diminishing returns.

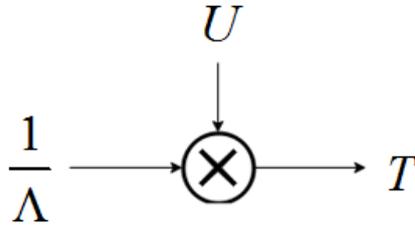


FIGURE 5.1: The neuron as an independent multiplicative noise channel. Here Λ is the intensity of the input of the neuron, T is the output of the neuron and U is the “noise”.

5.1 Relationship to Fading

From section 4.3, we can view the GIG conditional pdf as a multiplicative channel where $\frac{1}{\Lambda}$ is the input, T is the output, and U is the multiplicative noise independent of the input (see figure 5.1). Multiplicative noise is of interest in engineering problems, one of which is fading. Fading is relevant in wireless communications for portable receivers [27]. An independent multiplicative noise is used to model the effect of fading on a signal. The noise distribution is often Rayleigh or Rician instead of GIG. However, all three distributions share a unimodal shape. In the fading channel, if the transmitted data are encoded in the amplitude of the signal, then we have a similar phenomenon happening in the neuron. The multiplicative noise will alter the amplitude and interfere with the recovery of the symbols, which is similar to noise corrupting the timing of the spikes in neurons. The optimization technique developed here may be extended to the fading channel to achieve capacity given a constraint on the expectation of functions of the input and output. However, a major difference in fading channels is the signal is further corrupted by additive Gaussian noise and a decision is made at the receiver to recover a discrete alphabet, whereas the neuron receives a continuous alphabet without quantization. We have ignored thermal noise in this neuron model. Perhaps after incorporating thermal noise and jitter, our neuron model is even more similar to fading channels. Further studies are required to extend the theories here to fading channels.

Chapter 6

Conclusion

The information-energy function for optimal neurons is an increasing concave function. Increasing the energy increases the information, but at a decreasing rate. The Barndorff-Nielsen diffusion is an appropriate model for the PSP buildup as it captures the increasing rate caused by the fast Na^+ . The resulting marginal pdf of T given Λ is a GIG distribution. The resulting distribution of T is also GIG, but with different parameters. The input distribution can be numerically solved. Since not all parameters produce a valid input distribution, it is necessary to use the KKT conditions.

The primary contribution of this work is an improved model for analyzing the information-energy trade-off of neurons. We have added a basis for adopting the GIG distribution as the conditional of T given Λ by modeling the PSP buildup with the Barndorff-Nielsen diffusion. In addition, the marginal distribution of the input and output of a cortical neuron was derived and can be used as a framework in analyzing neurons. Finally, a method for determining the constrained capacity of continuous distributions were described. This method applies to general channels, not strictly to neurons.

A possible next step for research is to reevaluate the solution using the KKT conditions. This would allow for a more accurate information-energy curve, as well as the marginals of Λ and T . An alternative route to future research is to better relate the GIG distribution and its associated diffusion process to the biology of the neuron. Perhaps upon further study, we find that another distribution fit the reality better than the current model. It is difficult for experimental neuroscientists to measure the input distribution due to the nature of data collection. It is our belief that in order to study neurons in their natural working conditions, experiments should be done in vivo, as opposed to in vitro. Thus it is difficult to manipulate Λ in vivo as many factors affect the input intensity. Hence, further studies may involve studying appropriate neuron models.

Appendix A

Optimal Condition Derivation

Here, we present an alternative derivation of the optimality condition. The condition for a maximizing $I[f_\Lambda]$ constrained to a set of equations $G_i[f_\Lambda] = C_i$, $i = 0 \dots N - 1$ is [28]

$$\frac{\delta I[f_\Lambda]}{\delta f_\Lambda} - \sum_{i=0}^{N-1} \mu_i \frac{\delta G_i[f_\Lambda]}{\delta f_\Lambda} = 0, \quad (\text{A.1})$$

where μ_i 's are constants. For a general functional of the form $G[f_\Lambda] = \int_0^\infty \mathcal{G}(\lambda, f_\Lambda(\lambda)) d\lambda$, then its variational derivative is

$$\frac{\delta G[f_\Lambda]}{\delta f_\Lambda} = \frac{\partial}{\partial \nu} [\mathcal{G}(\lambda, \nu)]_{\nu=f_\Lambda(\lambda)}. \quad (\text{A.2})$$

For our constraint equations, it is clear what the variational derivative ought to be, i.e.

$$\frac{\delta G_0[f_\Lambda]}{\delta f_\Lambda} = \frac{\partial}{\partial \nu} [\nu]_{\nu=f_\Lambda(\lambda)} = 1 \quad (\text{A.3})$$

and

$$\frac{\delta G_1[f_\Lambda]}{\delta f_\Lambda} = \frac{\partial}{\partial \nu} [\nu \int_0^\infty f_{T|\Lambda}(t|\lambda) \sum_{i=0}^N g_i(\lambda, t) dt]_{\nu=f_\Lambda(\lambda)} = \int_0^\infty f_{T|\Lambda}(t|\lambda) \sum_{i=0}^N g_i(\lambda, t) dt. \quad (\text{A.4})$$

However, the definition of mutual information cannot be written in the same form as $G[f_\Lambda]$ because f_Λ is nested inside another integral. Hence, another method is used to obtain the variational derivative.

First, let $I[f_\Lambda]$ be the mutual information between Λ and T as a function of f_Λ . Suppose for a perturbation h in the function f_Λ , the difference $I[f_\Lambda + h] - I[f_\Lambda]$ can be written as

$$I[f_\Lambda + h] - I[f_\Lambda] = \delta I[f_\Lambda, h] + \epsilon(\|h\|), \quad (\text{A.5})$$

where $\epsilon(\|h\|) \rightarrow 0$ as $h \rightarrow 0$ and $\delta I[f_\Lambda, h]$ is linear in h . Then I is said to be differentiable and $\delta I[f_\Lambda, h]$ is the differential of I . I is maximized if $\delta I[f_\Lambda, h] = 0$ [28].

The definition of $I[f_\Lambda]$ is

$$I[f_\Lambda] = \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) f_\Lambda(\lambda) \log \left(\frac{f_{T|\Lambda}(t|\lambda)}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} \right) dt d\lambda. \quad (\text{A.6})$$

Then,

$$\begin{aligned} I[f_\Lambda + h] - I[f_\Lambda] &= \\ & \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) (f_\Lambda(\lambda) + h(\lambda)) \log \left(\frac{f_{T|\Lambda}(t|\lambda)}{\int_0^\infty f_{T|\Lambda}(t|\nu) (f_\Lambda(\nu) + h(\nu)) d\nu} \right) dt d\lambda \\ & - \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) f_\Lambda(\lambda) \log \left(\frac{f_{T|\Lambda}(t|\lambda)}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} \right) dt d\lambda. \end{aligned} \quad (\text{A.7})$$

The difference becomes,

$$\begin{aligned} I[f_\Lambda + h] - I[f_\Lambda] &= \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) \left[f_\Lambda(\lambda) \log \left(\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu \right) \right. \\ & \quad - f_\Lambda(\lambda) \log \left(\int_0^\infty f_{T|\Lambda}(t|\nu) (f_\Lambda(\nu) + h(\nu)) d\nu \right) + h(\lambda) \log(f_{T|\Lambda}(t|\lambda)) \\ & \quad \left. - h(\lambda) \log \left(\int_0^\infty f_{T|\Lambda}(t|\nu) (f_\Lambda(\nu) + h(\nu)) d\nu \right) \right] dt d\lambda. \end{aligned} \quad (\text{A.8})$$

Using the Taylor polynomial for logarithms, i.e. $\log(c+x) = \log(c) + \frac{x}{c} + o(x)$, then

$$\begin{aligned} I[f_\Lambda + h] - I[f_\Lambda] &= \\ & \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) \left[-f_\Lambda(\lambda) \frac{\int_0^\infty f_{T|\Lambda}(t|\nu) h(\nu) d\nu}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} + o \left(\int_0^\infty f_{T|\Lambda}(t|\nu) h(\nu) d\nu \right) \right. \\ & \quad + h(\lambda) \log(f_{T|\Lambda}(t|\lambda)) - h(\lambda) \log \left(\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu \right) \\ & \quad \left. - h(\lambda) \frac{\int_0^\infty f_{T|\Lambda}(t|\nu) h(\nu) d\nu}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} + o \left(\int_0^\infty f_{T|\Lambda}(t|\nu) h(\nu) d\nu \right) \right] dt d\lambda. \end{aligned} \quad (\text{A.9})$$

Note that,

$$\begin{aligned} & \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) f_\Lambda(\lambda) \frac{\int_0^\infty f_{T|\Lambda}(t|\nu) h(\nu) d\nu}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} d\lambda dt = \\ & \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) h(\lambda) \frac{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} d\lambda dt = \\ & \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) h(\lambda) d\lambda dt. \end{aligned} \quad (\text{A.10})$$

Collecting the terms that are not linear in $h(\lambda)$ and replacing it by $o(h(\lambda))$ yields,

$$\begin{aligned} I[f+h] - I[f] &= \int_0^\infty \int_0^\infty f_{T|\Lambda}(t|\lambda) \left[-h(\lambda) + h(\lambda) \log(f_{T|\Lambda}(t|\lambda)) \right. \\ &\quad \left. - h(\lambda) \log \left(\int_0^\infty f_{T|\Lambda}(t|\nu) f(\nu) d\nu \right) \right] dt d\lambda + o(h(\lambda)). \end{aligned} \quad (\text{A.11})$$

We have $o(h(\lambda)) \rightarrow 0$ as $\|h(\lambda)\| \rightarrow 0$, hence $I[f_\Lambda]$ is differentiable. Henceforth, we will drop the $o(h(\lambda))$ term. Then we have,

$$\begin{aligned} I[f_\Lambda + h] - I[f_\Lambda] &= \int_0^\infty h(\lambda) \int_0^\infty f_{T|\Lambda}(t|\lambda) \left[\log \left(\frac{f_{T|\Lambda}(t|\lambda)}{\int_0^\infty f_{T|\Lambda}(t|\nu) f_\Lambda(\nu) d\nu} \right) - 1 \right] dt d\lambda \\ &= \int_0^\infty h(\lambda) \mathbb{E} \left[\log \left(\frac{f_{T|\Lambda}(T|\Lambda)}{f_T(T)} \right) - 1 \middle| \Lambda = \lambda \right] d\lambda, \end{aligned} \quad (\text{A.12})$$

where $f_T(t) = \int_0^\infty f_{T|\Lambda}(t|\lambda) f_\Lambda(\lambda) d\lambda$. Then, $\frac{\delta I[f_\Lambda]}{\delta f_\Lambda}$ is defined to be [29],

$$\delta I[f_\Lambda, h] = \int_0^\infty h(\lambda) \frac{\delta I[f]}{\delta f} d\lambda. \quad (\text{A.13})$$

Therefore,

$$\frac{\delta I[f_\Lambda]}{\delta f_\Lambda} = \mathbb{E} \left[\log \left(\frac{f_{T|\Lambda}(T|\Lambda)}{f_T(T)} \right) \middle| \Lambda = \lambda \right] - 1. \quad (\text{A.14})$$

Hence, the optimality condition is

$$\mathbb{E} \left[\log \left(\frac{f_{T|\Lambda}(t|\lambda)}{f_T(t)} \right) \middle| \Lambda = \lambda \right] - 1 - \mu_1 \mathbb{E} \left[\sum_{i=1}^N g_i(\Lambda, T) \middle| \Lambda = \lambda \right] - \mu_0 = 0. \quad (\text{A.15})$$

This must hold for all values of λ we are interested in, namely $\lambda \geq 0$. In (4.1) We have redefined $\mu_0 := \mu_0 + 1$ to simplify.

Bibliography

- [1] A. L. Hodgkin and A. F. Huxley, “A quantitative description of membrane current and its application to conduction and excitation in nerve,” *The Journal of physiology*, vol. 117, no. 4, pp. 500–544, 1952.
- [2] L. C. Aiello and P. Wheeler, “The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution,” *Current Anthropology*, vol. 36, no. 2, pp. pp. 199–221, 1995. [Online]. Available: <http://www.jstor.org/stable/2744104>
- [3] D. F. Rolfe and G. C. Brown, “Cellular energy utilization and molecular origin of standard metabolic rate in mammals,” *Physiological Reviews*, vol. 77, no. 3, pp. 731–758, 1997.
- [4] 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>
- [5] 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>
- [6] 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>
- [7] D. MacKay and W. 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>
- [8] 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>

- [9] I. Nemenman, G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck, “Neural coding of natural stimuli: Information at sub-millisecond resolution,” *PLoS Comput Biol*, vol. 4, no. 3, p. e1000025, 03 2008. [Online]. Available: <http://dx.plos.org/10.1371%2Fjournal.pcbi.1000025>
- [10] J. E. Niven, J. C. Anderson, and S. B. Laughlin, “Fly photoreceptors demonstrate energy-information trade-offs in neural coding,” *PLoS Biol*, vol. 5, p. e116, 03 2007. [Online]. Available: <http://dx.doi.org/10.1371%2Fjournal.pbio.0050116>
- [11] W. B. Levy and R. A. Baxter, “Energy efficient neural codes,” *Neural Computation*, vol. 8, no. 3, pp. 531–543, April 1996.
- [12] —, “Energy-efficient neuronal computation via quantal synaptic failures,” *The Journal of Neuroscience*, vol. 22, no. 1, pp. 4746–4755, June 2002.
- [13] V. Balasubramanian, D. Kimber, and M. J. Berry, “Metabolically efficient information processing,” *Neural Computation*, vol. 13, no. 4, pp. 799–815, April 2001.
- [14] 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, February 2010.
- [15] J. Xing, T. Berger, and T. Sejnowski, “A berger-levy energy efficient neuron model with unequal synaptic weights,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 2964–2968.
- [16] J. Xing and T. Berger, “Energy efficient neurons with generalized inverse gaussian conditional and marginal hitting times,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 1824–1828.
- [17] A. Dimitrov, A. Lazar, and J. Victor, “Information theory in neuroscience,” *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 1–5, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s10827-011-0314-3>
- [18] J. Xing, T. Berger, M. Sungkar, and W. B. Levy, “Energy efficient neurons with generalized inverse gaussian conditional and marginal hitting times,” *Transactions on Information Theory*, 2015, submitted.
- [19] G. Ariav, A. Polsky, and J. Schiller, “Submillisecond precision of the input-output transformation function mediated by fast sodium dendritic spikes in basal dendrites of ca1 pyramidal neurons,” *The Journal of Neuroscience*, vol. 23, no. 21, pp. 7750–7758, 2003. [Online]. Available: <http://www.jneurosci.org/content/23/21/7750.abstract>

-
- [20] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of Neural Science*, 4th ed. McGraw-Hill, 2000.
- [21] T. F. Weiss, *Cellular Biophysics: Electrical Properties*. The MIT Press, 1996, vol. 2.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, N.J.: John Wiley and Sons, Inc., 2006.
- [23] 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.
- [24] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*. Springer Science & Business Media, 1992, vol. 23.
- [25] O. Barndorff-Nielsen, P. Blsild, 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. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0304414978900364>
- [26] F. Dyson, "A meeting with Enrico Fermi," *Nature*, vol. 427, no. 6972, pp. 297–297, jan 2004. [Online]. Available: <http://dx.doi.org/10.1038/427297a>
- [27] S. Haykin and M. Moher, *Modern Wireless Communications*. Upper Saddle River, N.J.: Pearson Prentice Hall, 2005.
- [28] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, ser. Selected Russian Publications in the Mathematical Sciences, R. A. Silverman, Ed. Englewood Cliffs, N.J.: Prentice-Hall, 1963, translated to English.
- [29] M. Giaquinta and S. Hildebrandt, *Calculus of Variations*, ser. Grundlehren der Mathematischen Wissenschaften. New York: Springer, 1996, vol. 1.