

# A Berger-Levy Energy Efficient Neuron Model with Unequal Synaptic Weights

Jie Xing

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

Toby Berger

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

Terrence J. Sejnowski

‡ Howard Hughes Medical Institute,  
Computational Neurobiology Laboratory,  
Salk Institute for Biological Studies,  
La Jolla, CA 92037;  
Division of Biological Sciences,  
University of California,  
San Diego, La Jolla, CA 92093.  
sejnowski@salk.edu

**Abstract**—How neurons in the cerebral cortex process and transmit information is a long-standing question in systems neuroscience. To analyze neuronal activity from an information-energy efficiency standpoint, Berger and Levy calculated the maximum Shannon mutual information transfer per unit of energy expenditure of an idealized integrate-and-fire (IIF) neuron whose excitatory synapses all have the same weight. Here, we extend their IIF model to a biophysically more realistic one in which synaptic weights are unequal. Using information theory, random Poisson measures, and the maximum entropy principle, we show that the probability density function (pdf) of interspike interval (ISI) duration induced by the bits per joule (bpj) maximizing pdf  $f_\Lambda(\lambda)$  of the excitatory postsynaptic potential (EPSP) intensity remains equal to the delayed gamma distribution of the IIF model. We then show that, in the case of unequal weights,  $f_\Lambda(\cdot)$  satisfies an inhomogeneous Cauchy-Euler equation with variable coefficients for which we provide the general solution form.

## I. INTRODUCTION

The human brain, only two percent of the body's weight, accounts for twenty percent of the body's energy consumption[2][3]. Brains have evolved that prodigiously compute and communicate information with remarkable efficiency. Energy minimization subject to functional constraints is a unifying principle[4]. Information theory often has been employed in neuroscientific data interpretation and system analysis during the last fifty years [5][6] [7][8]. However, energy-efficient neural codes have been studied for less than two decades [9][10][11]. Evidence for energy efficiency has been reported for ion channels [4], photoreceptors [12], retina [13] [14], and cortex [15] [16]. Laughlin and Sejnowski discussed the cortex as a communicating network from an energy efficiency perspective [17]; Mitchison et al. and Chklovskii et al. applied energy efficiency to analyze cortical wiring and brain mapping [18][19]; Berger and Levy proposed an energy efficient mathematical model for information transmission by a single neuron [1].

The goal of this paper is to find the capacity-achieving input distribution of incoming EPSP intensity for an extended Berger-Levy model in which their synapses have unequal weights, and the distribution of the output ISI durations that results from said capacity-achieving input distribution. Our main and perhaps striking result is that the same gamma distribution of ISI durations that Berger and

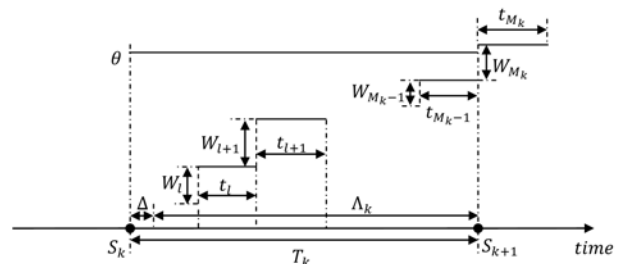


Figure 1. Single interspike interval (ISI) schematic with the illustration of all physical parameters.

Levy found resulted from the energy constrained capacity-achieving input distribution when the synaptic weights all were assumed to be equal continues to apply when these weights are allowed to be unequal. Also, we show that the optimal EPSP intensity for the case of unequal synaptic weights, which differs from its fixed-weight counterpart, is the solution of an inhomogeneous Cauchy-Euler equation with variable coefficients.

We first introduce a mathematical framework for how a single neuron stochastically processes and communicates information. Let us call the cortical neuron being modeled “neuron  $j$ ”, or simply “ $j$ ”. Let  $\underline{W}_k = (W_1, \dots, W_{M_k})$ , where  $W_l$  is the weight of  $l$ th synapse of  $j$  to receive a spike during the  $k$ th ISI and produce an EPSP in response thereto. We time-order the EPSP's according to the times at which they arrive at the somatic membrane and hence contribute to  $j$ 's PSP accumulation.  $M_k$  is the integer-valued random cardinality of  $\underline{W}_k$ .

We inherit from the B-L neuron model [1] the assumption that the only synaptic responses that  $j$  sums when generating its spikes are those of its excitatory synapses, the net strength of  $j$ 's inhibitory inputs serving only to scale this summation. However, we no longer assume that all of  $j$ 's excitatory synapses have the same weight. Rather, we assume that the components of  $\underline{W}_k$  are chosen independent and identically distributed (i.i.d.) according to a certain cumulative distribution function (cdf)  $F_W(w) = P[W \leq w], 0 \leq w < \infty$ . We model the  $l$ th contribution to  $j$ 's EPSP accumulation to be  $W_l \cdot u(t - t_l)$ ; here,  $W_l$  is a random variable (r.v.) with the aforementioned

cdf  $F_W(w)$  and  $u(t - t_i)$  equals 1 for  $t \geq t_i$  and equals 0 for  $t < t_i$ . We continue to assume as in [1] that each of  $j$ 's refractory periods has the same duration,  $\Delta$ . Since this theoretical extension connects the plasticity of a neuron's synaptic weights, widely considered essential to learning and memory, with the communication of information by the neuron, it may possess considerable practical significance [20].

Slightly differently from [1], we model the EPSP's produced in response to spikes from  $j$ 's afferent cohort as an inhomogeneous Poisson measure with instantaneous rate function,  $A(t)$ , defined by

$$A(t) := \lim_{dt \rightarrow 0} \frac{P[\text{one EPSP arrival in}(t, t + dt)]}{dt}. \quad (1)$$

Then as in [1] we take a time average operation over the rate function  $A(t)$  and obtain

$$\Lambda_k = \frac{1}{T_k - \Delta} \int_{S_{k-1} + \Delta}^{S_k} A(u) du, \quad (2)$$

where  $\Delta$  is the duration of  $j$ 's refractory period,  $T_k$  is the  $k$ th ISI duration of  $j$  and  $S_k = T_1 + \dots + T_k$ .

Henceforth, we suppress the ISI index  $k$  and just write  $\underline{W}$ ,  $M$ ,  $T$  and  $\Lambda$ . Thus, when  $\Lambda = \lambda$ , EPSP's are being assumed to arrive according to a homogeneous Poisson process with intensity  $\lambda$ .

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

$$I := \lim_{N \rightarrow \infty} \frac{1}{N} I(\Lambda_1, \dots, \Lambda_N; T_1, \dots, T_N), \quad (3)$$

lacking only an information decrement that addresses correlation among successive  $\Lambda_i$ 's; see [1].

Since  $T - \Delta$  is a one-to-one function of  $T$ , we have  $I(\Lambda; T) = I(\Lambda; T - \Delta)$ , which in turn is defined [21] [22] as

$$I(\Lambda; T - \Delta) = E \log \left[ \frac{f_{T-\Delta|\Lambda}(T - \Delta|\Lambda)}{f_{T-\Delta}(T - \Delta)} \right], \quad (4)$$

where the expectation is taken with respect to the joint distribution of  $\Lambda$  and  $T$ .

Toward determining  $I(\Lambda; T)$ , we proceed to analyze  $f_{T-\Delta|\Lambda}(t|\lambda)$  and  $f_{T-\Delta}(t)$  in cases with unequal synaptic weights.

## II. NATURE OF THE RANDOMNESS OF WEIGHT VECTORS

Even if the excitatory synaptic weights of neuron  $j$  were known,  $\underline{W} = (W_1, \dots, W_M)$  would still be random because the time-ordered vector  $\underline{R}$  of synapses excited during an ISI is random. However, for purposes of mathematical analysis of neuron behavior it is not fruitful to restrict attention to a particular neuron with a known particular set of synaptic weights. Rather, it is more useful to think in terms of the histogram of the synaptic weight distributions of neurons in whatever neural region is being investigated. When many such histograms have been ascertained, if their shapes almost all resemble one another closely, then they can be

arithmetically averaged to obtain a population histogram with fine resolution in the weights of its synaptic bins. This, in turn, would permit one to approximate this fine histogram by a continuous amplitude probability distribution of synaptic weights. (Then the analysis becomes more widely applicable than if one were to have used the exact weights of a particular neuron, especially considering that even that neuron will have a different set of weights in the future because of ongoing synaptic modification.) Moreover, the strong similarity of the synaptic weight distributions has been observed through experiments.[20] Therefore, in the analysis that follows we take the view that the components of  $\underline{W}$  are selected randomly from this continuous amplitude probability distribution. Said random distribution of synaptic weights also incorporates the random number of neurotransmitter-containing vesicles that are released when a spike is afferent to the synapse, the random number of neurotransmitter molecules in these vesicles, how many of those cross the synaptic cleft, bind to receptors and thereby generate EPSP's.

This model of random selection of weights comprising  $\underline{W}$  is applicable both to ISI's in which the afferent firing rate  $\Lambda$  is large and to those in which it is small. When the value  $\lambda$  assumed by  $\Lambda$  is large,  $\underline{W}$ 's components just get selected more rapidly than when  $\lambda$  is small, but they continue to come from the same distribution. This implies that the expected number of them in a single ISI remains the same. Hence, from now on, we assume that the weight vector components  $W_1, \dots, W_M$  are jointly independent of  $\Lambda$ .

## III. FINDING $f_{T-\Delta|\Lambda}(t|\lambda)$ : MIXTURES OF GAMMA DISTRIBUTIONS

We denote the spiking threshold by  $\theta$ . The contribution to  $j$ 's PSP accumulation attributable to the  $l$ th afferent spike during an ISI will be assumed to be a randomly weighted step function  $W_l \cdot u(t - t_l)$ , where  $t_l$  is the time at which it arrives at the postsynaptic membrane.<sup>1</sup>

It follows that the probability  $P_m = P(M = m)$  that exactly  $m$  PSP's are afferent to  $j$  during an ISI is

$$P_m = P(W_1 + \dots + W_{m-1} < \theta, W_1 + \dots + W_m \geq \theta). \quad (5)$$

Next, we write

$$\begin{aligned} & P(t \leq T - \Delta \leq t + dt | \Lambda = \lambda) \\ &= \sum_{m=1}^{\infty} P(t \leq T - \Delta \leq t + dt, M = m | \Lambda = \lambda) \\ &= \sum_{m=1}^{\infty} P_m \cdot P(t \leq T - \Delta \leq t + dt | \Lambda = \lambda, M = m), \quad (6) \end{aligned}$$

where (6) holds because of the assumption that  $\underline{W} = (W_1, \dots, W_M)$  is independent of  $\Lambda$ , which implies its random cardinality,  $M$ , is independent of  $\Lambda$ .

<sup>1</sup>In practice,  $u(t - t_l)$  needs to be replaced by a  $g(t - t_l)$ , where  $g(\cdot)$  looks like  $u(\cdot)$  for the first 15 or so ms but then begins to decay. This has no effect when  $\lambda$  is large because the threshold is reached before this decay ensues. For small-to-medium  $\lambda$ 's, it does have an effect but that could be neutralized by allowing the threshold to fall with time in an appropriate fashion. There are several ways to effectively decay the threshold, one being to decrease the membrane conductance.

It follows as in [1] that, given  $M = m$  and  $\Lambda = \lambda$ ,  $T - \Delta$  is the sum of  $m$  i.i.d. exponential r.v.'s with parameter  $\lambda$ , i.e., a gamma pdf with parameters  $m$  and  $\lambda$ . Summing over all the possibilities of  $M$  and letting  $dt$  become infinitesimally small, we obtain

$$f_{T-\Delta|\Lambda}(t|\lambda) = \sum_{m=1}^{\infty} P_m \cdot \frac{\lambda^m t^{m-1} e^{-\lambda t}}{(m-1)!} u(t). \quad (7)$$

It is impossible to determine  $P_m$  in the general case. However, we have been able to compute it exactly in the special case of an exponential weight distribution.

#### A. Case: Exponential Weight Distribution

Suppose the components of the weight vector are i.i.d. and have the exponential pdf  $\alpha e^{-\alpha w_i}$ ,  $w_i \geq 0$  with  $\alpha > 0$ .

Then we know that  $Y_m := W_1 + \dots + W_m$  has the gamma pdf

$$f_{Y_m}(y_m) = \frac{\alpha^m y^{m-1} e^{-\alpha y}}{(m-1)!} u(y).$$

So, referring to (5), we have

$$\begin{aligned} P_m &= P(Y_{m-1} < \theta, W_m \geq \theta - Y_{m-1}) \\ &= \int_0^\theta f_{Y_{m-1}}(u) du \int_{\theta-u}^\infty f_{W_m}(v) dv \\ &= \frac{(\alpha\theta)^{(m-1)}}{(m-1)!} e^{-\alpha\theta}. \end{aligned} \quad (8)$$

Therefore, it follows from (7) that

$$\begin{aligned} f_{T-\Delta|\Lambda}(t|\lambda) &= \sum_{m=1}^{\infty} \frac{(\alpha\theta)^{m-1}}{(m-1)!} e^{-\alpha\theta} \cdot \frac{\lambda^m t^{m-1} e^{-\lambda t}}{(m-1)!} u(t) \\ &= \lambda e^{-(\alpha\theta+\lambda t)} \sum_{k=0}^{\infty} \frac{(\alpha\theta\lambda t)^k}{(k!)^2} u(t). \end{aligned} \quad (9)$$

The summation in (9) equals  $I_0(2\sqrt{\alpha\theta\lambda t})$  where  $I_0$  is the modified Bessel function of the first kind with order 0 [23].

#### IV. $T - \Delta$ IS GAMMA DISTRIBUTED

For the conditional pdf  $f_{T-\Delta|\Lambda}(t|\lambda)$  as in (7), letting  $X = \lambda(T - \Delta)$ , we have the following equality

$$|f_{X|\Lambda}(x|\lambda) dx| = |f_{T-\Delta|\Lambda}(t|\lambda) dt|$$

in which  $x = \lambda t$ . It follows, in view of (7), that

$$f_{X|\Lambda}(x|\lambda) = \sum_{m=1}^{\infty} P_m \cdot \frac{x^{m-1} e^{-x}}{(m-1)!}, 0 \leq x < \infty. \quad (10)$$

Note that  $\lambda$  not only doesn't explicitly appear on the right-hand side of Eq. (10) but also does not appear there implicitly within any of the  $P_m$ 's; this is because, as noted earlier,  $M$  is independent of  $\Lambda$ , so  $P_m = P(M = m)$  cannot be  $\lambda$ -dependent. Accordingly,

$$f_{X|\Lambda}(x|\lambda) = f_X(x) = \sum_{m=1}^{\infty} P_m \cdot \frac{x^{m-1} e^{-x}}{(m-1)!}, x \geq 0. \quad (11)$$

Hence, although  $X$  equals  $\Lambda(T - \Delta)$ ,  $X$  nonetheless is independent of  $\Lambda$ .<sup>2</sup> We can rewrite the relationship as

$$T - \Delta = \frac{1}{\Lambda} \cdot X, \quad (12)$$

where  $X$  is marginally distributed according to Eq. (11).

Then by taking logarithms in Eq. (12), we have

$$\log(T - \Delta) = -\log \Lambda + \log X, \quad (13)$$

We see that (13) describes a channel with additive noise that is independent of the channel input. Specifically, the output is  $\log(T - \Delta)$ , the input is  $-\log \Lambda$ , and the additive noise is  $N := \log X$ , which is independent of  $\Lambda$  (and therefore independent of  $-\log \Lambda$ ) because  $X$  and  $\Lambda$  are independent of one another.

The mutual information between  $\Lambda$  and  $T - \Delta$  thus is

$$\begin{aligned} I(\Lambda; T - \Delta) &= I(\log \Lambda; \log(T - \Delta)) \\ &= h(\log(T - \Delta)) - h(\log(T - \Delta) | \log \Lambda) \\ &= h(\log(T - \Delta)) - h(N). \end{aligned} \quad (14)$$

Letting  $Z = \log(T - \Delta)$ , we have

$$h(\log(T - \Delta)) = h(Z) = -E \log f_Z(Z).$$

Since  $f_Z(z) = f_{T-\Delta}(t) |dt/dz| = f_{T-\Delta}(t) \cdot t$ , it finally follows that

$$\begin{aligned} h(\log(T - \Delta)) &= -E[\log(f_{T-\Delta}(T - \Delta) \cdot (T - \Delta))] \\ &= h(T - \Delta) - E \log(T - \Delta). \end{aligned} \quad (15)$$

Thus, after substituting Eq. (15) into Eq. (14) and adding the information decrement term as in [1], the long term average mutual information rate,  $I$ , we seek to maximize over the choice of  $f_\Lambda(\cdot)$  is

$$I = h(T - \Delta) - h(N) + (\kappa - 1)E \log(T - \Delta) - C. \quad (16)$$

Define  $T_- := T - \Delta$ ; then Eq. (16) becomes

$$I = h(T_-) + (\kappa - 1)E \log(T_-) - h(N) - C. \quad (17)$$

Since  $N$  is independent of  $\Lambda$ , the choice of  $f_\Lambda(\cdot)$  affects  $I$  in Eq. (17) only through  $f_{T_-}(\cdot)$  via the following equation

$$\int_0^\infty d\lambda f_\Lambda(\lambda) f_{T_-|\Lambda}(t|\lambda) = f_{T_-}(t), 0 \leq t < \infty. \quad (18)$$

Therefore, our bjp maximizing task has been reduced to maximizing the entropy rate  $h(T)$  subject to Lagrange multiplier constraints on information decrement,  $E \log T$ , and energy constraint  $ET$  which can be written as  $C_0(1 + bT)$  [1]. Taking  $C_0$  as the energy unit, we have  $b$  as the coefficient of  $ET$ . It is known that the entropy  $h(T)$  is maximized subject to constraints on  $E \log T$  and  $ET$  when

<sup>2</sup>A simple example in which  $X = AB$  is independent of  $A$  may be enlightening here. Let  $P(A = -1) = P(A = 1) = 1/2$ ,  $P(B = -2) = P(B = -1) = P(B = 1) = P(B = 2) = 1/4$ , and assume  $A$  and  $B$  are independent. Note that, given either  $A = -1$  or  $A = 1$ ,  $X$  is distributed uniformly over  $\{-2, -1, 1, 2\}$ , so  $X$  is independent of  $A$ .  $X$  is not independent of  $B$  in this example because  $|B| = 2$  implies  $|X| = 2$ , whereas  $|B| = 1$  implies  $|X| = 1$ . In our neuron model the two factors of  $X$ , namely  $\Lambda$  and  $T - \Delta$ , are not independent of one another but rather are strongly negatively correlated.

$T$  is a delayed gamma distribution with two parameters we denote by  $\kappa$  and  $b$ , i.e.,

$$f_{T_-}(t) = \left[ \frac{b^\kappa t^{\kappa-1} e^{-bt}}{\Gamma(\kappa)} \right] u(t), \quad (19)$$

which is the bpj-maximizing distribution of neuron  $j$ 's ISI duration,  $T$ , sans its refractory period which always has duration  $\Delta$ .

## V. FROM THE INTEGRAL EQUATION TO THE DIFFERENTIAL EQUATION

The integral equation below relates the following three quantities:

- 1) The to-be-optimized pdf  $f_\Lambda(\lambda)$  of the arithmetic mean of the net afferent excitation of neuron  $j$  during an ISI.
- 2) The conditional pdf  $f_{T_-|\Lambda}(t|\lambda)$  of  $j$ 's encoding of  $\Lambda$  into the duration  $T$  of said ISI.
- 3) The long term pdf  $f_{T_-}(t)$  of  $j$ 's ISI duration.

$$\int_0^\infty d\lambda f_\Lambda(\lambda) f_{T_-|\Lambda}(t|\lambda) = f_{T_-}(t), 0 \leq t < \infty. \quad (20)$$

Moreover, we have

$$f_{T_-|\Lambda}(t|\lambda) = \sum_{m=1}^{\infty} P_m \cdot \frac{\lambda^m t^{m-1} e^{-\lambda t}}{(m-1)!}, \quad (21)$$

and

$$f_{T_-}(t) = \left[ \frac{b^\kappa t^{\kappa-1} e^{-bt}}{\Gamma(\kappa)} \right] u(t). \quad (22)$$

Let's first consider a case in which  $P_m$  is nonzero only for two consecutive values of  $m$ , say  $n$  and  $n+1$  with respective probabilities  $P_n = p$  and  $P_{n+1} = 1-p := q$ . In such a case

$$\int_0^\infty d\lambda f_\Lambda(\lambda) \left( \frac{p\lambda^n t^{n-1}}{(n-1)!} + \frac{q\lambda^{n+1} t^n}{n!} \right) e^{-\lambda t} = f_{T_-}(t). \quad (23)$$

Substituting Eq. (22) into Eq. (23) and applying integration by parts and inverse Laplace transform on Eq. (23), it follows that

$$(1 + q/n)f_\Lambda(\lambda) + (\lambda q/n)f'_\Lambda(\lambda) = \lambda^{-n} \frac{\Gamma(n)}{\Gamma(\kappa)\Gamma(n-\kappa)} b^\kappa (\lambda - b)^{n-\kappa-1} u(\lambda - b). \quad (24)$$

Eq. (24) constitutes a conversion of the integral equation (23) for the maximum-bpj excitation intensity  $f_\Lambda(\lambda)$ , into a first-order linear differential equation. The differential equation has a  $\lambda$ -varying coefficient on its  $f'$  term. Nonetheless, it has an explicit analytical solution because there exists a general solution to any inhomogeneous first-order linear differential equation with variable coefficients.

In general, Eq. (24) turns out to be an inhomogeneous Cauchy-Euler equation with variable coefficients as Eq. (25).

$$\begin{aligned} & \sum_{j=0}^{n-1} \sum_{i=0}^j P_{j+1} \frac{(j+1)!}{(j-i+1)! i!(j-i)!} \lambda^{(j-i)} f_\Lambda^{(j-i)}(\lambda) \\ &= \frac{b^\kappa}{\Gamma(\kappa)\Gamma(1-\kappa)} \lambda^{-1} (\lambda - b)^{-\kappa} u(\lambda - b). \end{aligned} \quad (25)$$

Eq. (25) also has an analytical closed-form solution, which serves as the bpj-maximizing pdf of neuron  $j$ 's averaged afferent excitation intensity  $\Lambda$ .

## VI. CONCLUSION

We have shown that, when neuron  $j$  is designed to maximize bits conveyed per joule expended, even though  $j$ 's synapses no longer are being required to all have the same weight, the pdf of the ISI durations continues to be exactly the same gamma pdf as it was in [1] wherein all the weights were assumed to be equal. This happens despite the fact that the conditional distribution for  $T$  given  $\Lambda$  is now a mixture of gamma distributions instead of the pure gamma distribution that characterizes the special case of equal weights.

Additionally, we have implicitly determined the optimal distribution  $f_\Lambda(\lambda)$  that characterizes the afferent excitation intensity by (1) maximizing the Shannon mutual information rate given a constraint on the total energy cost that a cortical neuron expends for metabolism, postsynaptic potential accumulation, and action potential generation and propagation during one ISI; (2) converting the integral equation to a differential equation with a closed-form solution.

The energy efficiency of the human brain in terms of information processing is astonishingly superior to that of man-made machines. By extending the Berger-Levy information-energy efficient neuron model to an unequal synaptic weights case, the theory comes into closer correspondence with the actual neurophysiology of cortical networks, which might pave the way to wider applications in neuroscience and engineering.

## REFERENCES

- [1] T. Berger and W. B. Levy, *A mathematical theory of energy efficient neural computation and communication*, IEEE Trans. IT, vol. 56, No. 2, pp. 852-874, February 2010.
- [2] J. M. Kinney, H. N. Tucker, Clintec International Inc, *Energy Metabolism: Tissue Determinants and Cellular Corollaries* (Raven, New York), p xvi, 1992.
- [3] L. C. Aiello, P. Wheeler, *The expensive-tissue hypothesis—the brain and the digestive-system in human and primate evolution*, Curr Anthropol 36:199-221,1995.
- [4] A. Hasenstaub, S. Otte, E. Callaway, and T. J. Sejnowski, *Metabolic cost as a unifying principle governing neuronal biophysics*, Proc Natl Acad Sci USA 107: 12329-12334, 2010.
- [5] D. M. McCulloch, W. S. McCulloch, *The limiting information capacity of a neuronal link*, Bulletin of Mathematical Biophysics, vol. 14, 127-135, 1952.
- [6] W. S. McCulloch, *An upper bound on the informational capacity of a synapse*, In Proceedings of the 1952 ACM national meeting, Pittsburgh, Pennsylvania.
- [7] R. R. de Ruyter van Steveninck, W. Bialek, *Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences*, Proceedings of the Royal Society Series B, Biological Sciences, 234, 379-414, 1988.
- [8] W. Bialek, F. R. Rieke, R. R. de Ruyter van Steveninck and D. Warland, *Reading a neural code*, Science, 252,1854-7, 1991.

- [9] W. B. Levy, R. A. Baxter, *Energy efficient neural codes*, Neural Comput 8:531-543, 1996.
- [10] W. B. Levy, R. A. Baxter, *Energy-efficient neuronal computation via quantal synaptic failures*, J Neurosci 22:4746-4755, 2001.
- [11] V. Balasubramanian, D. Kimber, M. J. Berry, *Metabolically efficient information processing*, Neural Comput 13:799-815, 2001.
- [12] J. E. Niven, I. C. Anderson, S. B. Laughlin, *Fly photoreceptors demonstrate energy-information trade-offs in neural coding*, PLoS Biol 5:e116, 2007.
- [13] V. Balasubramanian, M. J. Berry, *A test of metabolically efficient coding in the retina*, Network 13:531-552, 2002.
- [14] S. B. Laughlin, R. R. de Ruyter van Steveninck, J.C. Anderson, *The metabolic cost of neural information*, Nat Neurosci 1:36-41, 1998.
- [15] D. Attwell, S. B. Laughlin, *An energy budget for signaling in the grey matter of the brain*, J Cereb Blood Flow Metab 21:1133-1145, 2001.
- [16] B. D. Willmore, J. A. Mazer, J. L. Gallant, *Sparse coding in striate and extrastriate visual cortex*, J Neurophysiol, 105(6):2907-19, 2011.
- [17] S. B. Laughlin, T. J. Sejnowski, *Communication in neuronal networks*, Science 301:1870, 2003.
- [18] G. Mitchison, *Neuronal branching patterns and the economy of cortical wiring*, Proc Biol Sci 245:151-158, 1991.
- [19] D. B. Chklovskii, A. A. Koulakov, *Maps in the brain: What can we learn from them*, Annu Rev Neurosci 27:369-392, 2004.
- [20] B. Barbour, N. Brunel, V. Hakim, J. Nadal, *What can we learn from synaptic weight distributions?*, Trends Neurosci 30(12): 622-629, 2007.
- [21] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.
- [22] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.
- [23] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Edited by A. Jeffrey and D. Zwillinger. Academic Press, New York, 7th edition, 8.447, p. 961, 2007.