Neural computation from first principles: Using the maximum entropy method to obtain an optimal bits-per-joule neuron

William B Levy¹, Toby Berger², Mustafa Sungkar²

¹Dept.’s of Neurosurgery and of Psychology, School of Medicine, University of Virginia, Charlottesville, VA 22908
²Dept. Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22903

Index Terms— energy limited, communication limited, neural computation, optimal computation

ABSTRACT
Optimization results are one method for understanding neural computation from Nature’s perspective and for defining the physical limits on neuron-like engineering. Earlier work looks at individual properties or performance criteria and occasionally a combination of two, such as energy and information. Here we make use of Jaynes’ maximum entropy method and combine a larger set of constraints, possibly dimensionally distinct, each expressible as an expectation. The method identifies a likelihood-function and a sufficient statistic arising from each such optimization. This likelihood is a first-hitting time distribution in the exponential class. Particular constraint sets are identified that, from an optimal inference perspective, justify earlier neurocomputational models. Interactions between constraints, mediated through the inferred likelihood, restrict constraint-set parameterizations, e.g., the energy-budget limits estimation performance which, in turn, matches an axonal communication constraint. Such linkages are, for biologists, experimental predictions of the method. In addition to the related likelihood, at least one type of constraint set implies marginal distributions, and in this case, a Shannon bits/joule statement arises.
I. INTRODUCTION

First principle approaches for understanding neural communication and computation are rare but possible with careful selection of constraints combined with an appropriate optimization method[1-8].

Energy-efficient sensing, computation, and communication are among these optimization schemes. The biological motivation for energy-based constraints is strong since (i) it is a common currency across every physiological levels — subcellular, cellular, organ-level, and even to an organism's behavior and (ii) it is generically relatable to biological fitness and reproductive success. In other words, quantified, energy-constrained optimizations move us closer to Darwin's paradigm reduced to mathematical models. From another direction, there are strong arguments from physics that computation is ultimately limited by available energy. Thus when successfully executed, the energy-constrained, optimization approach delivers the ultimate limits on engineered neural-like computation.

Earlier work pursuing bit/J optimizations use Shannon's mutual information exclusively. Here the ultimate goal is no different; however, the method is enhanced via a preliminary optimization. This enhancement expands the definition of a neuron and its costs in two important ways. First, it makes explicit the computation performed by a neuron, and second, it recognizes certain limits on pulse-based communication. The enhancement is the maximum-entropy method (MEM) — a conceptualization of probabilistic inference that arises nine years after Shannon's seminal work, and a conceptualization that intersects with Shannon theory over the last sixty years. After applying the MEM to find the optimal likelihood for the required estimation, which also turns out to be optimized, we return to a mutual information calculation explaining how to obtain a neuronal bits/J statement.

It is generally agreed that a neuron performs a computation. However in any specific situation, the exact nature of such a computation is a matter of conjecture. Here we conjecture that a neuron estimates the value of a scalar, latent variable, and we derive a family of optimal-inference probability densities for such an estimate. In other words, instead of viewing a neuron as a communication device [1,2,4,5,9], which optimizes mutual information just like a communication channel, the hypothesis here is that a neuron's estimate corresponds to the maximum entropy distribution constrained by a set of expectation constraints that define and/or limit the neuron's operation. Justification of the maximum entropy method as a way to discover the best density for inference has a lengthy history with the essential, pioneering work due to Jaynes [10,11]. Of four justifications claiming that the MEM finds the best density, perhaps the strongest is game-theoretic. (Three others are (i) axiomatic derivations [12-16], (ii) the entropy concentration of possible outcomes in which the best density is the one that is consistent with an exponentially vast number of the outcomes [17,18], and (iii) the Jaynes' perspective, use all the constraints that are credible but do not contradict oneself by using a distribution requiring additional constraints that one believes not to exist [19], i.e., logical consistency is superior to logical inconsistency). Via game theory, there are two salutary effects of a MEM inference: (i) this density maximizes the long-term, average pay-off while (ii) minimizing the maximum chance of going broke [20].

Another novelty here is the heterogeneous dimensionality of the constraint set. For example, such a set might include an energy-budget, an estimation-bias constraint, and a communication constraint. Perhaps counter-intuitively, this set need not include a minimum mean square error constraint on the estimation; instead, this minimum is inferred through the bias/non-bias constraint(s) and the energy-budget. More importantly, our MEM-inferred probability distributions are in the exponential class, implying the information-equivalence between the estimate itself (a minimal sufficient statistic of fixed dimension) and the inferred conditional probability distribution, i.e., a particular likelihood statement suitable for a Bayesian inference [21]. Given that Nature – via natural selection, internal stability, and enough time – implicitly incorporates prior probabilities into portions of an organism (the liver, the kidney, a neuron, the cerebellum); then, by combining this likelihood with the appropriate prior distribution, a neuron can perform a Bayesian inference and implicitly communicate this inference via transmitting a sufficient statistic.

In what follows, the general theory is developed in Sections II and III. Section IV introduces specific constraints deemed as basic neuroscience or as biophysically appealing, followed by the use of these
constraints in the three examples of Section V. This section points that two common neuron models can be used to solve and communicate a particular MEM-derived optimal inference while a third common model (the drifted Ornstein-Uhlenbeck model with fixed starting point and fixed threshold) cannot (see Section VII). For some, Section VI maybe the high point of the developments. This section explains how and why the third example of Section V implies the exact bits-per-joule performance of a neuron built to satisfy the constraint set of this third example.

II. PRIMARY ASSUMPTIONS

As in [5,9], the critical pair of assumptions here concern the input and output variables, \{A,T\}.

\textbf{A0.} A neuron, call it \textit{j}, communicates to other neurons (its direct recipients) with pulses. It begins an interpulse interval (IPI) with an initial excitation-state of zero; it sums its synaptic activations, charge-injections on to the membrane capacitor, until a threshold value, \(\theta\), is achieved. At this point it produces a pulse-out and then resets to zero, beginning the cycle anew. Note that the IPI is a first-hitting time (FHT). The neuron has an assigned, on-average, energy budget, \(E\). The pulse-out is of fixed amplitude and fixed shape regardless of the IPI. Moreover, assume that each recipient of a pulse must instantaneously decode the pulse as a particular charge-injection of short duration.

\textbf{Corollary to A0.} Assumption A0 implies that the only information available to a recipient is arrival-time and hence, the IPI, assuming an available timer. Thus, the output random variable (RV) is IPI \(T:=T_j(k)=t\) (here \(k\) indexes the IPI number within a longer, system-wide decision-cycle; upper case is used for RVs, lower case for any realization).

\textbf{A1.} Assume that the computational task of \(j\) is to estimate the value of a latent (hidden) RV, \(\Lambda:=\Lambda_j(k)=\lambda\). This variable is the rate of the union of point-processes governing each input \(i\). Assume that any input value, \(\Lambda=\lambda\), is stationary throughout \(j\)'s IPI and that there is Markovian conditional independence for successive IPIs with the necessary and locally available conditioning information stored by each of \(j\)'s recipients.

III. OPTIMIZING THE NEURON'S COMPUTATION

Given the above assumptions, the maximization of the differential entropy, \(h(T \mid \lambda)\), over \(p(t \mid \lambda)\) begins with a general statement of the Lagrange-multiplier equation:

\[ \mathcal{L} = h(T \mid \lambda) + \alpha_0(1 - \int p(t \mid \lambda) \, dt) + A + B + C, \]

where, beyond normalization, there are three kinds of constraints, \((A)\) budgeted energy, \((B)\) estimation restrictions or requirements, and \((C)\) communication constraints. Each constraint has a linear relationship with one or more expectations through its undefined multiplier \(\alpha\):

1. \(A := \alpha(\mathcal{E}_\infty - c\lambda E[T \mid \lambda]) + \sum_a^\infty \alpha_a(E_j - E[g_j(T)])\)
2. \(B := \sum_a^\infty \alpha_a(E_j - f_j(\lambda)E[g_j(T) \mid \lambda])\)
3. \(C := \sum_a^\infty \alpha_a(E_j - E[g_j(T)])\)

where the bold-script letters are constants and where the subscripted \(\mathcal{E}\)'s are the energy-budgets for the individual expectations, and there is a constant of energy-use, \(c_0\), accounting for the cost of pulse-generation and communication implicitly incorporated into the budget as \(\mathcal{E} = c_0 + \mathcal{E}_\infty + \sum_{k=2}^{m_k} \mathcal{E}_n\), or equivalently, \(\mathcal{E}_0 = \mathcal{E}_\infty + \sum_{k=2}^{m_k} \mathcal{E}_n\). For \(\lambda\)-conditional expectations, the expectation operator is constructed by integration over \(p(t \mid \lambda)\, dt\) with \(\lambda\) fixed; since an axon-communication constraint must be insensitive to \(\lambda\), i.e., an unconditional expectation, such an expectation is understood as \(\int p(\lambda) p(t \mid \lambda) g(t) \, dt \, d\lambda\).

Moreover, all expectations are finitely valued; energy-costs are positive; and each summation specifies a finite number of terms. Then take the derivative \(\frac{\partial}{\partial p(t \mid \lambda)}\) and re-arrange,
\( p(t | \lambda) = \exp(-\alpha_0 - 1 - \alpha_t \sum_b^{n_b} f_{a,b}(\lambda) g_{a,b}(T) - \sum_b^{n_b} \alpha_{b,h} f_{b,h}(\lambda) g_{b,h}(T) - \sum_b^{n_b} \alpha_{c,b} p(\lambda) g_{c,b}(T)) \)

That is, the optimal inference is in the exponential family of densities, implying at least two important conclusions. When a RV outcome of an experiment, or the likelihood from sampling, is in the exponential class, there exists a sufficient statistic of fixed and finite dimension [22, 23]. Indeed, such a statistic only exists when the distributional form is in the exponential class. From the neural perspective this is good news since it becomes possible to communicate all the information of the inference in this sufficient statistic (which certainly is minimal in the examples since its dimension is one) and since this form of a likelihood is suitable for conjugate Bayesian inference. The examples will clarify these two claims.

To delve a little deeper into the constraints leading to the exponential family, consider \( B, C \), and then the \( A \) constraints, all are multiplicative (separable) in any pairing of functions of \( \lambda \) and of \( T \) without any joint, non-separable function of both variables. The \( B \) constraints are always separable since they arise from estimates such as \( E[g_b(T) | \lambda] = f_b(\lambda) \) or \( E[\log(c / T | \lambda)] = \log(\lambda) \); The \( C \) constraints are axonal communication properties containing no explicit \( \lambda \)-terms; although such terms appear after differentiation, they still factor. Finally, regarding the \( A \) constraints two points should be made. The restrictions upon the allowable constraints that Nature or an engineer might implement are mild and relatively general. \textit{A priori} constraint knowledge will not be a polynomial such as \( E_h = E[\lambda^2 T^2 + 3 \lambda^2 T^4 | \lambda] \), rather than a single such polynomial constraint, there will be a set of constraints corresponding to each term of the polynomial, e.g., \( \{E_{h_1} = E[\lambda^2 T^2 | \lambda], E_{h_2} = E[3 \lambda^2 T^4 | \lambda] \} \). Moreover, in the case of constraints that are purely a functions of time, these differentiate into the \( \lambda \)-conditional expectation constraints. However, it is only by assumption (as motivated by biophysical intuition) that unfactorable forms such as \( E[\exp(\lambda T)] \) and \( E[\sin(\lambda T)] \) are purposely eliminated from consideration. When all is said and done, the actual energy-constraint set that fits our biological intuition is rather modest (see below).

IV. Constraint equations in the form of expectation equalities

Having achieved the central result (1), we move from the general to the specific. Explicit models, i.e., hypotheses, create particular neural-like constructions and may demand particular input distributions. Here a model begins with a specific constraint set, arrives at the optimal \( p(t | \lambda) \), and then identifies a neuronal construction that achieves this \( p(t | \lambda) \) while obeying \( A0 \) and \( A1 \).

(Because of the dimensional inhomogeneity of the constraints, we try to be careful with the constants needed for homogeneity and sensibility. Whenever \( \lambda \) and \( t \) (or \( T \)) appear, their product is dimensionless but whenever either appears alone, it is multiplied by an appropriate constant: in the case of rate \( \lambda \) (events/sec), it is multiplied by seconds, \( b_o \); and in the case of \( t \) (or \( T \)), it is divided by \( b_o \) or by a scaled version of this constant \( b_o \). The undetermined multipliers are dimensionless except for any multiplier arising in the \( A \) constraints, which are inverse joules, \( J^{-1} \); \( c_o, c_T, c_o \) and the \( E \)’s are joules; \( c_{ax} \), \( c_{ext} \) and \( c_o \) are dimensionless.)

Energy constraints. Energy constraints (per-pulse calculations) take a limited number of forms: (i) a constant \( c_o \) for axonal communication costs, (ii) time-linear costs such as maintenance, axonal leak, and defrayed construction costs, (iii) synaptic activation costs which are proportional to \( \lambda \cdot t \), (iv) a time-logarithmic cost for re-setting timing-devices that tick with exponential spacings, and depending on the neuron model, (v) an inverse-time cost to account for discarded energy between IPIs. Here we choose to associate the time-linear costs with the decision-maker (DM) that controls the duration of a decision-making cycle in which \( f \) might generate many IPIs (such a DM must know these time-proportional costs to decide when additional system computation is not improving the decision's expected utility, which takes into account the energy-costs). Thus here, time-linear costs do not constrain \( f \)'s MEM and (v), inverse-time costs, are also ignored. The most complicated energy-constraint expression used here is

\( A = \alpha_o (E_{syn} - c_o, \lambda E(T | \lambda)) + \alpha_{log} (E_{log} - c_c E[log(T / b_o)]) \) with

\( E = c_o + E_{syn} + E_{log} \) or equivalently, \( E_o = E_{syn} + E_{log} \).
Estimation constraints. Although \( B \) displays a very general set of possibilities including multiple estimates and estimates with \( \lambda \)-dependent bias, the constraints with intuitive appeal are more limited, \( f_1(\lambda) = E[f_2(1/T)\lambda] + c \). Generically, only two estimation forms occur below: the simplest being estimation of \( \lambda \) by \( c/t \) and the less simple, estimation of \( \log(\lambda) \) by \( \log(c/t) \). However for each of these, several specific forms can arise differing by a biasedness condition, e.g., the biased \( E[\log(T/b_t)\lambda] = \log(\lambda b_t) + c \) or the unbiased \( E[c_{es}/T\lambda] = \lambda \) with \( c_{es} \) determined by the energy-budget (see examples).

Communication constraints. The final example contains a communication constraint or more accurately, conjectures an unconditional expectation that reflects a communication constraint. Specifically, the specification \( E[b_t/T] = c_{es} \), minimizes intersymbol interference (ISI) by making closely spaced pulses highly unlikely.

In sum, constraints associated with \( A \) and \( C \) arise physical properties while a \( B \) constraint is essentially an appropriate selection of the pair \( \{f_{bs}\lambda), g_{bs}(t)\} \) to reflect \( f \)'s statistically-defined computation. There is no general solution in regard to this last, just specific examples. Here no more than three \( g(t) \) functions will appear, \( t, 1/t, \log(t) \). As a result, the two most complicated, and to us, most interesting inferred densities are in the class of generalized inverse Gaussian (GIG) distributions [24,25][26,27], which, before solving for the undetermined multipliers (UMs) take the forms \( p(t|\lambda) = t^{\nu-1}\exp\left(-\alpha_{t}\lambda t - \alpha_{s}\lambda / (\lambda t)\right) \) and \( p(t|\lambda) = t^{\nu-1}\exp\left(-\alpha_{o}(\lambda t) / (\lambda t)\right) \).

V. THREE EXAMPLES OF PROGRESSIVE RELEVANCE AND COMPLEXITY

All examples include a constant-per-pulse cost and the linear synaptic cost, \( wE[T\lambda\lambda] \) where \( w \) is the cost of an average synaptic activation (which is \( \lambda \)-independent when the conditional Poisson approximation [26] is good). The first example shows that a particularly simple constraint-set (rather selectively) allows a known and simple neural model [27]. Importantly, here the model is inferred while earlier work, e.g., [5,9,27] assume the model. Each example begins with its set of constraining expectations.

Ex 1, \( \{E[T|\lambda], E[\log(T)\lambda]\} \). In this example all synaptic weights have the same, unit-value. The neuron's job is to estimate \( \log(b_t, \lambda) \). In addition to normalization, there are two constraints, both conditioned on \( \Lambda = \lambda \): (i) an energy-cost (just two terms), \( \{c_o, w\lambda E[T\lambda]\} \), i.e., \( E_o := \exp/E - c_o = \lambda wE[T\lambda] = \lambda \) and (ii) the estimation with bias, \( E[\log(b_t/T)|\lambda] = \log(b_t/\lambda) + \log(1/c_3) \), or as a constraint, \( c_3E[\log(1/T)|\lambda] = 0 \) (i.e., \( 1/c_3 = \log(b_t/b_t) \)). The general result after differentiating the Lagrange equation is

\[
p(t|\lambda) = \exp(-\alpha_{o}(1))(\lambda t)^{\nu_{o}-1}\exp(-\alpha_{o}w\lambda t) , \tag{2}
\]

a \( \Gamma \)-distribution with \( \exp(-\alpha_{o}(1)) = \lambda(\alpha_{o}w)^{-\nu_{o}+1} / \Gamma(\alpha_{o}c_3 + 1) \).

A model that produces this FHT distribution begins with A0, a point-process with a mean, stationary rate \( \lambda \) and, in this example, generates excitatory synaptic activations without inhibition or leak: that is, the neuron just adds synaptic excitatory synaptic events, all valued as one. The threshold for this neuron is \( \theta \geq 1 \), with \( -\alpha_{o}(1) = \theta - 1 \). Thus

\[
p(t|\lambda) = (w\alpha_{o}\lambda)^{\theta}(t)^{\nu_{o}-1}\exp(-w\alpha_{o}\lambda t) / \Gamma(\theta) , \tag{3}
\]

This model makes further calculations easy. Specifically, the cost of reaching threshold every IPI is the same; therefore this cost must equal the average cost of the synaptic activations per pulse, \( w\theta = w\lambda E[T\lambda] = \lambda c_0 / \lambda \), or \( \theta = \lambda E[T\lambda] = \lambda c_0 / w = E_o^{\nu_{o}} \). Then noting that the mean of the \( \Gamma \)-distribution (3) is \( E[T\lambda] = \theta / (w\alpha_{o}\lambda) \), implies \( \alpha_{o} = w^{-1} \). Therefore, (3) becomes

\[
p(t|\lambda) = \lambda^{\nu_{o}}(t)^{\nu_{o}-1}\exp(-\theta t) / \Gamma(\theta) , \tag{4a}
\]

\[
p(t|\lambda) = \lambda^{\nu_{o}}e^{\nu_{o}-1}\exp(-\theta t) / \Gamma(\nu_{o}) \tag{4b},
\]

June 2016 IEEE submission Levy, Berger, Sungkar
Interestingly, the estimation constraint here must be biased. Nonetheless, fixing \( \varepsilon_0 \) and \( \theta \), there is a unique optimal bias, making it removable by a recipient; moreover, this removable bias and, more importantly, the variance of the estimate go to zero as the energy-budget increases.

With a bias-constraint statement as \(-c_1 E[\log(T/\lambda)|\lambda] - 1 = 0\), one infers that \( 1/c_1 = -Di\Gamma(\theta) = -Di\Gamma(\varepsilon_0) \).

Regarding the variance of the estimate, \( \text{var} [\lambda|\lambda] = E[(\log(T/b) - \log(\lambda b))]^2 | \lambda] = E[(\log(T/b) - E[\log(T/b)|\lambda])^2 | \lambda] = T\text{ri} [\varepsilon_0^2] = \varepsilon^{-1}_0 \). To some it may seem odd that a nominally optimal statistical computation should be biased, but remember that the original variable of interest cannot be directly sampled nor, for the IPI scheme, encoded without a nonlinear transformation. In particular, an estimate of a transformed RV is often biased. A less strange form of the bias statement is possible so long as a recipient has knowledge of \( \theta \) (or \( \varepsilon_0 \)), and in the same vein, (iv) Bayesian inference by the recipient, using \( p(\lambda) \) and equipped with a prior \( p(\lambda) \) arising from the previous IPI, is possible.

Below, our primary exhibits are a little more biological since they allow for synaptic inhibition and differently valued synaptic weights. Specifically, the resulting likelihoods are GIGs, and both examples are satisfied for an input that is a Skellam or a Poisson process when \( a = -\frac{1}{2} \). That is, after finding the general form of the MEM-optimal inference for the given constraint set, we find the necessary parameterization of the constraint set that are consistent with these models. (In essence for the Skellam, there are two point processes, one for excitation, \( \lambda_{ex} \), and one that runs a subtractive inhibition \( \lambda_{in} \); moreover, we require that inhibition is a known, fixed fraction of excitation \( \lambda_{in} = \Delta \lambda_{ex} \), \( \Delta \in (0.05) \), implying \( \lambda_{net} = \lambda_{ex}(1-\Delta) \) and variance \( \lambda_{net}(1+\Delta) \) see [27] for such an approach). The relations can be expanded to include the average charge-energy-injection by an active synapse, \( w \) and its variance, \( \sigma_w^2 \), so that the rate is \( \lambda_{ex}(1-\Delta)w \) and the variance parameter is \( \lambda_{net}(1+\Delta)\sigma_w^2 \). After discovering the MEM-solution, the model is completed as in [28] (G&M): the synaptic events are small, numerous, the dominant noise-source, and, as a good approximation conditional on \( \lambda \), independent. First we start with the constraints of energy and estimation.

Ex 2, \{ \( E[1/\lambda], E[T], E[\log(T)] \} \). The neuron’s computational task is an unbiased estimate of \( \lambda_{net} \) (from here on just \( \lambda \)) using \( c_{est}/t \); that is, there is the single scalar-constant \( c_{est} := b_1 / b_2 \), and there is the unbiased estimation statement \( E[b_1 T^{-1} | \lambda] = \lambda b_2 \); or in constraint form, \( E[c_{est}(\lambda T)^{-1} | \lambda] = 1 \). The cost-constraints are (i) the synaptic energy-cost, \( wE[\lambda T | \lambda] = \varepsilon_{syn} \), (ii) a logarithmic time-cost, which, e.g., arises from clocking-costs with exponentially spaced clock-ticks, \( E[\log(T/b) | \lambda] = \varepsilon_{log}(\lambda) \) and \( E[\log(T/b_2) | \lambda] = \varepsilon_{log}(\lambda) \int \varepsilon_{log}(\lambda) d\lambda \), and (iii) \( c_0 \), the constant per-pulse cost of communication; in constraint form, the statements are \( wE[\lambda T | \lambda] - \varepsilon_{syn} = 0 \) and \( c_2 E[\log(c_2 T/\lambda) | \lambda] - \varepsilon_{log}(\lambda) = 0 \), and overall \( \varepsilon(\lambda) = c_0 + \varepsilon_{syn} + \varepsilon_{log}(\lambda) \). The solutions to the full constraint-set, \{ \( E[1/(\lambda T)\lambda], E[\lambda T\lambda], E[\log(T)\lambda] \} \), are (5) a special subclass of all the GIG distributions [29]. Specifically, with \( a < 0 \) and where \( \alpha_1 \) and \( \alpha_2 \) are numerical, where \( a-1 := -c_2 \alpha_1 \), with \( c_2 > 0 \) because energy-constraints are positive, and where \( K_\alpha(A) \) is the modified Bessel function of the second kind with index value \( a \), \( p(t | \lambda) = \lambda^\alpha(\alpha_1/\alpha_2) \exp[-\alpha_1 \lambda t - \alpha_2 / (\lambda t)] / 2K_\alpha(2\sqrt{\alpha_2 \alpha_1}) \).

Note that all such densities are in the exponential family and that these special GIG forms avoid \( \lambda \) in the Bessel function [29]. This special subclass arises because of the paired pairs, \( \lambda t \) and \( 1/\lambda t \) in the exponent. More fundamentally, this required double pairing arises because (i) the first \( \lambda t \) unavoidably appears through the synaptic energy-constraint while (ii) the inverse pairing arises from the unbiased nature of the computational goal, i.e., the estimation constraint \( E[1/T\lambda] \) is directly proportional to \( \lambda \). Moreover, if
there existed a further constraint that forced a \(\lambda\)-term into the Bessel function, then all the standard conditional moments of \(t\), including the estimation constraint, would contain a Bessel function of \(\lambda\) making explicit the self-contradictory nature of such a constraint set (see APPENDIX/REMARKS).

In general, the UMs can be solved numerically, but since one goal here is to display conditions for which the G&M neuron is MEM-optimal, we must suppose \(\alpha = -\frac{1}{2}\). That is, the resulting likelihood specializes to the inverse Gaussian distribution, 
\[ p(t; \lambda) = \frac{1}{2\sqrt{\pi} \lambda^{3/2}} t^{3/2} \exp(2\sqrt{\alpha_2 \lambda t} - \alpha_1 t - \alpha_2 / (\lambda t)) \]
with explicitly solvable UMs. Indeed, there is enough here to specify the parameters in the constraint equations using the assumption of independent, additive synaptic events provided by an underlying Skellam (or Poisson process). As in the FHT of a drifted B-M [30], choose the multiplier of the \(1/t\) term in the exponent as 
\[ \alpha_2 / \lambda = \theta^2 + (1 + \Delta) \sigma^2 \lambda , \]
threshold-squared divided by the variance of the diffusion approximation. Then using the linear additivity, the cost-constraint of synaptic activation is 
\[ w E[(\lambda T) \mid \lambda] = \theta + (1 - \Delta) . \]
From 
\[ p(t; \lambda), E[\lambda T \mid \lambda] = \sqrt{\alpha_2 \lambda} / \alpha_1 = \theta(w(1 - \Delta))^{-1} \] and 
\[ E[1 / (\lambda T) \mid \lambda] = \alpha_2, \]
\[ \alpha_1 = w(1 - \Delta)^2 + (1 + \Delta) \sigma^2 \lambda \] and 
\[ 1 / c_{est} = ((1 + \Delta) \sigma^2 \lambda + \theta w(1 - \Delta)) / \theta^2 , \]
or 
\[ E[\theta / (\lambda T) \mid \lambda] = (1 + \Delta) \sigma^2 / \theta + w(1 - \Delta) . \] Thus, as the synaptic energy-budget grows, \(\theta\) can increase, and the estimation of \(\lambda\) has the sensible limit, 
\[ \lim_{\theta \to \infty} E[\theta(w(1 - \Delta))^{-1} / (\lambda T)] = \lambda . \]
However for finite energy, the diffusion coefficient exerts itself on the value of \(c_{est}\) (see below). (The logarithmic constraint leads to a rather complex form, 
\[ E[\log(T / b)] \mid \lambda] = \log \sqrt{\alpha_2 / \alpha_1} - \log(\lambda b) - \pi^{1/4} \alpha_2 / 4 \cdot \exp(2\sqrt{\alpha_2 \lambda}) K_{1/2} K_{1/2} \), \]
where \( K_{1/2} \) is the derivative of the index \(\lambda\) of the Bessel function. In sum, replacing the \(\alpha\)'s with the physical attributes of this neuron, 
\[ p(t; \lambda) = \theta(1 + \Delta) \sigma^2 \lambda T - w(1 - \Delta)^2 \sigma^2 \lambda t / (1 + \Delta - \theta^2) \]
and the same conclusions as in Ex 1, and more, follow: (i) \(t\) is a minimal sufficient statistic suitable for communication of all the information and for Bayesian updating, so long as the ancillary information of \(\Delta\) and \(\theta\) are available to a recipient, (ii) the estimator's variance goes to zero at \(\theta\), and necessarily synaptic energy, increases but even faster here: 
\[ \text{Var}[1 / \lambda] = \lambda^2 \sigma^2 (\theta w (1 - \Delta) + \sigma^2 (1 + \Delta)^2) / (2 \theta^4) \]
where \(E[1 / \lambda] = \lambda \) and \(c_{est} = (w(1 - \Delta) / \theta + (1 + \Delta) \sigma^2 / (2 \theta^4))^{-1} \), i.e., (iii) the estimate is unbiased for all energy-budgets.

Ex 3, \{ \[ E[1 / \lambda] \mid \lambda], E[1 / T] \mid \lambda], E[T] \mid \lambda], E[\log(T) \mid \lambda] \}. This is the final and most complex example with the unanticipated bonus of describing the input distribution \(p(\lambda)\). The constraint set includes all the constraints of Ex 2 plus an anti-Intersymbol interference constraint \(E[1 / T]\). The formal constraint statement for this last is 
\[ E[b_x / T] - c_{est} = 0 \]
with an additional, implicit consistency requirement arising from the estimation constraint 
\[ E[b_x / T \mid \lambda] = c_{est} b_x \]. Specifically,
\[ c_{est} = E[b_x / T] = \int p(\lambda) E[b_x / T \mid \lambda]d\lambda = c_{est} \int p(\lambda) \lambda b x d\lambda = E[\Lambda]b_x / c_{est} , \text{ i.e., } c_{est} = b_x E[\Lambda] . \] 
(7)

Also arising from the new constraint is an additional, seemingly problematic term in the exponent of the likelihood. That is, the MEM-likehood density resolves as 
\[ p(t; \lambda) = \lambda^\alpha (\alpha_2 + \alpha_2 \log(p(\lambda)) / \alpha_2)^{-a/2} t^{|a|} \cdot \exp(-\alpha_1 t - \alpha_2 / (\lambda t)) - \alpha_2 p(\lambda) \lambda b_x / (\lambda t)) / 2 K_{1/2} K_{1/2} \]
(8)
The new term in the exponent arises from the differentiation of the new Lagrange-equation term, 
\[ \frac{\partial}{\partial p(T = t | \Lambda = \lambda)} \int b_x t^{-1} p(t; \lambda) p(d\lambda) dt = p(\Lambda = \lambda) b_x / t = p(\Lambda = \lambda) b_x / (\lambda t) . \]
This density (8) has all the appearance of implying the undesirable Bessel-functions of \(\lambda\) in the moments. That is, using (8) and 
\[ S(\lambda) = \sqrt{\alpha_2 + \alpha_2 p(\lambda) b_x} \], the estimation constraint must be 
\[ E[(\lambda T)^{-1}] \mid \lambda] = \sqrt{\alpha_2 K_{1/2} K_{1/2}} (2 S(\lambda) \sqrt{\alpha_2}) / S(\lambda) K_{1/2} K_{1/2} \]. However, the original constraint denies this apparent \(\lambda\)-dependency,
and under just one condition, the $\lambda$-dependent Bessel functions do not occur. Because $\alpha_4$ is not a function of $\lambda$, by necessity $p(\lambda) \propto 1 / \lambda$; indeed, $p(\lambda) = c_2 / \lambda = (\lambda \log(\lambda_{\text{max}} / \lambda_{\text{min}}))^{-1}$, a bonus here requiring only the assumption of a continuous, proper probability density for the marginal in $\lambda$ (forcing the range, $\Lambda \in [0 < \lambda_{\text{min}}, \lambda_{\text{max}} < \infty]$). Then defining $\alpha_4 := \alpha_2 + \alpha_3 / \log(\lambda_{\text{max}} / \lambda_{\text{min}})$, the general form of Ex 2 reappears, $p(t|\lambda) = \lambda u(\alpha_4 / \alpha_i)^{\alpha_2/2} \cdot \exp(-\alpha_4 t - \alpha_4 / (\lambda t)) / 2K_0(2\sqrt{\alpha_4})$.

With this form, one can apply the assumptions that again yield the G&M result. Instead of performing these somewhat tedious calculations, there is still more to be implied from the constraint set for this example: it is now possible to specify the exact marginal $\lambda$-distribution if this system of constraints is to generate a consistent MEM-inference. In particular, the range of $\lambda$ is determined by the constraint set. First from (7)

$$c_{\text{est}} = b_2 \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} p(\lambda) d\lambda = b_2 (\lambda_{\text{max}} - \lambda_{\text{min}}) / \log(\lambda_{\text{max}} / \lambda_{\text{min}}).$$

(Note that LHS-increases imply increases of the mean which can occur with increases in either, or both of the bounds.) The second equation needed for the constraints to determine the range uses the expectation

$$E[\log(\hat{\lambda} b_2)] = \log(\lambda_{\text{min}} \lambda_{\text{max}} b_2^2) / 2$$

over the marginal distribution. Then noting that for any $\lambda$, $E[\log(T / b_2)] / \lambda = \log(\alpha_4 / \alpha_i) - \log(\lambda_{\text{max}} - \lambda_{\text{min}}) / \log(\lambda_{\text{max}} / \lambda_{\text{min}}) / \log(\lambda_{\text{max}} / \lambda_{\text{min}})$ again take an expectation over the marginal, yielding $E[\log(\hat{\lambda}) / c_2] = E[\log(T / b_2)] = \log(\alpha_4 / \alpha_i) - \log(\lambda_{\text{max}} - \lambda_{\text{min}}) / \log(\lambda_{\text{max}} / \lambda_{\text{min}})$ (numerically, this is a monotonic relation, and

$$\log(\alpha_4 / \alpha_i) - \log(\lambda_{\text{max}} - \lambda_{\text{min}}) / \log(\lambda_{\text{max}} / \lambda_{\text{min}}) = \log(\alpha_4 / \alpha_i) - E[\log(\hat{\lambda}) / c_2].$$

Thus we have two constraint-parameterized equations in two unknowns with $0 < \lambda_{\text{min}} < \lambda_{\text{max}}$, enough to determine the exact range of $\lambda$.

In sum, the MEM combined with sensible physical and statistical constraints identifies neurons that perform MEM-defined optimal inferences, i.e., the likelihood $p(t|\lambda)$, and identifies a minimal sufficient statistic, $t$, suitable for neuronal pulsed-based communication. Moreover, in Ex 3, the method specifies criteria, $\{\lambda_{\text{max}}, \lambda_{\text{min}}, p(\lambda) = (\lambda \log(\lambda_{\text{max}} / \lambda_{\text{min}}))^{-1}\}$, for constructing the input latent variable; that is, there is a required range and distribution of $\lambda$ that matches the MEM-constructed neuron relative to the neuron's estimation parameter, $c_{\text{est}}$, and output-signaling constraint, $c_{\text{est}}$. Regarding the implementation of some of these results, the Appendix contains remarks of relevance both for neural engineers and for neuroscientists. The next section leads back to earlier interest [31] in bits-per-joule conclusions.

VI. Mutual information, Bayes theorem, and bits/J

At this point – with both a likelihood and a marginal inferred from the constraint set – it seems plausible that there is a way to return to our older problem of describing a neuron as an optimal, bits/J device. Note that joules-per-pulse is exactly the given energy-budget, e.g., $E = c_s + E_{\text{syn}} + E_{\text{clock}}$. Moreover, this energy-budget forces the values of the other constraints (or vice versa): (i) $c_s$ is devoted to axonal signaling including pulse-velocity and avoidance of intersymbol-interference of successive pulses; (ii) as illustrated in the examples, $E_{\text{syn}}$ determines the precision of the estimation computation; and (iii) $E_{\text{clock}}$ accounts for clocking costs, and by implication, temporal resolution of clocking. Given the availability of the joules-per-pulse cost, we only need bits-per-pulse to get bits/J. In this regard Ex 3 stands out as yielding all the necessary pieces. That is, one needs to conjure Shannon-bits (mutual information, [32]) that is consistent with and that can upgrade the MEM-implied differential entropy, $h(T|\lambda)$ an early outcome of the preceding derivations.

To obtain a mutual information while remaining in the context of computation for the purpose of estimating $\Lambda$'s value, take the viewpoint of any recipient of $y$'s output. Then equip this recipient with the appropriate prior distribution and apply Lindley's insight: a Bayesian experiment is quantifiable just as the noisy channel problem. That is, the Bayes inference is $p(\lambda | t) = \frac{p(t|\lambda)p(\lambda)}{\int p(t|\lambda)p(\lambda) d\lambda} = \frac{p(t|\lambda)p(\lambda)}{p(t)}$ generates the
mutual information that can be written in either of two ways, \( E_{\Lambda,T} \left[ \log \frac{p(\Lambda|T)}{p(\Lambda)} \right] = E_{\Lambda,T} \left[ \log \frac{p(T|\Lambda)}{p(T)} \right] . \) Thus from any recipient's perspective, an IPI (\( T \rightarrow t \)) is the outcome of the experiment that advances a recipient's prior in restricting the probable values of \( \Lambda \). The average amount that \( T \) helps is the bits of mutual information, sensibly called information-gain. In the case of Ex 3, all the necessary terms are available; that is, using \( p(t \mid \lambda) \) and \( p(\lambda) \), \( p(t) = \int p(t \mid \lambda) p(\lambda) d\lambda \) \(^1\), and \( h(T \mid h(T \mid \Lambda) \) is the information-gain. Of course a neuron never does any of these calculations, but a biologist seeking evidence for optimization in this realm or an engineer wanting to build energy-optimized neural computation is advised to perform these computations.

\(^1\) For example using (8) with \( a = -\frac{1}{2} \), the \( p(t) \) marginal resulting from this integration is

\[
p(t) = (t \log(\lambda_{\text{max}} / \lambda_{\text{min}}))^{-1} (\Psi(A_{\text{max}}(t)) - \Psi(A_{\text{min}}(t)) - \exp(4)(\Psi(B_{\text{max}}(t)) - \Psi(B_{\text{min}}(t))))
\]

where \( \Psi(x) = (\text{erf}(x) + 1) / 2 \), \( A_0(t) := \sqrt{t \lambda_0} \alpha_1 - \frac{1}{t} \alpha_0 / \sqrt{\lambda_0} ; B_0(t) := \sqrt{t \lambda_0} \alpha_1 + \frac{1}{t} \alpha_0 / \sqrt{\lambda_0} ; h \in \{\text{min}, \text{max}\} . \)

The MEM-application introduced here extends to other neural computations including discrimination and prediction (e. g., [34]), to more complicated neurons with individual dendritically localized inferences, and with some work, may also extend to network-level analyses (certainly an energy-efficient network computation is built from energy-efficient neurons). Returning again to the two distinct perspectives of an engineer vs. a biologist, it is clear that the results here tell the engineer the minimum energy-costs of computation. For the biologists there are experimental predictions as in [4,35]. That is, Nature has already done the experiment by producing differently parameterized neurons across species and brain regions, and the results here predict how one measurable variable changes as a function of another variable. In the earlier work the variable pairs were relatively simple, energy-use vs. firing-rate and quantal-failure-rate vs. firing-rate; here the variables are the variables of various energy-uses vs. the constraint parameters, e.g., \( c_{\text{ax}} \) or \( \lambda_{\text{ax}} \), which this last can only be measured indirectly through observation of the highest observable twin-spike firing frequency and the number of input lines to a neuron. As noted in the remarks below, a constraint set \( \{f_h(\lambda) = E[g_h(T) \mid \lambda] ; h \in \{1, ..., m\} \} \) implies the equalities \( \lambda = f_h^{-1}(E[g_h(T) \mid \lambda]) \forall h \) when the \( f_h \)’s are invertible, as they are here (additionally, unconditional constraining expectation statements imply conditional statements after taking the derivative or, as in the case of the axonal constraint \( E[1/T] = c_{\text{ax}} \), must hold exactly for every \( \lambda \) since the axon is agnostic regarding the latent variable and its estimation). Through the \( \lambda \)-linked equalities and the inferred likelihood, every parameter (e.g., \( c_{\text{ax}}, c_{\text{ent}} \)) can be expressed as a function of energy and expectations. Thus not only are firing distributions predicted for experimental confirmation/falsification but so are the matchings between constraint parameters and measurable energy-use.

VII. APPENDIX/REMARKS

A. Non-redundant, non-contradictory constraints

A hallmark of the present treatment is the simultaneous application of multiple constraints, each of which may be fundamentally different (energy, estimation, axon-communication, and normalization). In contrast to textbook MEM-applications [36], here parametric as well as logical consistency between constraints is an important issue. Care should be taken to avoid contradictions among constraints. A variant of Ex. 2 (no estimation constraint, no axon constraint, and a term energy-constraint, \( c_1 E[\Lambda \mid \Lambda] + c_2 E[\log(T) \mid \Lambda] = \mathcal{E}_n \) ) creates the possibility of a contradiction through the logarithmic energy-constraint. Because energy is positive, \( c_2 > 0 \) must be greater than zero, but then the term \( T^{c_2} \) is worrisome because if the constraint forces \( \alpha \geq 1 \), then normalization is impossible. A more obvious contradiction is the constraint set \( \{E[1/T] = c_1, E[1/T] = c_2 \ ; c_1 \neq c_2 \} \). However, the set \( \{E[1/T] = c_1, E[1/T \mid \lambda] = c_2 \lambda \} \) can be allowed but only if \( c_1 = \int c_2 \lambda \cdot p(\lambda) d\lambda = c_2 E[\Lambda] \) as in Ex. 3. Another obvious acceptable condition, requiring \( c_1 \leq c_2 \) is \( \{E[1/T] = c_1, E[1/T] \leq c_2 \} \), just a single, non-contradictory constraint. Here is another
Finally, there is an unavoidable parametric consistency issue that, at the same time, allows one to see the relationship of every constraint parameter to the energy-budget. The constraint set \( \{f_h(\lambda) = E[g_h(T)|\lambda]; h \in \{1,\ldots,m\}\} \) implies the equalities \( \{\lambda = f_h^{-1}(E[g_h(T)|\lambda]) \ \forall h\} \) when the \( f_h \)'s are invertible, as they are here (additionally, unconditional constraining expectation statements imply conditional statements after taking the derivative or, as in the case of the axonal constraint \( E[1/T]=c_{ax}\), must hold exactly for every \( \lambda \) since the axon is agonistic regarding the latent variable and its estimation).

Through \( \lambda \), this set of equalities interlinks the constants associated with all the individual constraining expectation-statements. As illustrated in the examples, UM calculations link the energy-budget to estimation constants and communication constraints. Such parametric relationships between the constants are design requirements for an engineer and are experimental objects to be measured for the neuroscientist. That is, the values of the interrelated constraint parameters are predictions of the theory subject to measurement, falsification, and adjustment by empirical neuroscience.

**B. GIG distributions: a biophysical speculation**

In the G&M model, the primary assumptions are a high frequency Poisson process, very small synaptic events, and the additivity of these events: because these events are small, numerous, and independent, their sum approximates a drifted B-M. The jump-over threshold issue for non-infinitesimal events can never occur, physically, because actual synaptic charge-injections are smoothly rising with finite derivatives. The critical assumptions for approximating B-M with such small events is \( (\lambda\text{-conditional}) \) independence and additivity of the synaptic events. Indeed and in contrast to passive neurons which are subadditive and lead to an O-U process \([37]\) (and therefore fail the requirement for a finite number of moment constraints for the FHT distribution \([38]\)) , neurons with active conductances can be linear in time and voltage. The Skellam and Poisson processes, as would satisfy \( \text{Ex. 2&3} \), have the critical property of independence and an active-channel neuron can supply additivity, but the question can be asked: How do GIG-distributions arise as opposed to just the IG itself? First, recall that it is one or more logarithmic constraints that give rise to the term \( r^{-1} \), where the exponent \( a \) determines the specific member of the GIG family. Moreover, regardless of this exponent's value \((<0)\), every GIG is a FHT outcome \([25]\), i.e., the solution to the Laplace-transformed Chapman-Kolmogorov continuous, backward-equation. However, different family-members occur because of a positional nonstationarity even while temporal stationarity holds. For example, the background, macroscopic process (e.g., a Poisson or Skellam) is temporally stationary, but the microscopic parameters need not be positionally stationary. As the barrier \( (\theta) \) is approached, the microscopic parameters (drift and diffusion) of the backward-equation change their values. When \( a=-\frac{1}{2} \), there is no barrier interaction; when \( a > -\frac{1}{2} \), then the barrier exerts a distance-dependent repulsion; and when \( a <-\frac{1}{2} \), the distance-dependent force is attractive. That is, the amplitude, but not frequency, of the synaptic events are modulated, much as could happen at the action-potential initiation-site where voltage-sensitive, Na- and K-channels activate as threshold is approached.

Finally, the result just described is not the only way to get a GIG distribution. Clever selection of a moving barrier function, \( \theta(t) \), and/or for a particular noise-distribution associated with resets \([39]\) or the barrier itself, even an O-U process might produce a GIG FHT.
References

[1] V. Balasubramanian, D. Kimber, and M. J. Berry II. "Metabolically efficient information processing." *Neural Comp.*, vol. 13, no. 4, pp. 799-815, 2001.
[2] S. B. Laughlin, "Energy as a constraint on the coding and processing of sensory information." *Current opinion in neurobiology*, vol. 11, no. 4, pp 475-480, 2001.
[3] D. B. Chklovskii, T. Schikorski, and C. F. Stevens. "Wiring optimization in cortical circuits." *Neuron*, vol., no. 3, 341-347, 2002.
[4] W. B. Levy, and R. A. Baxter. "Energy-efficient neuronal computation via quantal synaptic failures." *J. of Neurosci.*, vol. 22, no. 11, pp. 4746-4755, 2002.
[5] T. Berger, and W. B. Levy. "A mathematical theory of energy efficient neural computation and communication." *IEEE Trans Inf. Theory*, vol 56, no. 2, pp. 852-874, 2010.
[6] K. Friston, S. Samothrakis, and R. Montague. "Active inference and agency: optimal control without cost functions," *Biol. Cybernet.*, vol 106, no. 8-9, pp. 523-541, 2012.
[7] M. D. McDonnell, S. Ikeda, and J. H. Manton, "An introductory review of information theory in the context of computational neuroscience," *Biol. Cybernet.*, vol 105, no. 1, pp. 55-70, 2011.
[8] L. Kostal, P. Lansky, and M. D. McDonnell, "Metabolic cost of neuronal information in an empirical stimulus-response model," *Biol. Cybernet.*, vol. 107, no. 3, pp.355-365, 2013.
[9] J. Xing, et al. "Energy efficient neurons with generalized inverse Gaussian conditional and marginal hitting times." *IEEE Trans Inf. Theory.*, vol. 61, no. 8, pp. 4390-4398, 2015.
[10] E. T. Jaynes, "Information theory and statistical mechanics." *Physical Rev.*, vol. 106, no. 4, pp. 620, 1957.
[11] E. T. Jaynes, "Information theory and statistical mechanics. II." *Physical Rev.*, vol. 108, no. 2, pp. 171, 1957.
[12] J. E. Shore and R. W. Johnson. "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy." *IEEE Trans Inf. Theory*, vol 26, no. 1, pp. 26-37, 1980.
[13] J. Skilling, "The axioms of maximum entropy." *Maximum-Entropy and Bayesian Methods in Science and Engineering*. Springer Netherlands, 1988. 173-187.
[14] I. Csiszar, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems." *Ann. Stat.*, pp. 2032-2066, 1991.
[15] J. B. Paris and A. Vencovská, "A note on the inevitability of maximum entropy." *Int. J. Approximate Reasoning*, vol 4, no. 3, pp. 183-223, 1990.
[16] J. B. Paris and A. Vencovská, "In defense of the maximum entropy inference process." *Int. J. Approximate Reasoning*, vol. 17, no. 1, pp. 77-103, 1997.
[17] J. M. Van Campenhout and T. M. Cover, "Maximum entropy and conditional probability IEEE Trans Inf. Theory*, vol. 27, no. 4, pp. 483-489, 1981.
[18] E. T. Jaynes, "On the rationale of maximum-entropy methods." *Proc. IEEE*, vol. 70, no. 9, pp. 939-952, 1982.
[19] E. T. Jaynes, The Maximum Entropy Formalism, (MIT, Cambridge, MA, 1979) p15
[20] P. D. Grünwald and A. P. Dawid, "Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory." *Ann. Stat.*, pp. 1367-1433, 2004.
[21] H. Riaffa and R. Schlaifer, *Applied Statistical Decision Theory*, (Harvard Business School, Boston, 1962
[22] G. Darmois, "Sur les lois de probabilité estimation exhaustive," *CR Acad. Sci. Paris* 260.1265, pp. 85, 1935.
[23] B.O. Koopman, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, vol. 39, no. 3, pp. 399-409, 1936.
[24] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, no. 3-4, pp. 237-264, 1953.
[25] O. Barndorff-Nielsen, P. Blæsild, and C. Halgreen, "First hitting time models for the generalized inverse Gaussian distribution." *Stoch. Processes Appl.*, vol. 7, no. 1, pp. 49-54, 1978.
[26] W. Tuckwell, "Poisson approximation for dependent trials." *Ann. Prob*, pp. 534-545, 1975.
[27] H. C. Tuckwell, *Introduction to Theoretical Neurobiology*, (Cambridge University, New York, 1988), Vol. 2.
[28] G. L. Gerstein and B. Mandelbrot, "Random walk models for the spike activity of a single neuron," *Biophys. J.* vol. 4, no. 1 Pt 1, pp. 41, 1964.
[29] E. Halphen 1941 as cited in Perreault, Bobée, Rasmussen PF. Halphen distribution system. I: Mathematical and statistical properties. Journal of Hydrologic Engineering. 1999 Jul;4(3):189-99.
[30] E. Schrödinger, "Zur theorie der fall-und steigversuche an teilchen mit brownscher bewegung." Physikalische Zeit., vol 16 (1915); pp. 289-295.
[31] S. B. Laughlin and T. J. Sejnowski TJ, "Communication in neuronal networks," Science. 2003 Sep 26;301(5641):1870-4.
[32] Shannon, Claude E. "A mathematical theory of communication", Bell Sys.Tech. J., pp. 623-656, 1948.
[33] D. V. Lindley, "On a measure of the information provided by an experiment," Ann. Math. Stat. 1956 Dec 1:986-1005.
[34] W. B Levy, "A computational approach to hippocampal function," Computational models of learning in simple neural systems 23 (1989): , pp. 243-305.
[35] W. B Levy and R. A. Baxter, "Energy efficient neural codes," Neural Comp., vol. 8, no. 3, pp. 531-543, 1996.
[36] M. Tribus, Rational Descriptions, Decisions and Designs, (Pergamon, New York, 1969).
[37] W. H. Calvin and C. F. Stevens, "A Markov process model for neuron behavior in the interspike interval," InProc. 18th Annual Conf. Engineering in Medicine and Biology 1965 (Vol. 7, p. 118). As cited in Ricciardi LM, Sacerdote L. The Ornstein-Uhlenbeck process as a model for neuronal activity. Biological cybernetics. 1979 Mar 1;35(1):1-9.
[38] L. M. Ricciardi and S. Sato, "First-passage-time density and moments of the Ornstein-Uhlenbeck process," J. Applied Prob., vol., pp. 43-57, 1988.
[39] P. Lánský and C. E. Smith, "The effect of a random initial value in neural first-passage-time models," Math. Biosciences, vol. 93, no. 2, pp. 191-215, 1989.