

# ***Mathematical and Neuroscientific Arguments for Generalized Inverse Gaussian ISI Durations of Cortical Neurons***

**Keynote Address**

**Neural Coding: Information Beyond Shannon**

**Prague, Czech Republic, July 3 and 4, 2013**

**Toby Berger**

**University of Virginia**

**Charlottesville, VA 22904 USA**

# ACKNOWLEDGMENTS

- I AM DEEPLY INDEBTED TO PROF. WILLIAM B LEVY FIRSTWHO INTRODUCED ME TO NEUROSCIENCE, MENTORED ME NEURO-SCIENTIFICALLY FOR THE LAST 16 YEARS, AND INTRODUCED ME TO THE GIG DISTRIBUTION.
- I AM ALSO INDEBTED TO MY PhD STUDENT, JIE “JACKIE” XING WHO CONTRIBUTED TO MUCH OF THE RESEARCH I SHALL PRESENT.
- FINALLY, I AM INDEBTED TO TERRY SEJNOWSKI OF SALK INSTITUTE FOR ACCOMMODATING AND ENCOURAGING JACKIE AND ME OVER THE YEARS.

# OVERVIEW

We focus on information processing in sensory cortex.

The “classical” view of cortical processing (Barlow; Hubel and Wiesel) was that 1) information gets compressed as it travels up the hierarchy of each sensory modality, and 2) feedback (a.k.a. top-down) signaling is inessential to information processing, this despite the then-known anatomical fact that more than half the communication links in sensory cortex carry top-down signals.

As neuroscience experimental techniques have become more sophisticated over the decades, this classical view has become increasingly repudiated. Indeed, even the briefest of sensory stimuli now are known to result in extensive neural firing along bottom-up, top-down and horizontal cortical pathways. Barlow himself humbly acknowledged this in a paper he wrote in 2001, forty years after its more celebrated predecessor, Barlow (1961).

# WHAT HAS REPLACED THE CLASSICAL VIEW?

THE ANSWER DEPENDS, OF COURSE, ON TO WHOM YOU ASK THE QUESTION. MY ANSWER, SHARED I BELIEVE BY MANY OTHERS, IS THAT INFORMATION IS NEITHER DESTROYED NOR CREATED BETWEEN ONE LEVEL OF CORTEX OR THE NEXT. RATHER, IN THE LOWEST LEVELS (PRIMARY CORTEX), IT IS “REARRANGED” INTO A FORM THAT IS BEST SUITED NOT ONLY TO INFORMATION PROCESSING IN THE GIVEN LEVEL BUT ALSO TO THAT OF THE IMMEDIATELY HIGHER AND THE IMMEDIATELY LOWER LEVEL TO WHICH THE GIVEN LEVEL IS HEAVILY TIED. THE TOP-DOWN SIGNALS LARGELY (NOT EXCLUSIVELY) SUPPRESS CERTAIN CELLS IN THE LEVEL BELOW VIA INHIBITORY CONNECTIONS, AND BOTTOM-UP SIGNALS LARGELY (AGAIN, NOT EXCLUSIVELY) STIMULATE CERTAIN CELLS IN THE LEVEL ABOVE VIA EXCITATORY CONNECTIONS.

PART WAY UP, SOME OF THE INFORMATION STARTS GETTING SENT TO OTHER SENSORY MODALITIES (E.G., SOME AUDITORY INFO IS SENT TO THE VISUAL SYSTEM HIERARCHY, AND VICE VERSA.) HOWEVER, THE NET AMOUNT OF INFO AT EACH LEVEL OF EACH MODALITY REMAINS MUCH THE SAME. AT THE HIGHEST LEVELS, NEAR THE SENSORY-MOTOR INTERFACE, INFO ALSO IS EXCHANGED WITH MEMORY AT WHICH STAGE THE MATHEMATICAL PURPOSE OF THE INFORMATION PROCESSING BECOMES MORE CONCERNED WITH DECISION THEORY (FISHER et al.) THAN WITH INFORMATION THEORY (SHANNON et al.).

# HOW DOES AN INDIVIDUAL NEURON (IN PRIMARY CORTEX) CONTRIBUTE TO THE PROCESSING OF INFORMATION DESCRIBED IN THE PREVIOUS SLIDE?

THIS IS CLEARLY A QUESTION OF GREAT CONCERN TO THOSE AT THIS WORKSHOP. I WILL ATTEMPT TO ANSWER IT, AT LEAST IN PART. IN SO DOING, I WILL DESCRIBE AND ANALYZE **GENERALIZED INVERSE GAUSSIAN (GIG) NEURON MODELS** FROM THE VIEWPOINT OF SHANNON THEORY. THE PRESENTATION WILL BE A MIXTURE OF NEUROSCIENCE AND INFORMATION THEORY. PLEASE UNDERSTAND THAT, ALTHOUGH I CLAIM TO BE AN INFORMATION THEORIST, I DO NOT CLAIM TO BE A NEUROSCIENTIST.

# Our Performance Criterion: bits/Joule

Let  $a$  and  $b$  denote the times of occurrence of two successive spikes generated by a neuron. The half-open interval  $(a, b]$  is called the interspike interval (ISI). **We seek to maximize the number of Shannon information bits that each of the neuron's targets receives upon learning the value of  $b-a$  per Joule of energy that the neuron expends to process its inputs during the ISI, generate the spike at  $b$ , and propagate that spike to each of its targets.**

(Actually, we seek the bit rate at which the targets receive information over a long sequence of successive ISIs per Joule expended by the neuron. This rate is less than the single-ISI information per Joule. The reason for this is that, although the energy expended is additive over successive ISI's, the information received in successive ISI's is subadditive in general. In earlier work we have shown how to correct analytically for this to first order.)

# WHY EMPHASIZE ENERGY?

**Reason 1.** In 2010-11 a team of eminent life science, computer science, electrical engineering and mathematics experts generated a simulation of 10 million neurons in cat visual cortex that processes its input stimuli in real time. When emailing congratulations to that team's leader, Dharmendra Modha of IBM, Chip Levy and I also included our estimate that Modha's team's simulation used  $10^8$  times more energy per 'neuron' than does a real cat to perform the tasks in question. Dr. Modha's candid reply was, "**No,  $10^9$ !**" (***A human brain runs on less energy than does a 40 watt light bulb.***)

**Reason 2.** Energy constraints are, in a sense, the only constraints there are. For example, when sending a satellite into orbit there appear to be volume and weight constraints, but volume and weight constraints become inconsequential if you have unlimited energy. Sputnik proved it was thrust energy that mattered most.

# Energy Costs

Neuroscientifically speaking, there probably are dozens if not hundreds of different kinds of energy expenditures that transpire within a neuron during an ISI. However, it is possible (I think it even likely) that there are only a few mathematical forms that most of these energy costs assume. In what follows, we assume that the major energy-expenditure costs during an ISI are functions of the intensity,  $\lambda$ , of the neuron's afferent excitation therein and the duration,  $t$ , of the ISI. (The definition of  $\lambda$  will be given on a subsequent slide.)

Our candidates for the three main energy costs are listed at the top of the next slide.



# FUNCTIONAL FORMS OF THE THREE MAIN ENERGY TERMS

$$g_1(\lambda, t) = A \log(1/t)$$

$$g_2(\lambda, t) = B/t$$

$$g_3(\lambda, t) = Ct$$

The first thing that occurs to one looking at this list is that none of the three allegedly main energy terms depends on  $\lambda$ . That's right; none of them does! We shall embark on a lengthy but worthwhile "digression" aimed at explaining why this is the case. (Note also that all three of these allegedly main energy constraints depend on the neuron's **output** variable,  $t$ , whereas in man-made communication systems the constraints on energy, or power, traditionally are imposed on the **input**.)

# DEFINING THE EXCITATION INTENSITY

THE SET OF ALL NEURONS WHOSE AXONS CONNECT TO ONE OR MORE SYNAPSES OF A GIVEN NEURON IS CALLED THAT NEURON'S **AFFERENT COHORT**. A NEURON'S EXCITATION DERIVES FROM THE SPIKE TRAINS IT RECEIVES FROM THE NEURONS THAT COMPRISE ITS AFFERENT COHORT. ONE MEASURE OF THE **EXCITATION INTENSITY** IS THE TOTAL NUMBER OF SUCH SPIKES THAT ARRIVE PER SECOND. ANOTHER CANDIDATE MEASURE MULTIPLIES EACH OF THE AFFERENT SPIKES BY THE 'WEIGHT' OF THE SYNAPSE TO WHICH IT ARRIVES, SUMS THESE WEIGHTED ARRIVALS, AND THEN DIVIDES BY THE DURATION OF THE INTERVAL DURING WHICH THEY ALL ARRIVED. MOREOVER, MOST NEURONS HAVE BOTH EXCITATORY AND INHIBITORY SYNAPSES, SO A SUM WITH BOTH POSITIVE AND NEGATIVE ALGEBRAIC SIGNS MAY BE MORE APPROPRIATE. UNFORTUNATELY, HOW NEURONS COMBINE THEIR EXCITATORY AND INHIBITORY ARRIVALS REMAINS QUITE MYSTERIOUS DESPITE CONSIDERABLE EXPERIMENTAL INVESTIGATION. ADD TO THIS THE FACT THAT A TYPICAL CORTICAL NEURON HAS CIRCA 10,000 SYNAPSES TIGHTLY PACKED IN 3-D BOTH ON ITS DENDRITE TREE AND ON ITS SOMA. FURTHERMORE, IT TURNS OUT TO BE MORE "NATURAL" TO MEASURE EXCITATION INTENSITY IN SECONDS/SPIKE THAN IN SPIKES/SECOND.

**WHAT SHOULD A THEORIST DO RE ALL THESE COMPLICATIONS?**

# DEFINING THE EXCITATION INTENSITY (Cont.-1)

OUR APPROACH (NOT UNIQUE TO US) IS TO DEFINE EXCITATION INTENSITY NOT AT THE SPIKE LEVEL BUT AT THE **ION LEVEL**. THE ARRIVAL OF AFFERENT SPIKES GALVANIZES MULTIPLE PHENOMENA IN THE SYNAPSES, DENDRITE AND SOMA. THESE PHENOMENA RESULT IN BOTH POSITIVELY AND NEGATIVELY CHARGED IONS CONTINUALLY BEING DEPOSITED ON THE POSTSYNAPTIC MEMBRANE, SOME OF WHICH LATER DECAY TO GROUND. SINCE THE MEMBRANE IS CAPACITATIVE, A VOLTAGE CALLED THE **POSTSYNAPTIC POTENTIAL (PSP)** IS PRODUCED THEREON THAT FLUCTUATES IN ACCORDANCE WITH THE ION FLUX. WE ASSUME THAT THE PSP HAS A POSITIVE DRIFT RATE DURING EACH ISI. THUS, IT EVENTUALLY REACHES A THRESHOLD,  $\theta$ , WHICH FOR NOW WE CONSIDER TO BE A FIXED LEVEL. THIS ENDS THE ISI BY TRIGGERING AN EFFERENT SPIKE [a.k.a ACTION POTENTIAL (AP)] THAT TRAVELS ALONG THE NEURON'S AXON TO ALL ITS TARGET NEURONS.

## DEFINING THE EXCITATION INTENSITY (Cont.-2)

WE CONSIDER THE POSITIVE DRIFT RATE OF ION FLUX OF THE PREVIOUS SLIDE TO BE THE NEURON'S NET EXCITATION INTENSITY. IF THE NEURON'S TARGETS WERE TO KNOW THIS EXCITATION INTENSITY *a priori*, THE NEURON'S *raison d'etra* WOULD CEASE TO EXIST; SAID *raison d'etra* IS TO CONTINUALLY REPORT TO ITS TARGETS INFORMATION ABOUT ITS AFFERENT EXCITATION INTENSITY. IN ACCORDANCE WITH THE ABOVE, THE EXCITATION INTENSITY AT TIME  $t$  IS A RANDOM VARIABLE WHICH WE SHALL DENOTE BY  $\Lambda(t)$ .

THE NEURON REPORTS ONLY AT THE DISTINCT TIME INSTANTS AT WHICH IT EMITS A SPIKE. SINCE ALL ITS SPIKES ARE NEARLY IDENTICAL IN SHAPE, THE INFORMATION IN A SPIKE TRAIN RESIDES ENTIRELY IN THE RANDOM DURATIONS  $T_1, T_2, \dots$  THAT SEPARATE ITS SUCCESSIVE SPIKES, i.e., ITS ISI DURATIONS.

# DEFINING THE EXCITATION INTENSITY (Cont. -3)

SINCE THE NEURON SENDS ONLY ONE RANDOM VARIABLE PER ISI, IT CANNOT SPECIFY THE ENTIRE HISTORY OF  $\Lambda(t)$  DURING AN ISI. WE TAKE THE VIEWPOINT THAT, FOR THE  $k$ th ISI, CALL IT  $ISI_k$ , THE NEURON USES THE DURATION  $T_k$  OF  $ISI_k$  TO ENCODE THE TIME AVERAGE OF  $\Lambda(t)$  OVER ALL OF  $ISI_k$  EXCEPT THE ABSOLUTE REFRACTORY PERIOD THAT OCCUPIES THE INITIAL  $\Delta$  SECONDS OF EVERY ISI. WE DENOTE THIS TIME AVERAGE BY

$$\Lambda_k := \frac{1}{T_k - \Delta} \int_{S_{k-1} + \Delta}^{S_{k-1} + T_k = S_k} \Lambda(t) dt,$$

WHERE  $S_k = T_1 + \dots + T_k$ .

WE PROCEED TO STUDY THE AVERAGE MUTUAL INFORMATION  $I(\Lambda_k; T_k)$  BETWEEN  $\Lambda_k$  and  $T_k$  WITH PARTICULAR EMPHASIS ON SITUATIONS IN WHICH  $\Lambda_k$  and  $T_k$  ARE JOINTLY ABSOLUTELY CONTINUOUS WITH jpdf

$$f_{\Lambda_k, T_k}(\lambda, t) = f_{\Lambda_k}(\lambda) f_{T_k | \Lambda_k}(t | \lambda).$$

# THE CONDITIONAL GIG DISTRIBUTION

The GIG conditional pdf  $f_{T|\Lambda}(t|\lambda)$  has three parameters,

$\alpha \in (-\infty, \infty)$ ,  $\beta \in [0, \infty)$  and  $\gamma \in (0, \infty)$ ; define  $\underline{\alpha} = (\alpha, \beta, \gamma)$ .

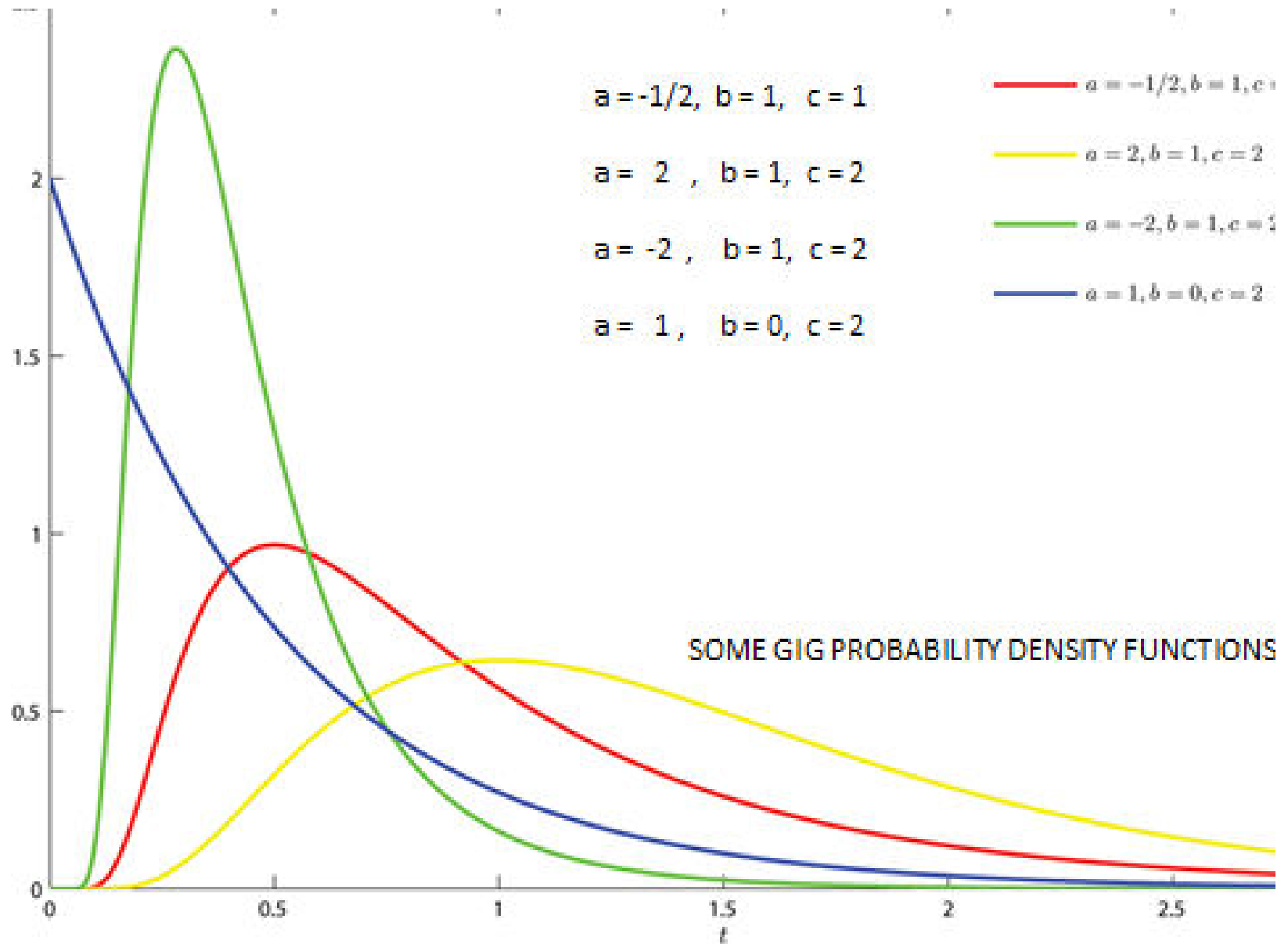
The mathematical formula for  $f_{T|\Lambda}(t|\lambda)$  is

$$f_{T|\Lambda}(t|\lambda) = C_{\underline{\alpha}}^{-1} \lambda^{\alpha} t^{\alpha-1} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right); t > 0, \lambda > 0.$$

The normalizing constant, which does not depend on  $\lambda$ , is

$$C_{\underline{\alpha}} := \int_0^{\infty} dt \lambda^{\alpha} t^{\alpha-1} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right) = 2\left(\frac{\beta}{\gamma}\right)^{\frac{\alpha}{2}} K_{\alpha}(2\sqrt{\beta\gamma}),$$

where  $K_{\alpha}(\cdot)$ , the modified Bessel function of the second kind and order  $\alpha$ , is strictly positive for positive real arguments.



## SOME FACTS ABOUT CONDITIONAL AND MARGINAL GIG pdf's

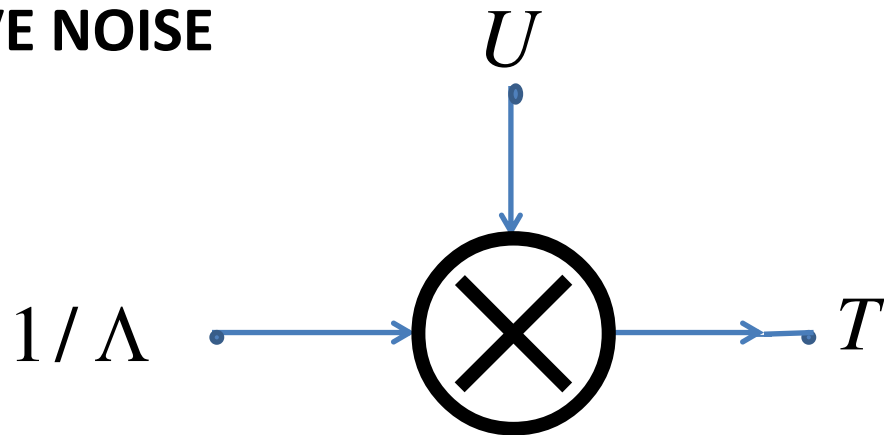
- The conditional GIG is in the exponential class, using Koopman's general definition thereof.
- Having three parameters is compatible with von Neumann's rule for not overfitting, namely:  
“With four parameters, I can fit an elephant, and with five I can make him wiggle his trunk.”
- The marginal GIG- $\underline{\alpha}$  pdf is the conditional GIG- $\underline{\alpha}$  with  $\lambda = 1$ . For  $\beta = 0$ , the GIG reduces to a Gamma conditional pdf.
- I strongly recommend the conditional GIG pdf for modeling neurons in primary cortex for reasons to be elucidated herein.



# The GIG Diffusion

In the late 1970's, O. Barndorff-Nielsen et al. found a positively drifting diffusion with infinitesimal increments that first hits a fixed threshold at a time that is GIG-distributed. That generalized a 1915 result of Schrodinger showing that a particle released at height  $h$  undergoing viscous Brownian motion in a gravitational field will first hit the ground at a time that today is said to be Inverse Gaussian (IG) distributed, i.e., GIG with  $\alpha = -\frac{1}{2}$ . For  $\alpha \neq -\frac{1}{2}$ , the increments in Barndorff-Nielsen's diffusion over disjoint time intervals are neither mutually independent nor Gaussian the way they are when  $\alpha = -\frac{1}{2}$ , but they remain infinitely divisible. In future work we hope to relate the parameters in Barndorff-Nielsen's diffusion to properties of real neurons.

# GIG CONDITIONAL AS A “CHANNEL” WITH INDEPENDENT MULTIPLICATIVE NOISE



$$U \perp\!\!\!\perp \text{GIG}(\underline{\alpha}); U \perp\!\!\!\perp \Lambda; \text{ hence } U \perp\!\!\!\perp (1/\Lambda).$$

Theorem:

The conditional pdf for  $T$  given  $1/\Lambda$  for the channel shown above is the GIG- $\underline{\alpha}$  conditional; i.e.,  $f_{T|1/\Lambda}(t|\lambda) = C_{\underline{\alpha}}^{-1} \lambda^{\alpha} t^{\alpha-1} \exp(-\frac{\beta}{\lambda t} - \gamma \lambda t)$ .

Proof: For the channel shown,  $P[T \in (t, t + dt) | \Lambda = \lambda] = P[\Lambda T \in (\lambda t, \lambda t + \lambda dt) | 1/\Lambda = 1/\lambda]$

Since  $U = \Lambda T$ , the right-hand side of the above probability reads

$$P[U \in (\lambda t, \lambda t + \lambda dt) | 1/\Lambda = 1/\lambda] = P[U \in (\lambda t, \lambda t + \lambda dt)], \text{ equality holding because } U \text{ is independent of } 1/\Lambda. \text{ So, } f_{T|1/\Lambda}(t|1/\lambda) = f_{T|\Lambda}(t/\lambda) = \lambda f_U(\lambda t) = \lambda C_{\underline{\alpha}}^{-1} (\lambda t)^{\alpha-1} \exp(-\frac{\beta}{\lambda t} - \gamma \lambda t) = C_{\underline{\alpha}}^{-1} \lambda^{\alpha} t^{\alpha-1} \exp(-\frac{\beta}{\lambda t} - \gamma \lambda t). \text{ QED } 19$$

## AN ADDITIVE INDEPENDENT NOISE CHANNEL

IT FOLLOWS FROM THE PREVIOUS SLIDE THAT THE CHANNEL WITH INPUT  $\log(1/\Lambda)$ , “NOISE”  $\log U$ , AND OUTPUT  $\log T$  IS AN **ADDITIVE** INDEPENDENT NOISE CHANNEL, WHICH FAMILY OF CHANNELS IS PARTICULARLY AMENABLE TO MATHEMATICAL ANALYSIS. WE SHALL RETURN TO THIS FACT AT AN OPPORTUNE TIME.

# PERFORMANCE CRITERION REDUX

WE HAVE SAID THAT OUR OBJECTIVE IS TO MAXIMIZE BITS PER JOULE (bpJ). BUT

## **WHAT IS IT WE'RE GOING TO VARY IN ORDER TO ACHIEVE THIS MAXIMUM?**

FOR MORE THAN A DECADE , I HAVE DONE A POOR JOB OF EXPLAINING THIS TO OTHERS. FROM THE OUTSET I CONSIDERED THAT THE CHANNEL MODEL, i.e., THE pdf OF  $T$  CONDITIONAL ON  $\Lambda$  , WAS FIXED/GIVEN. SINCE THE JOINT DISTRIBUTION OF  $\Lambda$  AND  $T$  DETERMINES BOTH THE AVERAGE MUTUAL INFORMATION AND THE AVERAGE ENERGY EXPENDITURE, THIS LEFT THE MARGINAL DISTRIBUTION OF  $\Lambda$  AS THE ONLY ENTITY OVER WHICH TO PERFORM THE MAXIMIZATION, SO THAT'S WHAT I DID. THE TROUBLE WITH THIS IS THAT THE NEURON HAS ALMOST NO ABILITY, ESPECIALLY IN THE SHORT TERM, TO VARY THE  $\Lambda$  MARGINAL. THE AFFERENT COHORT IS THE OBJECT THAT GENERATES THE  $\Lambda$  - MARGINAL.

OF LATE, IT HAS BECOME CLEAR THAT WHAT REALLY IS GOING ON NEUROSCIENTIFICALLY IS A **JOINT SOURCE-CHANNEL OPTIMIZATION** THAT I ADDRESS IN THE NEXT SLIDE.

# ENERGY-EFFICIENT JOINT SOURCE-CHANNEL OPTIMIZATION FOR NEURONS

IF YOU SIMPLY SAY THAT YOU WILL OPTIMIZE OVER THE JOINT DISTRIBUTION OF THE SOURCE AND THE CHANNEL, THEN YOU HAVE NO MATHEMATICALLY USEFUL STRUCTURE; EVERY JOINT DISTRIBUTION OVER ANY PRODUCT SPACE IS AT YOUR DISPOSAL.

I BELIEVE THE CORRECT STRUCTURE FOR THE PROBLEM AT HAND IS AS FOLLOWS: FIRST, FIX THE FAMILY OF CHANNELS (CONDITIONAL DISTRIBUTIONS) YOU ARE GOING TO USE. (FOR NEURONS, MY FAVORITE SUCH CHOICE IS THE FAMILY OF GIG CONDITIONAL MODELS.) HOWEVER, DO NOT FIX **WHICH** CONDITIONAL pdf IN THE FAMILY YOU ARE CONSIDERING. INSTEAD, ASSUME THAT THE NEURON IS **ADAPTIVE** IN THE SENSE THAT IT CAN VARY ITS PARAMETERS, EVEN IF ONLY OVER A LIMITED DYNAMIC RANGE. FOR GIG NEURONS, THIS MEANS THAT  $\underline{\alpha} = (\alpha, \beta, \gamma)$  CAN BE VARIED BY THE NEURON OVER A 3-D SUBSET. ASSOCIATED WITH THIS SUBSET, THERE IS A FAMILY OF  $\Lambda$ -MARGINALS EACH OF WHICH IS bpJ-OPTIMIZING FOR ONE SUCH  $\underline{\alpha}$ . IF THE AFFERENT COHORT IS, OF LATE, GENERATING A MARGINAL FOR  $\Lambda$  WHICH IS NEAR IN DISTRIBUTION TO ONE OF SAID bpJ-OPTIMIZING MARGINALS, THE NEURON WILL IN REASONABLY SHORT ORDER ADAPT ITS  $\underline{\alpha}$  VALUE THERETO. **THIS IS WHY FINDING THE bpJ-MAXIMIZING  $\Lambda$ -MARGINAL FOR EACH  $\underline{\alpha}$  IS SO INTERESTING!**

## **A SELF-SERVING BRIEF HISTORY OF JOINT SOURCE-CHANNEL MATCHING**

IN A FEBRUARY 1998 INVITED LECTURE DELIVERED AT SALK INSTITUTE, I STATED PUBLICLY FOR THE FIRST TIME MY BELIEF THAT MANY BIOLOGICAL SYSTEMS ARE DOUBLY MATCHED IN THAT THE SOURCE OF THE SYSTEM'S INPUT DRIVES THE SYSTEM MODEL, OR CHANNEL, AT A RATE THAT EQUALS ITS (ENERGY-CONSTRAINED) SHANNON CAPACITY, WHILE SIMULTANEOUSLY THE CHANNEL'S CONDITIONAL DISTRIBUTION OF OUTPUT GIVEN INPUT ACHIEVES AN OPERATING POINT ON SHANNON'S RATE-DISTORTION FUNCTION WITH RESPECT TO A FIDELITY CRITERION APROPOS OF THE JOINT (SOURCE,USER) PAIR. THIS MEANS THAT SHANNON-OPTIMAL END-TO-END PERFORMANCE IS OBTAINED SIMPLY BY PUTTING THE SOURCE DIRECTLY INTO THE CHANNEL AND THE FEEDING THE CHANNEL OUTPUT DIRECTLY TO THE USER. (I LATER DISCOVERED THAT DECADES EARLIER PETER ELIAS HAD PUBLISHED AN EXAMPLE OF SUCH A CODING-FREE OPTIMUM SYSTEM IN AN OBSCURE NON-BIOLOGICAL CONTEXT). AT ISIT 2000 IN SORRENTO I DISCUSSED SUCH DOUBLE MATCHING AT LENGTH WITH BIXIO RIMOLDI. ON JULY 4 2001 I WAS INVITED BY EPFL LAUSANNE, WHERE RIMOLDI HAD RECENTLY JOINED THE FACULTY, TO GIVE A PLENARY LECTURE IN WHICH I ALSO TOUCHED ON SUCH DOUBLE MATCHING. A YEAR LATER TO THE DAY, I HAD THE HONOR OF PRESENTING THE SHANNON LECTURE ENTITLED "LIVING INFORMATION THEORY" AT ISIT 2002, LAUSANNE, A PORTION OF WHICH WAS DEVOTED TO BIOLOGICAL DOUBLE MATCHING . I NEVER PUBLISHED ANYTHING ABOUT DOUBLE MATCHING. RIMOLDI'S STAR STUDENT, MICHAEL GASTPAR, SUBSEQUENTLY WROTE A THESIS ENTITLED "To Code or Not To Code" THAT EXPANDED DOUBLE MATCHING THEORY. "To Code or Not To Code" ALSO WAS PUBLISHED AS A PAPER IN IEEE TRANSACTIONS ON INFORMATION THEORY WITH CO-AUTHORS RIMOLDI AND VETTERLI.

# WHY DON'T THE ENERGY CONSTRAINTS DEPEND ON $\Lambda$ ?

WE ARE NOW IN A POSITION TO ADDRESS THIS INTERESTING QUESTION. MANY NEUROSCIENTISTS WOULD CORRECTLY STATE THAT THE AMOUNT OF SIGNAL PROCESSING (ION MOVING) THAT THE NEURON PERFORMS DURING AN ISI IS PROPORTIONAL TO THE PRODUCT OF  $\Lambda$  and  $T$ , WHICH CERTAINLY SUGGESTS THE NEED FOR AN ENERGY TERM PROPORTIONAL TO  $\Lambda T$ . WE HAVE JUST EXPLICATED WHY OUR PROBLEM IS TO MAXIMIZE bpj BY VARYING THE  $\Lambda$ -MARGINAL. WE DON'T CONSTRAIN THE ENERGY IN EACH ISI, THOUGH. WE CONSTRAIN THE **AVERAGE** ENERGY EXPENDED IN AN ISI. THAT IS, bpj IS A RATIO OF TWO AVERAGES, AN AVERAGE MUTUAL INFORMATION AND AN AVERAGE ENERGY. ON THE NEXT SLIDE WE WILL SHOW THAT FOR A GIG NEURON, AMONG OTHERS, ANY FUNCTION OF  $\Lambda$  and  $T$  THAT IS EXPRESSIBLE AS A FUNCTION SOLELY OF  $\Lambda T$  HAS AN EXPECTED VALUE THAT DOES NOT DEPEND ON HOW THE  $\Lambda$ -MARGINAL IS CHOSEN. THEREFORE, IT IS NOT NECESSARY TO INCLUDE ANY ENERGY TERM THAT IS A FUNCTION SOLELY OF THE  $\Lambda T$  PRODUCT IN THE SOLUTION PROCEDURE. SUCH TERMS HAVE AN EFFECT ON THE TOTAL ENERGY EXPENDED BUT NOT ON THE bpj MAXIMIZING  $\Lambda$ -MARGINAL.

# PROOF THAT, FOR A GIG NEURON MODEL, $E[g(\Lambda T)]$ DOESN'T DEPEND ON THE $\Lambda$ -MARGINAL

$$\begin{aligned} E[g(\Lambda T)] &= \int_0^\infty dF_\Lambda(\lambda) \int_0^\infty dt f_{T|\Lambda}(t|\lambda) g(\lambda t) \\ &= \int_0^\infty dF_\Lambda(\lambda) \int_0^\infty dt C_{\underline{\alpha}}^{-1} \lambda^\alpha t^{\alpha-1} \exp\left(-\frac{\beta}{\lambda t} - \gamma \lambda t\right) g(\lambda t) \end{aligned}$$

IN THE INNER INTEGRAL, LET  $x = \lambda t$ . CONTINUING THE EQUALITY CHAIN,

$$= \int_0^\infty dF_\Lambda(\lambda) \int_0^\infty dx C_{\underline{\alpha}}^{-1} x^{\alpha-1} \exp\left(-\frac{\beta}{x} - \gamma x\right) g(x)$$

THE INNER INTEGRAL IS SEEN TO BE A CONSTANT, CALL IT  $C$ . IT FOLLOWS THAT THE ANSWER IS  $C$  REGARDLESS OF THE FORM OF THE CHOICE OF THE  $\Lambda$ -MARGINAL  $F_\Lambda(\cdot)$ . QED



## OTHER FUNCTIONS THAT CAN'T BE ENERGY FUNCTIONS

NO ENERGY FUNCTION CAN DEPEND SOLELY ON  $\Lambda$  AND NOT AT ALL ON  $T$ . THAT'S BECAUSE  $\Lambda$  IS GENERATED ENTIRELY BY THE NEURON'S AFFERENT COHORT, NOT THE NEURON ITSELF. THEREFORE, THE MEMBERS OF THE AFFERENT COHORT GET CHARGED FOR THE ENERGY THEY EXPEND IN GENERATING  $\Lambda$ . TO CHARGE IT AGAIN TO OUR NEURON CONSTITUTES FAULTY ACCOUNTING. THE FIXED COSTS PER ISI, SUCH AS THAT FOR PROPAGATING THE AP GENERATED AT THE END OF THE ISI TO ALL THE TARGETS, ALSO ARE NOT OF INTEREST TO bpj MAXIMIZATION, SINCE A CONSTANT,  $C$ , ALSO IS A FUNCTION OF  $\Lambda T$ , NAMELY  $C(\Lambda T)^0$ .

THIS LEAVES AS THE ONLY REMAINING POSSIBILITY FUNCTIONS THAT INVOLVE BOTH  $\Lambda$  AND  $T$  IN A WAY THAT DOES NOT REDUCE TO A FUNCTION SOLELY OF THEIR PRODUCT. SUCH FUNCTIONS SEEM RATHER FAR-FETCHED, SO I WON'T THINK ABOUT THEM UNLESS AND UNTIL SOME ONE FINDS ONE THAT IS SHOWN TO HAVE NEUROSCIENTIFIC SIGNIFICANCE IN THE ENERGY SENSE.

**THUS, I'M STICKING TO MY CLAIM THAT THE ONLY ENERGY TERMS THAT MATTER ARE FUNCTIONS OF  $T$  ONLY.**

# A MATHEMATICAL PROCEDURE FOR DETERMINING THE bpJ-MAXIMIZING $\Lambda$

WE NOTED EARLIER THAT,  $Z = X + Y$ , WHERE  $X = \log(1/\Lambda)$ ,  $Y = \log U$ , AND  $Z = \log T$ , CONSTITUTES AN ADDITIVE INDEPENDENT NOISE CHANNEL.

IT FOLLOWS THAT THE CHARACTERISTIC FUNCTION OF  $Z$  IS THE PRODUCT OF THE CHARACTERISTIC FUNCTION OF  $X$  AND THAT OF  $Y$ . IN SYMBOLS,  $\varphi_Z(v) = \varphi_X(v) \cdot \varphi_Y(v)$

THEREFORE,  $\varphi_X(v) = \varphi_Z(v) / \varphi_Y(v)$ .

SINCE  $Y = \log U$  AND WE KNOW THAT  $U \sim \text{GIG}(\underline{\alpha})$ ,

AT LEAST IN THEORY WE KNOW HOW TO FIND  $\varphi_Y(v)$ . WE DON'T KNOW IN GENERAL WHAT  $\varphi_Z(v)$  IS. BUT, IT TURNS OUT THAT IN THE CASE OF INTEREST TO US, NAMELY THAT IN WHICH  $\Lambda$  IS bpJ-MAXIMIZING, WE CAN FIND  $\varphi_Z(v)$ ! (SEE NEXT SLIDE) THAT MEANS WE CAN FIND  $\varphi_X(v) = \varphi_{\log(1/\Lambda)}(v)$ , THEN  $f_{\log(1/\Lambda)}(\cdot)$ , THEN  $f_{1/\Lambda}(\cdot)$ .

# FINDING $f_T(\cdot)$ WHEN bpJ IS MAXIMIZED

GALLAGER'S TEXT CONTAINS THE EQUATIONS ONE MUST SOLVE TO FIND THE INPUT DISTRIBUTION THAT ACHIEVES A CHANNEL'S CAPACITY. THESE HAVE BEEN EXTENDED BY JIMBO-KUNISAWA AND OTHERS TO TIME AND AMPLITUDE CONTINUOUS SIGNALS AND TO CONSTRAINTS ON THE INPUT AND/OR OUTPUT.

FOR OUR bpJ MAXIMIZATION PROBLEM, ONE OBTAINS THE FOLLOWING PAIR OF INTEGRAL EQUATIONS:

$$\int_0^{\infty} dt f_{T|\Lambda}(t|\lambda) \left[ \log\left(\frac{f_{T|\Lambda}(t|\lambda)}{f_T(t)}\right) - A \log(1/t) - B/t - Ct \right] = 0 \text{ for all } \lambda \geq 0 \quad (1)$$

$$f_T(t) = \int_0^{\infty} d\lambda f_{\Lambda}(\lambda) f_{T|\Lambda}(t|\lambda) \quad (2)$$

Equation (1) above is the one that results from the energy-constrained version of the Jimbo-Kunisawa extension of Gallager's result. Note that (1) does not even explicitly involve the unknown  $f_{\Lambda}(\lambda)$ . That's why equation (2) is needed as well. The procedure is first to solve (1) for the  $f_T(t)$  that the optimizing  $f_{\Lambda}(\lambda)$  generates, then to solve (2) for  $f_{\Lambda}(\lambda)$  itself.

# SOLVING EQUATION (1):

IN THE SPECIAL CASE OF A GAMMA CHANNEL MODEL OBTAINED FROM THE GIG MODEL BY SETTING  $\beta = 0$ , IT IS STRAIGHTFORWARD TO SOLVE Eq. (1) USING LAPLACE TRANSFORMS, AMONG OTHER METHODS.

The result is that  $f_T(t)$  also is Gamma-distributed, but with different parameters than  $\alpha$  and  $\gamma$  of the conditional pdf with  $\lambda = 1$ .

LAPLACE TRANSFORMS ARE UNABLE TO SOLVE Eq. (1) IN THE GENERAL GIG CASE, BUT THE ABOVE RESULT IN THE GAMMA SUBCASE SUGGESTED THE CONJECTURE THAT, FOR SOME  $\underline{a} \neq \underline{\alpha}$ ,  $f_T(t)$  IS GIG ( $\underline{a} = (a,b,c)$ ) FOR A GIG ( $\underline{\alpha} = (\alpha,\beta,\gamma)$ ) CONDITIONAL.

SUBSTITUTING THIS CONJECTURE INTO Eq. (1) PRODUCES TERMS ON THE LHS EACH OF WHICH IS PROPORTIONAL EITHER TO  $\lambda, 1/\lambda$ , or  $\log \lambda$ . UPON GROUPING TERMS OF OF THESE THREE TYPES, IT IS POSSIBLE TO ZERO THE NET COEFFICIENTS OF EACH GROUP BY SETTING  $a = A, b = B$ , and  $c = C$ , THEREBY "VALIDATING" THE CONJECTURE.

I PUT "VALIDATING" IN QUOTES BECAUSE THE ZEROING CHOICES  $(a,b,c) = (A,B,C)$  SEEM TO HOLD *INDEPENDENTLY* OF THE VALUE OF  $\underline{\alpha}$ . THIS IS NOT REALLY THE CASE, THOUGH. WE KNOW FROM THE GAMMA SUBCASE THAT THERE IS ALSO THE REQUIREMENT THAT  $0 < a < \alpha$ . THAT CONDITION, AND PERHAPS ANOTHER WE'RE UNSURE OF AT PRESENT, ARE REQUIRED FOR THE GENERAL GIG.

# HYPERGEOMETRIC FUNCTIONS AND BARNES INTEGRALS

WHEN WE FOLLOW THROUGH THE PROCEDURE FOR FINDING  $f_T(t)$  FOR THE GAMMA SUBCASE, WE OBTAIN FOR THE CHARACTERISTIC FUNCTION OF  $1/\Lambda$  AN INTEGRAL REPRESENTATION OF A CONFLUENT HYPERGEOMETRIC FUNCTION  ${}_1F_1(g, h; z)$  WITH  $z$  PURELY IMAGINARY, WHICH REPRESENTATION IS CALLED A **BARNES INTEGRAL** (1908). THERE IS ANOTHER INTEGRAL REPRESENTATION FOR  ${}_1F_1(g, h; jv)$  THAT'S INTERPRETABLE AS THE CHARACTERISTIC FUNCTION OF A BETA DISTRIBUTION. IT FOLLOWS THAT THE bpJ-MAXIMIZING  $f_{1/\Lambda}(\lambda)$  IS A SCALED **BETA DISTRIBUTION**, THE SCALE FACTOR BEING  $\frac{\gamma}{c}$ .

THERE IS A SIMPLER WAY TO SHOW IN THE GAMMA SUBCASE THAT  $f_{1/\Lambda}(\lambda)$  IS A BETA DISTRIBUTION, NAMELY VIA LAPLACE TRANSFORMS (Berger-Levy, 2010). NONETHELESS, THE ABOVE IS QUITE INTRIGUING BECAUSE 1) LAPLACE TRANSFORMS DO NOT WORK FOR GENERAL GIG, AND 2) THERE IS ALSO A BARNES INTEGRAL REPRESENTATION OF THE HYPERGEOMETRIC FUNCTION  ${}_2F_1(g, h, l; z)$ . WE ARE INVESTIGATING IF THIS CAN LEAD TO AN "EXPLICIT" ANSWER FOR  $\varphi_{1/\Lambda}(v)$  and thus  $f_{1/\Lambda}(\cdot)$  and  $f_\Lambda(\cdot)$  FOR THE GENERAL GIG CASE.

WE REMARK THAT BARNES INTEGRALS ARE BEING EMPLOYED IN BRANCHES OF MODERN PHYSICS, INCLUDING GAUGE THEORY AND STRING THEORY.

# PROFFERED NEUROSCIENTIFIC EXPLANATION OF B/T ENERGY TERM

DURING THE ABSOLUTE REFRACTORY PERIOD (ARP) IMMEDIATELY FOLLOWING THE TRIGGERING OF A SPIKE, THE NEURON CANNOT FIRE ANOTHER SPIKE REGARDLESS OF HOW STRONG ITS AFFERENT EXCITATION MAY BE. WE HAVE DENOTED THE DURATION OF THE ARP BY  $\Delta$ .

NEUROSCIENTISTS CAN EXPERIMENTALLY INDUCE/EVOKE AN EFFERENT ACTION POTENTIAL (AP) SHORTLY AFTER  $t = \Delta$ , IN THE SO-CALLED RELATIVE REFRACTORY PERIOD (RRP), BY DIRECTLY INSERTING ELECTRONIC STIMULATION INTO THE AXON'S INITIAL SEGMENT WHERE AP'S ARE GENERATED. THE STRONGER THESE SIGNALS ARE, THE SOONER AFTER  $t = \Delta$  IT IS THAT THE NEURON FIRES. HOWEVER, IT TAKES LARGER AND LARGER INPUT ENERGY TO GET CLOSER AND CLOSER TO  $t = \Delta$ .

IN RESONSE TO AFFERENT AP'S, NEURONS GENERATE EXCITATORY POSTSYNAPTIC POTENTIALS (EPSP'S) WHOSE DURATIONS ARE SOME 2 ms LONG. AT ITS SYNAPSES A NEURON HAS NO WAY OF KNOWING WHETHER THE NEXT SPIKE WILL BE GENERATED DURING THE RRP OR WELL AFTER THE END OF THE RRP, SO EPSP'S HAVE THE SAME DURATION IN EITHER CASE.

# PROFFERED EXPLANATION OF THE B/T ENERGY TERM (Cont.)

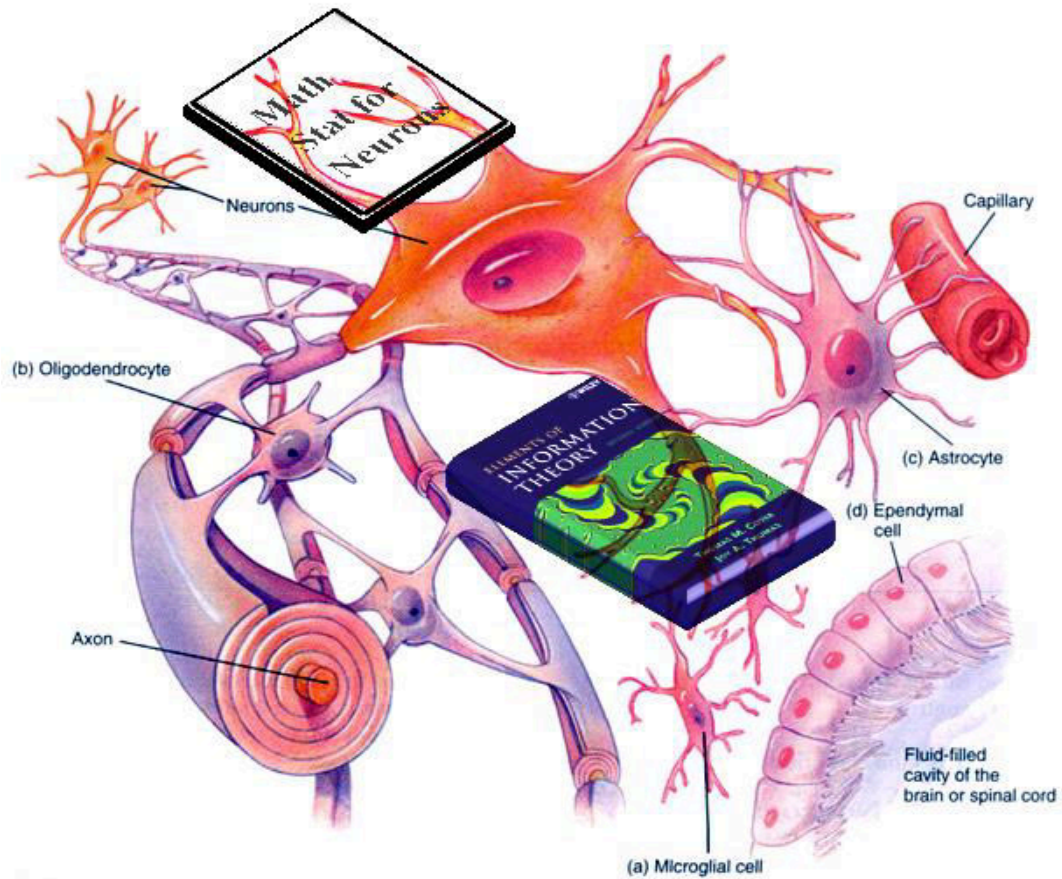
THE ONLY EPSP'S THAT CAN CONTRIBUTE TO AN AP GENERATED AT  $t = \Delta + \delta$  MUST ARRIVE AT THE AXON HILLOCK AFTER  $\Delta - \tau$  AND BEFORE  $t = \Delta + \delta$ , WHERE  $\tau$  DENOTES THE DURATION OF AN EPSP. SINCE THE PSP CANNOT LEAVE ITS RESET VALUE UNTIL THE ARP IS OVER, AT MOST A FRACTION  $\delta / \tau$  OF ANY PSP CAN CONTRIBUTE TO A SPIKE GENERATED AT  $t = \Delta + \delta$ . THE UPSHOT OF THIS IS THAT, IF  $T - \Delta < \tau$ , IT TAKES  $\tau / (T - \Delta)$  TIMES AS MUCH ENERGY, MUCH OF WHICH IS WASTED,\* TO PRODUCE A SPIKE AT  $T - \Delta$  AS IT DOES TO PRODUCE ONE AT ANY TIME GREATER THAN  $\Delta + \tau$ .

SAYING THE SAME THING SLIGHTLY DIFFERENTLY, IF THE ACTIVE PROCESSING TIME  $T^* := T - \Delta$  IS LESS THAN  $\tau$ , IT TAKES  $\tau / T^*$  TIMES AS MUCH ENERGY TO SPIKE AT  $T^*$  AS IT DOES TO SPIKE AT ANY TIME GREATER THAN  $\tau$ .

THIS EXPLAINS THE PRESENCE OF AN ENERGY TERM THAT VARIES INVERSELY WITH THE ACTIVE PROCESSING TIME IN AN ISI. IN PRACTICE, THIS TERM BECOMES NEGLIGIBLE ONCE THE ACTIVE PROCESSING TIME EXCEEDS  $\tau$ ; IN OTHER WORDS, THE RRP IS APPROXIMATELY  $\tau$  SECONDS LONG.

\* IT'S QUITE POSSIBLE THE NEURON FINDS SOME WAY TO RECOVER A GOODLY FRACTION OF THIS ENERGY IN SOME MANNER AKIN TO REGENERATIVE BRAKING, BUT EVEN IF 90% IS RECOVERED, THE OTHER 10% STILL IS PRORTIONAL TO  $\tau / T^*$ .<sup>32</sup>

# HOW NEURONS LEARNED HOW TO DO IT ALL !!





**THE END**