Energy-Efficient Neuronal Computation via Quantal Synaptic Failures

William B Levy[1] and Robert A. Baxter[1,2]

[1]*University of Virginia Health System, Department of Neurosurgery, Charlottesville, Virginia 22908, and*
[2]*Baxter Research Company, Bedford, Massachusetts 01730*

Organisms evolve as compromises, and many of these compromises can be expressed in terms of energy efficiency. For example, a compromise between rate of information processing and the energy consumed might explain certain neurophysiological and neuroanatomical observations (e.g., average firing frequency and number of neurons). Using this perspective reveals that the randomness injected into neural processing by the statistical uncertainty of synaptic transmission optimizes one kind of information processing relative to energy use. A critical hypothesis and insight is that neuronal information processing is appropriately measured, first, by considering dendrosomatic summation as a Shannon-type channel (1948) and, second, by considering such uncertain synaptic transmission as part of the dendrosomatic computation rather than as part of axonal information transmission. Using such a model of neural computation and matching the information gathered by dendritic summation to the axonal information transmitted, $H(p^*)$, conditions are defined that guarantee synaptic failures can improve the energetic efficiency of neurons. Further development provides a general expression relating optimal failure rate, f, to average firing rate, p^*, and is consistent with physiologically observed values. The expression providing this relationship, $f \approx 4^{-H(p^*)}$, generalizes across activity levels and is independent of the number of inputs to a neuron.

Key words: computation; efficiency; energy; entropy; information theory; mutual information; optimization; quantal failures; Shannon

This paper interrelates three topics: synaptic failure rates, dendrosomatic information processing, and neuronal energy use. As an introduction we briefly address each topic.

In the hippocampus and in neocortex, excitatory synaptic connections dominate and are remarkably unreliable. Each synapse transmits, at most, a single standardized package called a quantum ($\sim 10^4$ neurotransmitter molecules). When an action potential arrives presynaptically, the probability of evoking the release of one such quantal package is reported to range from 0.25 to 0.5 with 0.5 being less common and 0.25 being quite common (Thomson, 2000), especially when one takes into account the spontaneous rates of neurons (Stevens and Wang, 1994; Destexhe and Paré, 1999). The failure of quantal synaptic transmission is a random process (Katz, 1966) and is counterintuitive when it exists under physiological conditions. After all, why go to all the trouble, and expense, of transmitting an action potential if a synapse does not use it. The observation of synaptic failures is particularly puzzling in light of observations outside of neocortex showing failure-free, excitatory synaptic transmission can exist in the brain (Paulsen and Heggelund, 1994, 1996; Bellingham et al., 1998). One insight that clarifies this puzzle is that systems with low failure rates tend to form clusters of synapses on a single postsynaptic target neurons that are *boutons terminaux*, whereas those that fail are predominantly forming *en passage* synapses with multiple (thousands or tens of thousands) of postsynaptic neurons. For *en passage* systems, a particular spike works at some synapses but not at others. Thus, in such *en passage* situations, failure at the axon hillock is not equivalent to random synaptic failure because a large number of synapses will transmit, just not a large percentage. Still failures have the feeling of inefficiency. Here we show that the quantal failures can be viewed as an energy efficiency mechanism relative to the information that survives neuronal information processing. That is, under certain circumstances failures will not lower the transmitted computational information of a postsynaptic neuron, but they will lower energy consumption and heat production. The relationship between energy and information has, at least implicitly, been an issue in physics since the time of Maxwell (Leff and Rex, 1990). Today this relationship continues to be discussed particularly because energy consumption, or heat generation, may place the ultimate limits on manmade computation. In the context of biological computation, such issues also seem particularly relevant because of the large fraction of our caloric intake that goes directly and indirectly toward maintaining brain function (Sokoloff, 1989; Attwell and Laughlin, 2001). Indeed because of such costs, we proceed under the hypothesis that, microscopically, natural selection has approximately optimized energy use as well as information processing in constructing the way neurons compute and communicate. Recent successes and interest arising from this hypothesis of a joint optimization (Levy and Baxter, 1996; Laughlin et al., 1998; Andreou, 1999; Abshire and Andreou, 2001; Balasubramanian et al., 2001, Schreiber et al., 2001) encourage us to continue examining the possibility that neuronal communication and computation are efficient when considered in the dual context of energy and information rather than either context

alone. Particularly encouraging is the energy audit of Attwell and Laughlin (2001). This work concludes that >85% of the energy consumed by the neocortical neuropil goes toward recovering from the ion fluxes that are, in effect, all of the computation and communication within the neocortex.

MEASURING INFORMATION

At the level of an individual neuron, neuronal computation can be sensibly quantified by viewing a computational transformation as a communication system. The aptness of using information-theoretic ideas for analyzing analog computation was pointed out by von Neumann (see Bremermann, 1982) and by Bremermann (1982) and is one of several possible measures that seems worth calculating to quantify analog computation. Although they give us no details, most simply an analog computation is just a transformation as $X \rightarrow f(X)$ so that mutual information, $I(X; f(X))$, is obviously relevant and can be aptly called the information available from neuronal integration.

As is traditional (Shannon, 1948), mutual information is defined as:

$$I(X;Y) \overset{def}{=} E_{xy}\left[\log \frac{P(XY)}{P(X)P(Y)} \right] = H(X) - H(X|Y) = H(Y) - H(Y|X),$$

where

$$H(X) \overset{def}{=} H(P(X)) = - \sum_x P(X) \log P(X)$$

and

$$H(X|Y) \overset{def}{=} - \sum_y P(Y = y) \sum_x P(X|Y = y) \log P(X|Y = y)$$

are Shannon entropies, and logarithms are base two.

When the conditional entropy is zero (e.g., $H(X|Y)$), then mutual information, $I(X;Y)$, equals the entropy $H(X)$. Because this is true for neocortical axons (Mackenzie and Murphy, 1998; Cox et al., 2000; Goldfinger, 2000), we were able to use entropy rather than mutual information when studying axons.

Previously (Levy and Baxter, 1996) we noted that, solely in the context of signaling information capacity, or equivalently representational capacity, information alone is not optimized by neocortical neurons. In the neocortex, where the maximum spike frequency of a pyramidal neuron is ~400 Hz, the average rate of axonal spiking is 10–20 Hz, not the 200 Hz optimal for information transmission alone. At the other extreme, there would be no energetic cost if a neuron did not exist, so energy alone is not optimally conserved. However, forming the ratio of information transmitted by an axon to the energy it consumes (a measure whose ultimate dimension is bits per joule) leads to an optimal spike rate value that fits with observed values of spike rates and energy consumption (Levy and Baxter, 1996). This particular optimization is critical to what follows.

The information flow for a single neuron is depicted in Figure 1*A*. The notation and its biological correspondence are as follows. The random multivariate binary input to a neuron is X, and the output of this neuron is Z, a univariate random binary variable, {no spike, spike} ≡ {0, 1}. The spike generator, which in our model absorbs many dendritic nonlinearities, determines when dendrosomatic excitation exceeds threshold. Then $Z = 1$, and the spike is conducted by the axon away from the neuron, eventually arriving presynaptically as Z' where the cycle begins again. Our specific interest here is the computational transformation that includes the quantal release process. As depicted in Figure 1*A*, input signals to a neuron undergo three information-losing transformations before the transformation by the spike generator: (1) quantal release–failure, (2) quantal amplitude variation, and (3) dendrosomatic summation. The release–failure process (Fig. 1*B*) produces a new binary random variate $\phi(X_i)$. The probability of a quantal failure is denoted by f, whereas the probability of a successful quantal release is denoted by s, and $f = 1 - s$. The random variate Q_i denotes the amplitude of the ith input when release occurs. Using this notation, the information passing through the computation is explicitly expressed as the mutual information $I^C \overset{def}{=} I(X; \Sigma \phi(X_i)Q_i) = H(X) - H(X|\Sigma_i \phi(X_i)Q_i)$. Also, the lack of spontaneous spikes and the faithful conduction of neocortical axons implies $Z = Z'$ so that $I(Z;Z') = H(P(Z))$, as mentioned earlier.

It is also useful to introduce the notation for the energy-optimal capacity of the axon, C^E, which occurs at $\max_{P(Z=1)}[H(P(Z))/axonal\ energy\ use]$, and as well p^* the value of $P(Z = 1)$ that produces C^E. From our earlier calculations (Levy and Baxter, 1996) and from neurophysiological observations of sensory and association cortex, p^* ranges from .025 to 0.05 per minimum interspike interval (approximated as 2.5 msec for a synaptically driven pyramidal neuron (Levy and Baxter, 1996)). This produces C^E values ranging from 0.169 to 0.286 bits per 2.5 msec. Importantly, we will suppose that both input and output neurons adhere to the same optimum.

We explicitly assume that there is an inconsequential information loss by the spike generator and that the cost of generating extra spikes is negligible. This later assumption is justified by the energy audit of Attwell and Laughlin (2001). Attwell and Laughlin (2001) showed that the energetic costs associated with action potential production in a functioning brain are highest in axons, with ~47% of the total energy consumed (which is in agreement with Levy and Baxter, 1996) The next highest cost is associated with dendritic excitation, which is ~34% of the total energy consumed. A relatively small amount goes to the presynaptic aspects of synaptic transmission. Perhaps the lowest cost (which is negligible) is associated with the cell body because cell bodies have such small surface areas relative to axons and dendrites. In our model, we assume that the spike generator is part of the cell body. Therefore, the cost of generating extra spikes is negligible, whereas the cost of conducting the spike down the axon is quite high. Regardless of that cost, information must still be transmitted. That is, even if one were compelled to postulate failure at the spike generator, one is still left with an average axonal usage (firing) rate of p^*. Thus, it is our explicit hypothesis that information is transmitted at the optimal rate, $H(p^*)$, and we are now in a position to be much more explicit about energetically efficient computation.

Conjecture. Maximize the computational information developed by a neuron and its inputs to no more than the limit imposed by the information capacity of the axon whose capacity is set by optimizing the energy efficiency of its signaling.

That is, if the axonal transmitting system is energy optimized to $H(p^*)$, then this rate is an upperbound constraint on the computational information that can be transmitted given the hypothesized spike-generating process. Moreover, when failure rates are zero, the computational information will always have a potential to be greater than $H(p^*)$ because this is the amount that would be available after noise free processing by a neuron with more than just a single input. Because failure rates are not zero, this conjecture leads to the hypothesis that failure rates reduce the energy consumption of computation while not wasting any of (that is, while using all of) the axonal capacity.

Quantal failures are an excellent mechanism to create this matching because of the energy they save. [If every successive step from the arrival of an action potential presynaptically down to the depolarization of the cell body is energy consuming (Attwell and Laughlin, 2001) then a mechanism that eliminates as many of these steps as possible will save the most energy. Specifically, failure of synaptic transmission saves the cost of vesicle recycling, transmitter reuptake and repackaging, and most of all it saves on the cost of postsynaptic depolarization.] Moreover, because both the information of computation and of optimal channel use are both controlled by p^*, we can determine a failure rate that brings computational information exactly to its maximally transmittable rate. This failure rate then saves as much energy as possible while still allowing the neuron to develop the maximally transmittable information. Curiously the optimal failure rate quickly becomes independent of the number of inputs, and it is in the range of number of inputs that neocortical (and indeed, many other neurons) operate.

In sum, it is our explicit hypothesis that neural computation, as well as neural communication, can be measured from the Shannon perspective of sources and channels. In pursuing an overall analysis, we have opted to partition function. As a result of this partitioning, the physical correspondence between source or channel changes as the separate parts or functions of a neuron are sequentially analyzed. For example, in our previous work an axon is a channel, whereas here the set of axons going into a neuron are an information source. Here the synapses and dendritic summation process are analyzed as if they are a channel. But as we shall see, they will also be viewed as a source for the next stage. But first, let us develop some quantitative intuition by considering a bounding case.

Special case. $f = 0$ and $Q_i = 1$ for all i. If we consider the case with a zero failure probability and all $Q_i = 1$ then $I^C = H(\Sigma X_i) - H(\Sigma X_i|X)$, and we easily obtain $H(\Sigma X_i)$ as an upper bound on the mutual informa-

Figure 1. Partitioning communication and computation for a single neuron and its inputs. *A,* The presynaptic axonal inputs to the postsynaptic neuron is a multivariate binary vector, $X = [X_1, X_2, \ldots, X_n]$. Each input, X_i, is subject to quantal failures, the result of which is denoted by $\phi (X_i)$, another binary vector that is then scaled by quantal amplitude, Q_i. Thus, each input provides excitation $\phi(X_i)Q_i$. The dendrosomatic summation, $\Sigma_i \phi(X_i)Q_i$ is the endpoint of the computational process, and this sum is the input to the spike generator. Without specifying any particular subcellular locale, we absorb generic nonlinearities that precede the spike generator into the spike generator, $g (\Sigma_i \phi(X_i)Q_i)$. The spike generator output is a binary variable, Z, which is faithfully transmitted down the axon as Z'. This Z' is just another X_i elsewhere in the network. In neocortex, experimental evidence indicates that axonal conduction is, essentially, information lossless, as a result $I(Z; Z') \approx H(Z)$. The information transmitted through synapses and dendrosomatic summation is measured by the mutual information $I(X; \Sigma \phi(X_i)Q_i) = H(X) - H(X|\Sigma_i \phi(X_i)Q_i)$. Given the assumptions in the text combined with one of Shannon's source-channel theorems implies that, $H(X) - H(X|\Sigma_i \phi(X_i)Q_i) = H(p^*)$, where $H(p^*)$ is the energy-efficient maximum value of $H(Z)$. *B,* The model of failure prone synaptic transmission. An input value of 0, i.e., no spike, always yields an output value of 0, i.e., no transmitter release. An input value of 1, an axonal spike, produces an output value of 1, transmitter release, with probability success $s = 1 - f$. A failure occurs when an input value of 1 produces an output value of 0. The probability of failure is denoted by f.
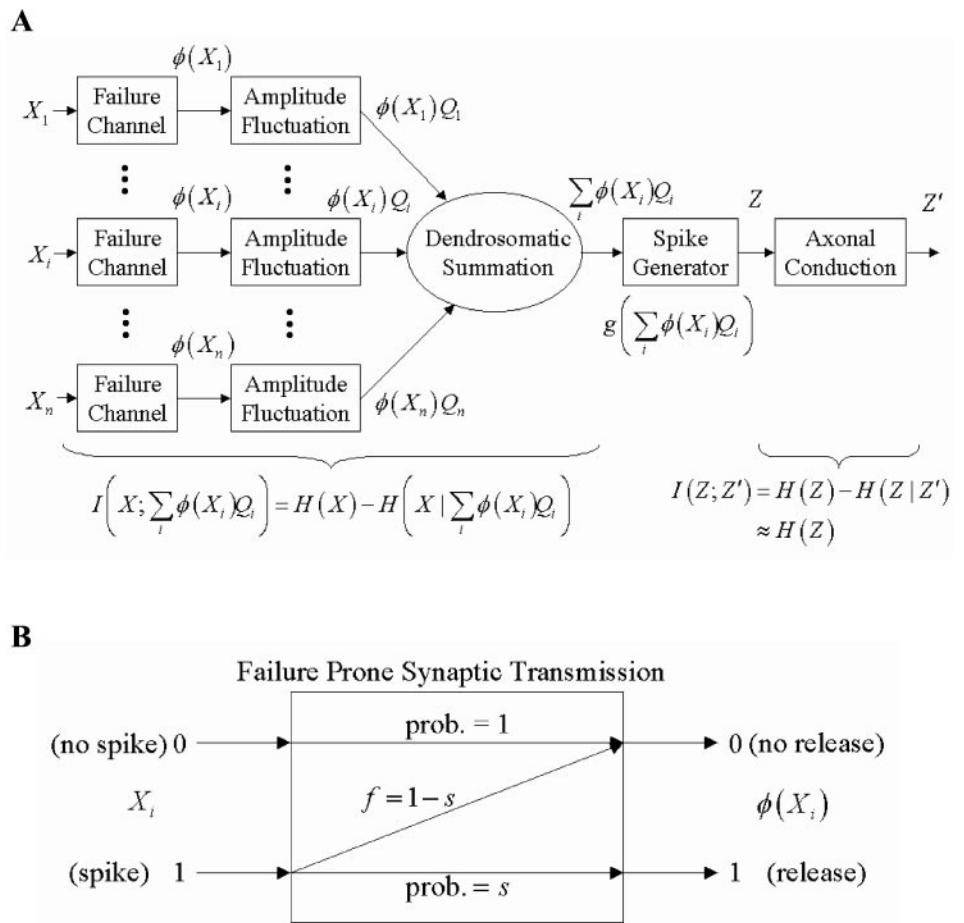
tion of the computation. This upper bound occurs when the second term is zero, i.e. in the failure-free, noise-free situation with all quanta the same size. Appealing to the central limit theorem, this entropy is well approximated by the entropy of a normal distribution. Therefore, if we suppose each of the *n* inputs is an independent Bernoulli process with the same parameter $p = p^*$, we get:

$$H(\Sigma X_i) = \frac{1}{2} \log_2[2\pi e n p^*(1 - p^*)] = 6.5 \text{ bits for } n = 10^4 \text{ and } p^* = 0.05,$$

where this value of p^* comes from the Levy and Baxter (1996) calculations as well as the actual observed value of average firing rates in neocortex. Although 6.5 bits is a tremendous drop from $H(X)$, which under these assumptions is 2860 bits (10,000 inputs each with 0.286 bits), this 6.5 bits is still a very large number of bits to be transmitted per computational interval compared to the energy-efficient channel capacity of $H(p^*) = 0.286$ bits.

The reason why 6.5 bits is a tremendous excess arises when we consider Shannon's source/channel theorems. These say that the channel limits the maximum transmittable information to its capacity. As a result, any energy that goes toward producing $I(X; \Sigma X_i)$ that exceeds the channel capacity $H(p^*)$ is wasted information. This idea is at the heart of the analysis that follows. Because the total information of the computation is many times the energy-efficient channel capacity, much waste is possible. Indeed, even if we dispense with the independence assumption (while still supposing some kind of central limit result holds for the summed inputs) and suppose that statistical dependence of the inputs is so bad that every 100 inputs act like 1 input, an approximation that strikes us as more than extreme, there still are too many bits (~3.2 bits) being generated by the computation compared with what can be transmitted. Thus, the computation is not going to be energy-efficient if it takes energy (and it does) to develop this excess, nontransmittable computational information.

RESULTS

We now begin the formal analysis that substantiates and quantifies the conjecture and that brings to light a set of assumptions making the conjecture true.

Assumptions

A0: A computation by an excitatory neuron is the summation of its inputs every computational interval. The mutual information of such information processing is closely approximated as:

$$I^C = I(X; \sum_i \phi(X_i)Q_i).$$

A1: Axons are binary signaling devices carrying independent spikes and used at their energy optimum; that is, each axon is used at the information rate C^E bits per computational interval, which implies firing probability p^*.

A2: The number of inputs to a neuron is not too small—say $n > 2/p^*$. Clearly this is true in neocortex; see Fig. 3 for evaluation of this assumption.

A3: With the proviso that A1 and A2 must be obeyed, a process requiring less energy is preferred to a process requiring more energy.

A4: The spike generator at the initial segment, which incorporates generic nonlinearities operating on the linear dendritic summation, creates a bitwise code suitable for the axonal channel, and this encoding is nearly perfect in using the information received from the dendrosomatic computation. That is, as an

information source the spike generator produces information at a rate of nearly $H(p^*)$.

From these assumptions we have a lemma.

Lemma 1: $I^C \geq H(p^*)$. That is, the only way to use an axon at its energy optimal rate, $H(p^*)$, is to provide at least that much information to it for possible transmission.

Proof by contradiction: Providing anything less would mean that the axon could be run at a lower rate than implied by p^* and as a result save energy while failing to obtain its optimal efficiency which contradicts (A1).

The importance of this lemma is the following: no process that is part of the computational transformation or part of energy saving in the computation or part of interfering fluctuations arising within the computation should drive I^C below $H(p^*)$. In particular, this lemma dictates that quantal failures, as an energy saving device, will be used (or failure rates will be increased) only when I^C is strictly greater than $H(p^*)$.

With this lemma and assuming increased synaptic excitation leads to monotonically increasing energy consumption (Attwell and Laughlin, 2001), we can prove a theorem that leads to an optimal failure rate. Thus, the averaged summed postsynaptic activation, $E[\Sigma_i \phi(X_i)Q_i]$, should be as small as possible because of energy savings (A3), whereas (A2) maintains n and (A1) maintains p^*. This restricted minimization of average synaptic activation implies processes, including synaptic failures, that reduce energy use. But when operating on the energy-efficient side of the depolarization versus information curve, reducing the average summed activation monotonically reduces I^C as well as reducing energetic costs with this reduction of I^C unrestricted until Lemma 1 takes force. That is, this reduction of I^C should go as far as possible because of A3-(energy saving) but no lower than $H(p^*)$ because of the lemma. As a result, energy optimal computation is characterized by:

$$I^C = C^E = H(p^*),$$

an equality that we call "Theorem G." Accepting Theorem G leads to the following corollary about synaptic failures:
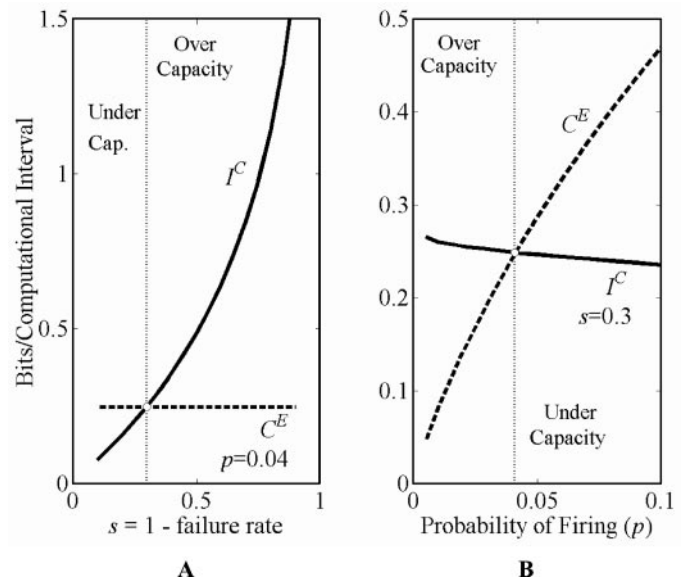
Corollary F

Provided $np^* > 2$, neuronal computation is made more energy-efficient by a process of random synaptic failures (see Appendix and below).

Obviously failures are in the class of processes that lower average postsynaptic excitation in part because I^C is reduced uniformly as f increases and, in part, because the associated energy consumption is also reduced uniformly. Just below and in the appendix we prove a quantified version of Corollary F that shows that the failure rate f producing this optimization is approximated purely as a function of p^*; specifically,

Quantified Corollary F

$$f \approx \left(\frac{1}{4}\right)^{H(p^*)}.$$

Figure 2A illustrates the existence of a unique, optimal failure rate by showing the intersection between C^ϵ, the energy-efficient capacity of the axon, with I^C, the information of the computation. Here we have used $n = 10^4$, $p^* = 0.041$. From another perspective, Figure 2B shows how one might take some physiologically appropriate failure rate, $f = 0.7$, and determine the optimal p. In either case we note the single intersection of the two monotonic curves.

Figure 2. A, The optimal failure rate $(1 - s)$ of theorem G and corollary F is obtained by noting the intersection of the two curves, I^C (the computational information) and $C^E = H(p^*)$ (the output channel capacity). At higher values of s, any input information greater than $H(p^*)$ that survives the input-based computational process of summation is wasted because the information rate out cannot exceed $H(p^*)$, the output axonal energy-efficient channel capacity. These values define an overcapacity region. For lower values of s, neuronal integration is unable to provide enough information to the spike generator to fully use the available rate of the axon. This is the undercapacity region. Of course, changing p^* changes the optimal failure rate because the C^E curve will shift. These curves also reveal that a slight relaxation of assumption A4 will not change the intersection value of s very much (e.g., a 10% information loss at the spike generator produces a <3% change in the value of s). The success rate s equals one minus the failure rate. The optimal success rate is demarcated by the *vertical dotted line*. In this figure the output channel capacity, $H(p^*)$, uses $p^* = 0.041$; $n = 10,000$ inputs. B, An alternative perspective. Assuming the failure rate is given as 0.7 by physiological measurements, then we could determine p^*, the p that matches computational information I^C to the energy-efficient channel capacity. Again the *vertical dotted line* indicates the predicted value; $n = 10,000$. Both A and B are calculated using the binomial probabilities of the Appendix.

The generality of what this figure shows is established in the Appendix. Specifically, Appendix Part A assumes equal Q_i values, whereas Parts B and C allow for Q_i to vary; they show:

$$-\frac{1}{2}\log(f) \approx I(X; \Sigma_i \phi(X_i)Q_i).$$

Because Theorem G requires $I(X, \Sigma_i \phi(X_i)Q_i) = H(p^*)$, the two results combine, yielding

$$f \approx \left(\frac{1}{4}\right)^{H(p^*)},$$

a statement that is notable for its lack of dependence on n, the number of inputs to a neuron. This lack of dependence, illustrated for one set of values in Figure 3, endows the optimization with a certain robustness. Moreover, the predicted values of f also seem about right. For example, $p^* = 0.05$, implies $f = 0.67$, whereas other values can be read off of Figure 4. So, by choosing a physiologically observed p^*, the relationship produces failure rates in the physiologically observed range. Thus, on these two accounts (the robustness and the prediction of one physiological observation from another nominally independent experimental
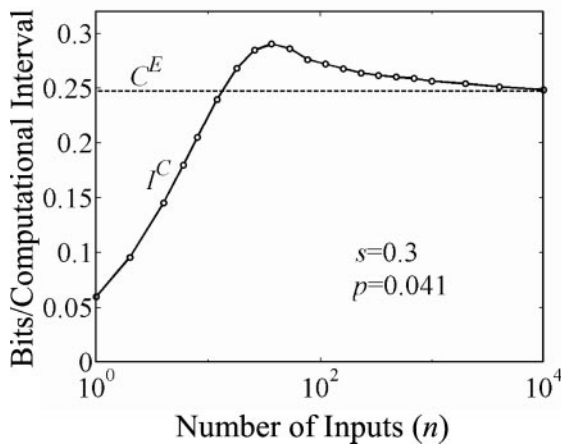
Figure 3. At the optimal failure rate, matching I^C to C^E is increasingly robust as number of inputs, n, increases. Nevertheless I^C, the mutual information measure of computation, attains the approximate value of output capacity, C^E, for n as small as 200. Calculations used the binomial distributions of the Appendix with failure rate fixed at 0.7 and p^* set to 0.041. The *dashed line* indicates $H(p^*)$.
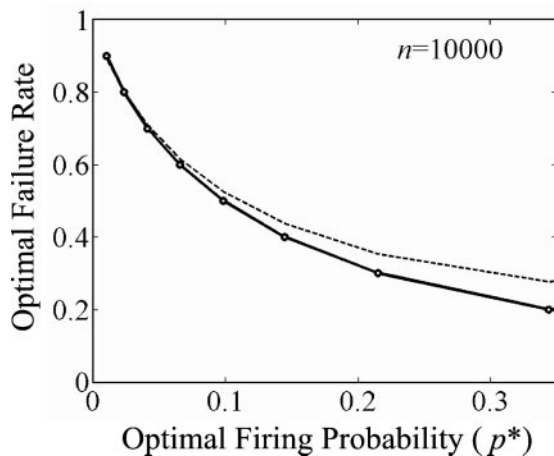
Figure 4. Optimal failure rate as a function of spike probability in one computational interval. The optimal failure rate decreases monotonically as firing probability increases so that this theory accommodates a wide range of firing levels. The vicinity of physiological p^* (0.025–0.05 for nonmotor neocortex and limbic cortex) predicts physiologically observed failure rates. The *dashed line* plots $f = (1/4)^{H(p^*)}$, whereas the *solid line* is calculated without the Gaussian approximations described in the Appendix. Note the good quality of the approximation in the region of interest ($p^* \approx .05$), although for very active neurons the approximation will overestimate the optimal failure rate. More important than this small approximation error, we would still restrict this theory to places where information theoretic principles, as opposed to decision theoretic or control theoretic principles, best characterize information processing.

observation), we reap further rewards from the analysis of microscopic neural function in terms of energy-efficient information.

The more involved proof of Appendix Part B sheds light on the size of one source of randomness (quantal size) relative to another (failure rate). Taking the SD of quantal size to be 12.5% of the mean quantal size leads to an adjustment of about $s/65$ in the implied values of f. For example, suppose no variation in Q_i produces an optimal failure rate of 70%, then taking variation of Q_i into account adjusts this value up to 70.46%. Clearly the effect of quantal size variation is inconsequential relative to the failure process itself.

DISCUSSION

In addition to the five assumptions listed on page 11, we made two other implicit assumptions in the analysis. First, we assumed additivity of synaptic events. While this assumption may seem unreasonable, recent work (Magee, 1999, 2000; Andrásfalvy and Magee, 2001) and (Cook and Johnston, 1997, 1999; Poolos and Jonston, 1999) make even a linear additivity assumption reasonable. The observations of Destexhe and Paré (1999), showing a very limited range of excitation, also makes a linear assumption a good approximation. Even so, we have explicitly incorporated any nonlinearities that might operate on this sum and then group this nonlinearity with the spike generator. Second, we have assumed binary signaling. Very high temporal resolution, in excess of $2-10^4$ Hz, would allow an interspike interval code that outperforms the energetic efficiency of a binary code. Our unpublished calculations (which of necessity must guess at spike timing precision including spike generation precision, spike conduction dither, and spike time decoding precision; specifically, a value of 10^{-4} msec was assumed) indicate a p^* for such an interspike interval code would be ~50% greater than the p^* associated with binary coding as well as being more energetically efficient. However, we suspect such codes exist only in early sensory processing and at the input to cerebellar granule cells. Systems, such as considered here, with single quantum synapses, quantal failures, and 10–20 Hz average firing rates, would seem to suffer inordinately using interspike interval codes; a quantal failure can cause two errors per failure and observed firing rates are suboptimal for interspike interval code but fit the binary hypothesis.

The relationship $f \approx 4^{-H(p^*)}$ partially confirms, but even more so, corrects the intuition that led us to do this analysis. That is, we had thought that the excess information in the dendrosomatic computation could sustain synaptic failures and still be large enough to fully use the energy-efficient capacity of the axon, C^E. However, this same intuitive thinking also said that the more information a neuron receives, i.e., as either p^* or as n grows, the more a failure rate can be increased, and this thought is wrong with regard to both variables.

First, the relationship $f \approx 4^{-H(p^*)}$ tells us that the optimal failure rate actually decreases as p^* increases, so intuitive thinking had it backwards. We had thought in terms of the postsynaptic neuron adding up its inputs. In this case, the probability of spikes is like peaches and dollars, the more you possess the less each one is worth to you. This viewpoint led to the intuition that, when there are more spikes, any one of them can be more readily discarded; i.e., f can be safely increased when p increases. However, this intuition ignored the output spike generator that neuronal integration must supply with information. Here at the generator (and its axon and each of its synapses) the probability of spikes is very different than peaches and dollars: because the curve for binary entropy, $H(p)$, increases as p increases from 0 to 1/2, increasing probability effectively increases the average worth of each spike and, as well, nonspikes; so it is more costly to discard one. This result, one that only became clear to us by quantifying the relationships, leads to optimal failure rates that are a decreasing function of p^*.

Second, in the neocortically relevant situation, where n is in the thousands, if not tens of thousands, changing n has essentially no effect on the optimal failure rate (Fig. 3). Indeed, the lower bound, (A3), is so generous relative to actual neocortical connectivity, that there is no way to limit connectivity (and thus, no way to optimize it) based on saving energy in the dendrosomatic

computation modeled here. To say it another way, one should look elsewhere to explain the constraints on connectivity [e.g., ideas about volume constraints as in Mitchison, 1991 or Ringo, 1991, or ideas about memory capacity (Treves and Rolls, 1992).]

Thus, we are forced to conclude that, to a good approximation, once n times p^* is large enough, I^C depends only on the failure rate, an observation that is visualized by comparing the I^C curves of Figures 2, A and B, and 3.

In sum, the failure channel can be viewed as a process for lowering the energy consumption of neuronal information processing, and synaptic failures do not hurt the information throughput when the perspective is broad enough. More exactly, the optimized failure channel decreases energy consumption by synapses and by dendrites while still allowing the maximally desirable amount of information processing. This result is achieved when $I^C = C^E = H(p^*)$, and this condition implies the optimal failure rate is solely a function of p^*.

Finally, optimizations such as these support the long, strong (Barlow, 1959; Rieke et al., 1999; Dayan and Abbott, 2001) and now increasingly popular [e.g., inter alia (Bialek et al., 1991; Tovee and Rolls, 1995; Theunissen and Miller, 1997; Victor, 2000; Atwell and Laughlin, 2001), see also articles in Abbott and Sejnowski (1999)] tradition of analyzing brain function using Shannon's ideas. Successful parametric optimizations like the one presented here (and those produced in some of the previously cited references), reinforce the validity of using entropy-based measures to describe and analyze neuronal information processing and communication. Such results also stimulate hypotheses (Weibel et al., 1998): e.g., not only does natural selection take such measures to heart but often does so in the context of energy efficiency.

APPENDIX

In this part we relate the quantal failure rate, f, to H(p*), the energy-efficient channel capacity. Parts A, B, and C produce essentially the same result, but Parts A and C are simpler. Part A develops the result when the failure process is assumed to be by far the largest source of noise. Parts B and C relax this assumption to include the effect of variable quantal size which (as shown in Part B) turns out to be negligibly small. Throughout, we assume all synaptic weights are identically equal to one. However, a small number of multiple synapses can accommodate variable synaptic strength without changing the result.

Part A

A neuron receives an n-dimensional binary input vector X. Each component of the input vector, X_i, is a Bernoulli random variable, $P(X_i = 1) = p^*$. Define $y = \Sigma X_i$ as a realization of the random variables summed (without quantal failures). The failure process, $\phi()$, produces a new random variable denoted $\phi(X_i)$. Then denote $y^\varphi = \Sigma \phi(X_i)$ as a realization of the summed input subject to the failure process. The quantal success rate, the complement of the failure rate, is $P(\phi(X_i) = 1|X_i = 1) \overset{def}{=} s = 1 - f$ and the other characteristic of such synapses is $P(\phi(X_i) = 0|X_i = 0) = 1$. We want to examine the mutual information between X and $\Sigma \phi(X_i)$, when Theorem G is obeyed. That is, when $I(X; \Sigma \phi(X_i)) = H(p^*)$.

First note that because the failure channel at one synapse operates independently of all other inputs defined by X and

because the sums $\Sigma X_i = y$ partition the X values and that $P(\Sigma \phi(X_i)|X = x$ where $x \Rightarrow \Sigma X_i = y) = P(\Sigma \phi(X_i)|\Sigma X_i = y)$, then

$$H(\sum \phi(X_i)|X) = H(\sum \phi(X_i)| \sum X_i)$$

so that

$$I(X; \sum \phi(X_i)) = I(\sum X_i; \sum \phi(X_i)).$$

We assume that ΣX_i can be modeled as a Poisson random variable with parameter $\lambda = np$, where n is the number of inputs (e.g., $n \approx 10,000$ and $p \approx 0.05$ so $\lambda \approx 500$). That is,

$$P(\sum X_i = y) = \frac{e^{-\lambda}\lambda^y}{y!}$$

and, likewise when the failure mechanism is inserted:

$$P(\sum \phi(X_i) = y^\phi) = e^{-\lambda s}\frac{(\lambda s)^{y^\phi}}{y^\phi!},$$

another outcome evolving from the fact that quantal failures occur independently at each activated synapse. On the other hand, the summed response conditioned on ΣX_i is binomial with parameters (y, s); that is,

$$P(\Sigma \phi(X_i) = y^\phi|\Sigma X_i = y) = \frac{y!}{(y - y^\phi)!y^\phi!} s^{y^\phi}(1 - s)^{y-y^\phi}.$$

But this inconvenient form, with its normalization term depending on the conditioning variable can be reversed. Using the definition of conditional probabilities and our knowledge of the marginals: Thus,

$$P(\Sigma X_i = y|\Sigma \phi(X_{i_i}) = y^\phi) = e^{-\lambda(1-s)}\frac{[\lambda(1 - s)]^{y-y^\phi}}{(y - y^\phi)!}$$

$$= \frac{e^{-\lambda(1-s)}(\lambda(1 - s))^t}{t!}$$

where $t = y - y^\phi$, and note that $t \geq 0$ because of the way the failure channel works (i.e., $y \geq y^\phi$).

Now it can be seen that both $P(\Sigma X_i = y^\phi)$ and $P(Y = y|\Sigma \phi(X_i) = y^\phi)$ are Poisson distributions with parameters λ and $\lambda(1 - s)$, respectively. Greatly simplifying further calculations, this second Poisson parameter is independent of its conditioning variable y^ϕ, and is particularly easy to conditionally average because all summations occur over the same range; i.e., note that:

$$P(\sum X_i = y|\sum \phi(X_{i_i}) = y^\phi) = P(\sum X_i = y, \sum \phi(X_i)$$

$$= y^\phi|\sum \phi(X_i) = y^\phi) = P(t|\sum \phi(X_i) = y^\phi)$$

$$= \frac{e^{-\lambda(1-s)}(\lambda(1 - s))^t}{t!} \text{ with } t \in \{0, 1, 2, \ldots\} \text{ regardless of } y^\phi.$$

To compute $I(\Sigma X_i; \Sigma \phi(X_i))$ we will use the relation:

$$I(\sum X_i; \sum \phi(X_i)) = H(\sum X_i) - H(\sum X_i| \sum_i \phi(X_i))$$

and the normal approximation for the entropy; that is, for a Poisson distribution with parameter (i.e., variance) λ large

enough, the entropy is very nearly $\log_2 \sqrt{2\pi e \lambda}$. So for the two-distributions with the parameters λ and $\lambda(1 - s)$, respectively, subtracting one entropy from the other yields

$$I(\sum X_i | \sum \phi(X_i)) = -\frac{1}{2}\log_2(1 - s) = -\frac{1}{2}\log_2(f).$$

Thus, $I(\Sigma X_i | \Sigma \phi(X_i))$ matches the energy-efficient channel capacity p^* when $f = (1/4)^{H(p^*)}$. So the quantal failure rate is uniquely determined by p^* and is independent of the number of inputs n provided that n is sufficiently large. [In fact, sufficiently large is not very large at all. When the Poisson parameter is two, the relative error of the normal approximation is <4% (Frank and Öhrvik, 1994).] Furthermore, because $H(p) \leq 1$ the quantal failure rate has the lower bound, $f \geq 0.25$.

Part B

The following mathematical development accounts for the variation in synaptic excitation caused by the failure channel and plus the variance in quantal size. A number of approximations are involved, but they are all of the same type. That is, we will go back and forth between discrete and continuous distributions (and back and forth between summation and integration) with the justification that if two distributions are approximated by the same normal distribution, they can be used to approximate each other.

Each successful transmission, $\phi(X_i) = 1$, results in a quantal event size $Q_i = q_i$.

The number of synapses is n, and the failure rate is $f = 1 - s$, where s is the success rate. Each input is governed by a Bernoulli variable p.

$X \in \{0,1\}^n$, the vector of inputs; $\phi(X) \in \{0,1\}^n$, the vector of active inputs passing through the failure channel; $\Sigma X_i \in \{0, 1, \ldots, n\}$, the number of active input axons; $\Sigma \phi(X_i) \in \{0, 1, \ldots, n\}$, the number of successful transmitting synapses; $Q_i \in \{0, 1, \ldots, m\}$, a quantal event with a discrete random amplitude; $\Sigma \phi(X_i)Q_i \in \{0, 1, \ldots, n\cdot m\}$ the sum of the quantal responses with quantal amplitude variation.

Upper case indicates a random variable and lower case a specific realization of the corresponding random variable. The actual (biological) sequences of variables and transformation is:

$$X \rightarrow \phi(X) \rightarrow \begin{bmatrix} \phi(X_1)Q_1 \\ \vdots \\ \phi(X_i)Q_i \\ \vdots \\ \phi(X_n)Q_n \end{bmatrix} \rightarrow \sum \phi(X_i)Q_i \quad \text{(B.1.1)}$$

Because the Q_i are independent of i and identically distributed, there is a mutual information equivalent sequence:

$$X \rightarrow \phi(X) \rightarrow \sum \phi(X_i) \rightarrow \sum \phi(X_i)Q_i, \quad \text{(B.1.2)}$$

which is what we will use to calculate:

$$I(X; \sum \phi(X_iQ_i) = H(\sum \phi(X_i)Q_i) - H(\sum \phi(X_i)Q_i | X). \quad \text{(B.1.3)}$$

Part B.1

Find $H(\Sigma \phi(X_i)Q_i)$. To get $H(\Sigma \phi(X_i)Q_i)$, we need $P(\Sigma \phi(X_i)Q_i)$, which we get via $P(\Sigma \phi(X_i)Q_i = h) = \Sigma_k P(\Sigma \phi(X_i)Q_i = h|\Sigma \phi(X_i) = y^\phi) P(\Sigma \phi(X_i) = y^\phi)$ and some approximations.

The first approximation

When nps is large and ps is small, approximate the discrete distribution $P(\Sigma \phi(X_i) = y^\phi) = \binom{n}{y^\phi} (ps)^{y^\phi} (1 - ps)^{n - y^\phi}$ by the continuous distribution

$$\approx \frac{e^{-y^\phi}(y^\phi)^{nps - 1}}{Gamma(nps)} \quad \text{(B.1.4)}$$

because each is nearly normally distributed and has the same mean and nearly the same variance.

To take quantal sizes into account, start with the assumption that, the single event, Q_i, is distributed as a Poisson that is large enough to be nearly normal and that has the same shape. We can do this if the value of the random variable at the mean and at one SD on either side of the mean yields the same approximate relationship (it will) while the Poisson parameter is large enough for the normal approximation. For now let $P(Q_i = q) = e^{-\alpha} \alpha^q/q!$. As noted below experimental observations allow us to place a value of approximately 64 on α. Note that the approximation of quantal amplitude is usually assumed to be normal, but here we assume a Poisson distribution of about the same shape. Indeed, the normal assumption produces biologically impossible negative values that the Poisson approximation avoids. Now note that the Q_i are independent of each other and the X_i values, so $P(X_iQ_i = q_i | X_i = 1) = P(Q_i = q_i)$, and when we sum the independent events another Poisson distribution occurs:

$$P(\sum \phi(X_i)Q_i = h | \sum \phi(X_i) = y^\phi) = \frac{e^{-\alpha y^\phi}(\alpha y^\phi)^h}{h!}. \quad \text{(B.1.5)}$$

Now following Bayes's procedure using these two approximations (B.1.1 and B.1.2):

$$P(\sum \phi(X_i)_iQ_i = h, \sum \phi(X_i) = y^\phi)$$

$$= P(\sum \phi(X_i)Q_i = h | \sum \phi(X_i) = y^\phi) \cdot P(\sum \phi(X_i) = y^\phi)$$

$$\approx \frac{e^{-(\alpha+1)y^\phi}(y^\phi)^{h + nps - 1} \alpha^h}{h! Gamma(nps)}, \quad \text{(B.1.6)}$$

where the approximation is caused by the approximation noted in B.1.1. This joint distribution is then marginated, by summing over y^ϕ via another approximation (*see B.1. footnote just below), to yield a negative binomial distribution with $h \in \{0, 1, \ldots\}$

$$P(\sum \phi(X_i)Q_i = h) = \binom{nps + h - 1}{h}\left(\frac{\alpha}{\alpha + 1}\right)^h\left(\frac{1}{\alpha + 1}\right)^{nps}. \quad \text{(B.1.7)}$$

This marginal distribution has mean = $nps \cdot (\alpha/\alpha + 1)(1/\alpha + 1)^{-1} = \alpha nps$, and variance = $mean \cdot (1/\alpha + 1)^{-1} = \alpha(\alpha + 1)nps$.

Because the negative binomial is nearly a Gaussian under parameterizations of interest here, $H(\Sigma\phi(X_i)Q_i) \approx 1/2 \log 2\pi e \alpha(\alpha + 1)nps$.

Part B.1. footnote

*Approximate the summation (i.e., the margination) with the integration

$$\int_{y^\phi=0}^{\infty} \frac{e^{-\alpha y^\phi}(\alpha y^\phi)^h}{h!} \cdot \frac{e^{-y^\phi}(y^\phi)^{nps-1}}{Gamma(nps)} dy^\phi = \int_{y^\phi=0}^{\infty} \frac{e^{-(\alpha+1)y^\phi}\alpha^h(y^\phi)^{h+nps-1}}{h!Gamma(nps)} dy^\phi$$

$$= \frac{\alpha^h}{h!Gamma(nps)(\alpha+1)^{h+nps-1}} \int e^{-(\alpha+1)y^\phi}((\alpha+1)$$

$$\cdot (y^\phi))^{h+nps-1} dy^\phi.$$

Now let $t = (\alpha + 1)y^\phi$, then $dt/dy^\phi = \alpha + 1$ and $dy^\phi = dt/\alpha + 1$ and substitute to get a recognizable integral:

$$= \frac{\alpha^h}{h!Gamma(nps)(\alpha+1)^{h+nps}} \int_{t=0}^{\infty} e^{-t}t^{h+nps-1}dt$$

$$= \frac{1}{h!Gamma(nps)} \cdot \left(\frac{1}{\alpha+1}\right)^{nps} \cdot \left(\frac{\alpha}{\alpha+1}\right)^h Gamma(h+nps)$$

Part B.2

$$H(\sum \phi(X_i)Q_i|X)$$

To find $H(\Sigma \phi(X_i)Q_i|X)$ we need $P(\Sigma \phi(X_i)Q_i|X = x)$.

Because the Q_i are independent of the particular i that generate them and only depend on how many synapses transmitted,

$$P\left(\sum \phi(X_i)Q_i = h \middle| X = x\right) = \sum_{y^\phi}$$

$$P\left(\sum_{j=1}^{y^\phi} Q_j = h, \sum_{x_i\in x} \phi(X_i) = y^\phi \middle| X = x\right). \quad (B.2.1)$$

Via the earlier description of the signal flow of computation, we can use the Markov property of conditional probabilities to write:

$$P\left(\sum_{j=1}^{y^\phi} Q_j = h, \sum \phi(X_i) = y^\phi \middle| X = x\right) = \sum_{y^\phi}$$

$$P\left(\sum_{j=1}^{y^\phi} Q_j = h, \sum \phi(X_i) = y^\phi \middle| X = x, \sum_{x_i\in x} X_i = y\right)$$

$$= \sum_{y^\phi} P\left(\sum_{j=1}^{y^\phi} Q_j = h, \sum \phi(X_i) = y^\phi \middle| \sum_{x_i\in x} X_i = y\right)$$

$$= \sum_{y^\phi} P\left(\sum_{j=1}^{y^\phi} Q_j = h \middle| \sum_{x_i\in x} X_i = y, \sum_{x_i\in x} \phi(X_i) = y^\phi\right)$$

$$P\left(\sum_{x_i\in x} \phi(X_i) = y^\phi \middle| \sum_{x_i\in x} X_i = y\right)$$

$$= \sum_{y^\phi} P\left(\sum_{j=1}^{y^\phi} Q_j = h \middle| \sum_{x_i\in x} \phi(X_i) = y^\phi\right) P\left(\sum_{x_i\in x} \phi(X_i) = y^\phi \middle| \sum_{x_i\in x} X_i = y\right),$$

where the last two steps follow by the definition of conditional probability, multiplying and dividing by the joint probability and then using Markov's idea again.

Now apply to this last result the quantal size approximation already introduced, approximate summation by integration, and replace a binomial distribution by a gamma distribution where they both are nearly approximating the same normal distribution.

$$P\left(\sum \phi(X_i)Q_i = h|X = x\right) \approx \sum_{y^\phi=0}^{\infty} \frac{e^{-\alpha y^\phi}(\alpha y^\phi)^h}{h!}$$

$$\left(\sum_{y^\phi} x_i\right) s^{y^\phi}(1-s)^{-y^\phi+\sum x_i}$$

$$\approx \int_{y^\phi=0}^{\infty} \frac{e^{-\alpha y^\phi}(\alpha y^\phi)^h}{h!} \frac{e^{-(y^\phi)(1-s)^{-1}} y^{\phi-1+s(1-s)^{-1}\sum x_i}}{Gamma(s(1-s)^{-1}\sum x_i)} dy^\phi$$

$$= \frac{\alpha^h\left(\alpha+\dfrac{1}{1-s}\right)^{-h+1-(1-s)s\sum x_i}}{Gamma((1-s)s\sum x_i)h!} \int_{y^\phi} e^{-(y^\phi)(\alpha+(1-s)^{-1})}$$

$$\cdot \left(\left(\alpha+\frac{1}{1-s}\right)y^\phi\right)^{h-1+(1-s)s\sum x_i} dy^\phi,$$

where the approximate equality arise just as in the previous subsection. However, in contrast to the calculation of $P(\Sigma \phi(X_i)Q_i)$, for the case of replacing the binomial distribution, here we must accommodate the mean and variance being different. Thus, a more complicated gamma distribution is used. Still, the integration is made easy by a change of variable.

$$t = \left(\left(\alpha+\frac{1}{1-s}\right)y^\phi\right), \text{ or}$$

$$\frac{dt}{dy^\phi} = \alpha + \frac{1}{1-s}, \text{ or } dy^\phi = \left(\alpha+\frac{1}{1-s}\right)^{-1} dt,$$

which substitutes to give:

$$P(\sum \phi(X_i)Q_i|X = x)$$

$$\approx \frac{\alpha^h\left(\alpha+\dfrac{1}{1-s}\right)^{-h-(1-s)s\sum x_i}}{Gamma((1-s)s\sum x_i)h!} \int_{t=0}^{\infty} e^{-t}t^{h-1+(1-s)s\sum x_i} dt$$

$$= \frac{\alpha^h\left(\dfrac{(\alpha+1-\alpha s)}{1-s}\right)^{-h-(1-s)s\sum x_i}}{Gamma((1-s)s\sum x_i)h!} Gamma(h+(1-s)s\sum x_i)$$

$$= \left(\frac{\alpha(1-s)}{(\alpha+1)-\alpha s}\right)^h \left(\frac{1-s}{(\alpha+1)-\alpha s}\right)^{(1-s)s\sum x_i}$$

$$\frac{Gamma(h+(1-s)s\sum x_i)}{Gamma((1-s)s\sum x_i)h!}$$

Once again, the resulting distribution, $P(\Sigma \, \phi(X_i)Q_i | X = x)$, is a negative binomial with mean:

$$\frac{(1-s)\cdot s \, \sum x_i \, \dfrac{\alpha(1-s)}{(\alpha+1)-\alpha s}}{\left(\dfrac{(1-s)}{(\alpha+1)-\alpha s}\right)} = \alpha(1-s)s \, \sum x_i$$

and variance $= \text{mean} \cdot \left(\dfrac{1-s}{(\alpha+1)-\alpha s}\right)^{-1} = \alpha \cdot (\alpha+1) \cdot$

$$\left(1 - \frac{\alpha s}{(\alpha+1)}\right)s \, \sum x_i.$$

Part B.3.

Now calculate the mutual information with normal approximations for each of the two negative binomials.

$$H\left(\sum \phi(X_i)Q_i\right) - H\left(\sum \phi(X_i)Q_i \,\Big|\, X\right)$$

$$= \frac{1}{2} \log 2\pi e(\alpha+1)\alpha n p s - \sum_{X=x} P(X$$

$$= x) \frac{1}{2} \log 2\pi e(\alpha+1)\alpha\left(1 - \frac{\alpha s}{(\alpha+1)}\right)s \, \sum x_i$$

$$= \frac{1}{2} \log np - \frac{1}{2} E_x\left[\log \sum X_i\right] - \frac{1}{2}\log\left(1 - \frac{\alpha s}{(\alpha+1)}\right).$$

Because the first two terms combine to approximately zero and when

$$\frac{\alpha s}{(\alpha+1)} \approx s,$$

then

$$I\left(X; \sum \phi(X_i)Q_i\right) \approx -\frac{1}{2}\log(1-s) = -\frac{1}{2}\log(f).$$

But $I(X; \Sigma \, \phi(X_i)Q)$ is set to $H(p^*)$ if we obey Theorem G. Then $H(p^*) = -1/2 \log(f)$, implying $f = 2^{-2H(p^*)}$.

Two comments are in order.

The approximate $E[\log\Sigma X_i/np] \approx 0$ is better as np gets larger but is good enough down to $np = 4$ (Fig. 3). Second, the variance of quantal amplitudes expresses itself in the term $\alpha s/(\alpha+1)$ as opposed to just s when there is no quantal variance. That is, if we leave out quantal amplitude variations, we get nearly the same answer without even using these approximations.

The particular parameterization of α is arrived at from experimental data. When k equals one, we have the distribution of quantal amplitudes for a single quantum. From published reports (Katz, 1966 their Fig. 30), we estimate that at 1 SD from the mean, the value of the quantal amplitude changes ~12.5%. Thus, a Poisson distribution, with a nearly normal distribution of the same shape, has a mean amplitude proportional to $\alpha = 64$ and an amplitude proportional to 72 at 1 SD from the mean. That is, 12.5% to either side of the mean value is ~1 SD of quantal size given the mean has a value of α. With $\alpha = 64$, $\alpha/\alpha+1 = 64/65$, which is reasonably close to one.

Part C

The following simple proof was suggested by Read Montague.

Let Y_1, Y_2, and Y_3 be random variables formed from the sums:

$$Y_1 = \sum_i X_i,$$

$$Y_2 = \sum_i \phi\,(X_i),$$

$$Y_3 = \sum_i Q_i\phi\,(X_i),$$

where each X_i represents the activity of an input signal and is a binary random variable with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p = q$. The function $\phi(X_i)$ is associated with quantal failures such that $P(\phi(X_i) = 1 | X_i = 1) = s$, $P(\phi(X_i) = 0 | X_i = 1) = 1 - s = f$, $P(\phi(X_i) = 0 | X_i = 0) = 1$, and $P(\phi(X_i) = 1 | X_i = 0) = 0$.

We will assume that the quantal amplitudes are Gaussian distributed with mean μ and variance σ^2 and let such a Gaussian distribution be denoted by $P(Q_i) = N(\mu, \sigma^2)dQ_i$. Furthermore, we will assume that the number of inputs (i.e., the number of terms in the sums) is large enough such that $P(Y_1) = N(np, npq)dY_1$ and $P(Y_2) = N(nps, nps)dY_2$.

We seek to determine the mutual information between Y_3 and Y_1:

$$I(Y_3; Y_1) = \sum_{Y_1, Y_3} P(Y_1) P(Y_3|Y_1) \log_2\left[\frac{P(Y_3|Y_1)}{P(Y_3)}\right].$$

To determine $I(Y_3; Y_1)$, we compute $P(Y_3|Y_1)$ as:

$$P(Y_3|Y_1) = \int_{Y_2} P(Y_3, Y_2|Y_1)dY_2$$

$$= \int_{Y_2} P(Y_3|Y_2, Y_1) P(Y_2|Y_1)dY_2$$

$$= \int_{Y_2} P(Y_3|Y_2) P(Y_2|Y_1)dY_2$$

$$= \int_{Y_2} N(\mu Y_2, \sigma^2 Y_2)dY_3 N(sY_1, sfY_1)dY_2$$

$$= N(s\mu Y_1, nps(\sigma^2 + f\mu^2))dY_3,$$

where we have used the approximation $N(\mu Y_2, \sigma^2 Y_2) \approx N(\mu Y_2, nps\sigma^2)$.

Then we compute $P(Y_3)$ as

$$P(Y_3) = \int_{Y_1} N(\mu s Y_1, nps(\sigma^2 + f\mu^2))dY_3 N(np, npq)dY_1$$

$$= N(nps\mu, nps(\sigma^2 + f\mu^2 + qs\mu^2))dY_3,$$

which yields

$$I(Y_3;Y_1) = \frac{1}{2}\log_2\left(\frac{\sigma^2 + f\mu^2 + qs\mu^2}{\sigma^2 + f\mu^2}\right).$$

If we let $\sigma^2 = \mu, p \ll 1$ $(q \sim 1)$, and $\mu \gg 1$, then we again obtain the result of Appendix Part B.

REFERENCES

Abbott L, Sejnowski TJ (1999) Neural codes and distributed representations: foundations of neural computation. Cambridge, MA: MIT.

Abshire P, Andreou AG (2001) Capacity and energy cost of information in biological and silicon photoreceptors. Proc IEEE 89:1052–1064.

Andrásfalvy BK, Magee JC (2001) Distance-dependent increase in AMPA receptor number in the dendrites of adult hippocampal CA1 pyramidal neurons. J Neurosci 21:9151–9159.

Andreou AG (1999) Energy and information processing in biological and silicon sensory systems. In: Proceedings of the Seventh International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems. Los Alamitos, CA, April.

Attwell D, Laughlin SB (2001) An energy budget for signalling in the grey matter of the brain. J Cereb Blood Flow Metab 21:1133–1145.

Balasubramanian V, Kimber D, Berry MJ II (2001) Metabolically efficient information processing. Neural Comput 13:799–815.

Barlow HB (1959) Symposium on the mechanization of thought processes, No. 10, pp 535–559. London: H. M. Stationary.

Bellingham MC, Lim R, Walmsley B (1998) Developmental changes in EPSC quantal size and quantal content at a central glutamatergic synapse in rat. J Physiol (Lond) 511:861–869.

Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. Science 252:1854–1857.

Bremermann HJ (1982) Minimum energy requirement of information transfer and computing. Intl J Theor Physics 21:203–217.

Cook EP, Johnston D (1997) Active dendrites reduce location-dependent variability of synaptic input trains. J Neurophysiol 78:2116–2128.

Cook EP, Johnston D (1999) Voltage-dependent properties of dendrites that eliminate location-dependent variability of synaptic input. J Neurophysiol 81:535–543.

Cox CL, Denk W, Tank DW, Svoboda K (2000) Action potentials reliably invade axonal arbors of rat neocortical neurons. Proc Natl Acad Sci USA 97:9724–9728.

Dayan P, Abbott LF (2001) Theoretical neuroscience. Cambridge, MA: MIT.

Destexhe A, Paré D (1999) Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J Neurophysiol 81:1531–1547.

Frank O, Öhrvik J (1994) Entropy of sums of random digits. Comp Stat Data ANal 17:177–194.

Goldfinger MD (2000) Computation of high safety factor impulse propagation at axonal branch points. NeuroReport 11:449–456.

Katz B (1966) Nerve, muscle, and synapse. New York: McGraw-Hill.

Laughlin SB, de Ruyter van Steveninck RR, Anderson JC (1998) The metabolic cost of neural information. Nat Neurosci 1:36–41.

Leff HS, Rex AF (1990) Maxwell's demon entropy, information, computing. Princeton, NJ: Princeton UP.

Levy WB, Baxter RA (1996) Energy-efficient neural codes. Neural Comput 8;531–543.

Mackenzie PJ, Murphy TH (1998) High safety factor for action potential conduction along axons but not dendrites of cultured hippocampal and cortical neurons. J Neurophysiol 80:2089–2101.

Magee JC (2000) Dendritic integration of excitatory synaptic input. Nat Rev Neurosci 1:181–190.

Magee JC (1999) Dendritic I_h normalizes temporal summation in hippocampal CA1 neurons. Nat Neurosci 2:508–514.

Mitchison G (1991) Neuronal branching patterns and the economy of cortical wiring. Proc R Soc Lond Biol Sci 245:151–158.

Paulsen O, Heggelund P (1994) The quantal size at retinogeniculate synapses determined from spontaneous and evoked EPSCs in guinea-pig thalamic slices. J Physiol (Lond) 480:505–511.

Paulsen O, Heggelund P (1996) Quantal properties of spontaneous EPSCs in neurones of the guinea-pig dorsal lateral geniculate nucleus. J Physiol (Lond) 496:759–772.

Poolos NP, Johnston D (1999) Calcium-activated potassium conductances contribute to action potential repolarization at the soma but not the dendrites of hippocampal CA1 pyramidal neurons. J Neurosci 19:5205–5212.

Rieke F, Warland D, Bialek W (1999) Spikes: exploring the neural code. Cambridge, MA: MIT.

Ringo JL (1991) Neuronal interconnection as a function of brain size. Brain Behav Evol 38:1–6.

Schreiber S, Machens CK, Herz AVM, Laughlin SB (2001) Energy-efficient coding with discrete stochastic events. Neural Comput, in press.

Shannon CE (1948) A mathematical theory of communication. Bell System Tech J 27:379–423.

Sokoloff L (1989) Circulation and energy metabolism of the brain. In: Basic neurochemistry: molecular, cellular, and medical aspects (Siegel GJ, Agranoff BW, Albers RW, Molinoff PB, eds) Ed 4, pp 565–590. New York: Raven.

Stevens CF, Wang Y (1994) Changes in reliability of synaptic function as a mechanism for plasticity. Nature 371:704–707.

Theunissen F, Miller JP (1997) Effects of adaptation on neural coding by primary sensory interneurons in the cricket cercal system. J Neurophysiol 77:207–20.

Thomson A (2000) Facilitation, augmentation and potentiation at central synapses. Trends Neurosci 23:305–312.

Tovee MJ, Rolls ET (1995) Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. Vis Cognit 2:35–58.

Treves A, Rolls E (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus 2:189–200.

Victor JD (2000) Asymptotic bias in information estimates and the exponential (Bell) polynomials. Neural Comput 12:2797–2804.

Weibel ER, Taylor CR, Bolis L (1998) Principles of animal design: the optimization and symmorphosis debate. Cambridge, UK: Cambridge UP.