Energy-Efficient Interspike Interval Codes

Patrick Crotty and William B Levy

University of Virginia Health System

P.O. Box 800420, Department of Neurosurgery

Charlottesville, VA 22908

Abstract

We investigate the energy efficiency of interspike interval (ISI) neural codes. Using the hypothesis that nature maximizes the energy efficiency of information processing, it is possible to derive neuronal firing frequencies which maximize the information/energy ratio. With simple assumptions about the encoded ISI and noise distributions, we show that ISI codes can be at least as efficient as discrete binary and frequency codes and that their predicted optimal frequencies are in the same range.

1. Introduction

In 1996, Levy and Baxter [4] proposed that information processing in the brain is optimized with respect to energy efficiency rather than representational capacity. That is, evolution has maximized not the amount of information processed by neurons, but the amount of information scaled by the energy required to transmit it. Neurons convey information by means of action potentials, though the way in which the information is encoded remains unknown. For any given code, however, the formalisms of information theory can be used to quantify the

amount of information being transmitted, thereby making it possible to determine whether the information processing is at an energy-efficient optimum.

Assuming binary and frequency codes, Levy and Baxter were able to derive optimal mean firing frequencies which maximize the information/energy ratio. These frequencies are considerably lower than those which, in their model, maximize information capacity alone. But mean neuronal firing rates as observed in the rat hippocampus and subiculum are around 10 Hz or lower [7, 6, 3], in the lower range of Levy and Baxter's results.

In this study, we determine the optimal frequencies for neurons using interspike interval (ISI) codes, in which the information is encoded as the time between successive spikes. We compare the frequency and energy-efficiency results to those obtained by using binary and frequency codes. ISI codes have also been studied by Goldberg et al. [2], who derived an upper limit on the energy efficiency by assuming an entropy-maximizing distribution for the observed ISIs. In contrast, we begin with a distribution for the generated ISIs, and obtain the distribution for the observed ISIs by convolving this distribution with that for the noise.

2. Energy efficiency of interspike interval codes

Interspike interval codes, in contrast to those studied by Levy and Baxter, are continuous, and therefore the versions of Shannon's mathematical formalisms [5] for continuous probability distributions must be used to quantify the information transmitted. We review those briefly here, following (and using the notation) of Cover and Thomas [1].

Our setup consists of a single neuron generating action potentials, with the axon connected to something which measures the time between each pair of action potentials. We do not concern ourselves here with the physiological mechanism behind this "clock." We assume only that the detected ISI length X consists of the true ISI, T, plus a noise term Z (called the "jitter" noise):

$$X = T + Z \tag{1}$$

We assume that T is bounded from below by a biophysical minimum ISI t_0 (which we will take as 2.5 ms). We also restrict the noise to being additive, $Z \geq 0$, so it produces only delays.

The amount of information transmitted when the receiver measures not the original signal but some transformation of it, as in our case, is given by the mutual information. The mutual information per spike is given by

$$I^* = h(X) - h(Z \mid T) \tag{2}$$

where the differential entropies are

$$h(X) \equiv -\int_{x_0}^{x_f} dx\, \rho_X(x) \log_2(\rho_X(x))$$

$$h(Z \mid T) \equiv -\int_{t_0}^{t_f} dt\, \rho_T(t) \int_{z_0(t)}^{z_f(t)} dz\, \rho_{Z\mid T}(z \mid t) \log_2(\rho_{Z\mid T}(z \mid t)) \tag{3}$$

and where $\rho_T(t)$, $\rho_X(x)$, and $\rho_{Z\mid T}(z \mid t)$ are the marginal probability densities for T and X and the conditional probability density for Z given $T = t$, respectively. The marginal distribution for X may be derived by convolving ρ_T with $\rho_{Z\mid T}$ over the appropriate range of T. These integrals are defined only over the support sets of the distributions, i.e., where they are non-zero. (When T is not bounded above, we take $t_f \to \infty$.) The corresponding bounds on X are x_0 and x_f.

The marginal p.d.f. for X may be written as

$$\rho_X(x) = \int_{t_0}^{x} dt\, \rho_T(t) \rho_{Z\mid T}(x - t \mid t) \tag{4}$$

where $x \geq t_0$. If Z and T are independent, we have $h(Z \mid T) = h(Z)$, so $I^* = h(X) - h(Z)$. It is important to note that although the differential entropies can be negative (in contrast to the entropies of discrete probability distributions), this is never true of the mutual information.

The mutual information per second, I, is related to I^* by $I = I^*/\langle T \rangle$, where $\langle T \rangle$ is the mean ISI. In order to determine energy efficiency, we must also quantify the mean amount of metabolic energy used per second. Most of the metabolic energy consumed by biological neurons is used by the Na^+/K^+ pump in the membrane to maintain or restore ionic concentrations.

We define c_0 as the amount of metabolic energy used to maintain the neuron over a time interval t_0 (the minimum ISI) when it is at rest. We then define r to be the ratio such that $c_0 r$ is the amount of energy used during an interval of length t_0 containing an action potential. We neglect energy costs associated with the noise. On average, a time interval $\langle T \rangle$ contains $\langle T \rangle / t_0 - 1$ intervals of length t_0 when the neuron is at rest, each of which is associated with an energy cost c_0, plus one containing a spike and associated with energy cost $c_0 r$. The mean amount of energy used per second is obtained by dividing this sum by $\langle T \rangle$; i.e.,

$$\frac{dE}{dt} = \frac{c_0}{\langle T \rangle} \left(\frac{\langle T \rangle - t_0}{t_0} + r \right) \tag{5}$$

We note in passing that r is implicitly a function of the minimum ISI t_0. Levy and Baxter assumed that r ranged from 10 to 200; we set $r = 100$.

The energy efficiency hypothesis states that the ratio

$$\frac{I}{E} = \frac{I *}{c_0 \left(\langle T \rangle / t_0 - 1 + r \right)} \tag{6}$$

is maximized in nature. In our calculations below, we neglect the numerical factor c_0. Although c_0 would have to be known in order to obtain numerical values, we can still compare the relative efficiencies of different codes if c_0 is defined the same way and assumed to have the same value.

3. Exponentially distributed signal and noise

We consider exponential, and independent, distributions for T and Z:

$$\rho_T(t) = \Theta(t - t_0) \mu^{-1} e^{-(t - t_0)/\mu} \tag{7}$$

$$\rho_Z(z) = \Theta(z) \nu^{-1} e^{-z/\nu} \tag{8}$$

where we use the Heaviside function $\Theta(y)$, defined as

$$\Theta(y) = \begin{cases} 1 & y \geq 0 \\ 0 & y < 0 \end{cases} \tag{9}$$

to indicate that t and z are restricted to the ranges $t \geq t_0$ and $z \geq 0$. The distribution of $X = T + Z$ is

$$\rho_X(x) = \Theta(x - t_0)(\mu - v)^{-1}\left(e^{-(x-t_0)/\mu} - e^{-(x-t_0)/v}\right) \quad (\mu \neq v)$$
$$= \Theta(x - t_0)\frac{(x - t_0)}{\mu^2}e^{-(x-t_0)/\mu} \quad (\mu = v) \tag{10}$$

where $x \geq t_0$. We plot these distributions for $\mu = v = 1$ ms in Figure 1. The mutual information per spike is

$$I^* = h(X) - h(Z) = -\int_0^\infty ds (\mu - v)^{-1}\left(e^{-s/\mu} - e^{-s/v}\right)\log_2\left[(\mu - v)^{-1}\left(e^{-s/\mu} - e^{-s/v}\right)\right]$$
$$- \log_2(v e) \tag{11}$$

with $s = x - t_0$ and $\mu \neq v$. The integral does not have a simple closed-form solution (except with the X distribution for $\mu = v$) and must be calculated numerically. We note that for $v \to \infty$, i.e., when the signal is completely overwhelmed by the noise, $I^* \to 0$ as one would intuitively expect. Conversely, when $v \to 0$, I^* increases without bound. This reflects the fact that a noiseless continuously valued signal could perfectly transmit arbitrary real numbers, and therefore an infinite number of bits.

The optimum mean firing frequencies increase as noise decreases. Setting the noise distribution parameter v (which is also the mean noise) to 0.1, 1, and 10 ms yields optima of, respectively, 18.2, 13.0, and 7.5 Hz. Comparing $v = 0.1$ ms to $v = 1$ ms, the smaller noise increases the energy efficiency by $0.02 - 0.025$ bits/c_0. For $v = 10$ ms versus $v = 1$ms, the opposite is true: the much larger noise decreases the energy efficiency by about $0.015 - 0.025$ bits/c_0. Of course, we can make the information rate as large as we like by making v arbitrarily small, so it is important to know what range of this parameter is biologically realistic.

In Figure 2, we plot I/E for these three choices of the noise distribution parameter. Curves (a), (b), and (c) correspond, respectively, to $v = 0.1$, 1, and 10 ms. The two thick curves are for binary and frequency codes (see below). We set $t_0 = 2.5$ ms, which establishes a maximum frequency of 400 Hz. The spike energy consumption parameter is fixed at $r = 100$; the

curves for the other r values show a similar relationship with respect to v. The I/E ratio is plotted as a function of mean firing frequency, which is related to the ISI distribution parameter μ as $1/\langle T \rangle = (\mu + t_0)^{-1}$. Since we are neglecting c_0 in the numerical calculation, the units of the ordinate are bits/c_0.

In Figure 3, we show the mutual information per second I alone. Curves (a), (b), and (c) again correspond to the exponential distributions. The energy-efficient optima are all under 20 Hz, which is far below the frequencies (and produces far smaller bit rates) than those which maximize I by itself.

4. Comparison with noiseless binary and frequency codes

For a $\{0,1\}^1$ binary code, information is conveyed by a spike or non-spike over a minimum ISI, t_0. Since t_0 is the minimum interval between spikes, it is not possible to transmit any information over a shorter time interval than t_0, and thus using t_0 as the time step gives the maximum information rate for this code. The probability p of spiking during this time is considered to be independent from one interval to the next, and the time between successive firings is not considered. The information/energy ratio can be shown to be

$$\frac{I(p)}{E} = \frac{H(p)}{c_0[1 + p(r-1)]} = \frac{-p\log_2 p - (1-p)\log_2(1-p)}{c_0[1 + p(r-1)]} \tag{12}$$

where $H(p)$ is the (discrete) Shannon entropy for a Bernoulli process. The Bernoulli parameter p is related to the mean firing frequency, f, by $f = p/t_0$, and the maximum possible frequency is $f_{max} = 1/t_0$. We assume the binary and frequency codes to be noiseless, so the mutual information equals the Shannon entropy.

In the frequency code we consider, the signal is the number of times a neuron fires over some time window $T > t_0$. The time windows are considered successively, i.e., they do not overlap. If we assume $T = N \cdot t_0$, where N is an integer greater than one, and that the probability of spiking over each smaller interval t_0 is constant, then the resulting probability distribution for

6

the frequency is binomial. The Shannon entropy can be computed numerically using this distribution. The efficiencies for binary and frequency codes are equal (as are the codes themselves) when $N = 1$, and for higher N the binary code is more efficient.

In Figure 2, we plot I/E for binary and frequency (with $N = 10$) codes as, respectively, the thick solid and thick dashed curves. We see that the binary code is just about as energy efficient as the exponential ISI code with $v = 1$ ms, though decreasing v even slightly makes the ISI code more efficient. In Figure 3, we plot I for the two discrete codes, and see once again that the binary code occupies an intermediate position with respect to the ISI codes. The question of whether or not nature uses ISI codes to maximize energy efficiency therefore depends heavily on how much noise there is.

5. Conclusions

We have shown that ISI codes with reasonable assumptions about the noise can be as or more efficient than discrete binary and frequency codes, but their relative efficiencies depend strongly on the noise levels. The predicted energy-efficient optimum firing frequencies in ISI codes are similar to those of binary and frequency codes over a large range of average noise values. We have also made fairly simple assumptions about the ISI distributions, and further research should investigate the efficiencies of more complicated distributions as well as establish more rigorous constraints on the jitter noise.

References

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory (John Wiley & Sons, New York, 1991).

[2] D. H. Goldberg, A. P. Sripati, and A. G. Andreou, Energy efficiency in a channel model for the spiking axon, Neurocomput. 52-54 (2003) 39-44.

[3] S. A. Hollup, S. Molden, J. G. Donnett, M. Moser and E. I. Moser, Accumulation of hippocampal place fields at the goal location in an annular watermaze task, J. Neurosci. 21 (2001) 1635-1644.

[4] W. B. Levy and R. A. Baxter, Energy-efficient neural codes, Neural Comput. 8 (1996) 531-543.

[5] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27 (1948) 379-423, 623-656.

[6] E. R. Wood, P. A. Dudchenko and H. Eichenbaum, The global record of memory in hippocampal neuronal activity, Nature Let. 397 (1999) 613-616.

[7] P. E. Sharp and C. Green, Spatial correlates of firing patterns of single cells in the subiculum of the freely moving rat, J. Neurosci. 14 (1994) 2339-2356.

Biographical Sketches

Patrick Crotty has a BS in physics and mathematics from the College of William and Mary and an MS and PhD in physics from the University of Chicago. He is presently a postdoctoral research fellow in the Laboratory of Systems Neurodynamics at the University of Virginia.

William B Levy earned a BA in Psychology from Princeton and a PhD in Psychobiology from the University of California Irvine. He was a Psychology professor at the University of California Riverside from 1974 until 1979 at which point he joined the faculty at the University of Virginia, where he is currently a professor in the Neurological Surgery department and in the Psychology department.

Figure Captions

Figure 1: The probability density function for an observed ISI is a function of the encoded ISI and the jitter noise. The density function for the observed ISI, *X*, is the convolution of the density function for the encoded ISI, *T*, and the density function for the noise, *Z*. In this example we set $\mu = v = 1$ ms for the distributions for *T* and *Z*. **dashed curve**: p.d.f. for noise, $\rho_Z(z)$. **thin curve**: p.d.f. for generated ISI, $\rho_T(t)$. **thick curve**: p.d.f. for observed ISI, $\rho_X(x)$.

Figure 2: An information/energy ratio curve, as a function of the mean firing frequency $1/\langle T \rangle$, has a single maximum. The maximum of each curve gives the most energy-efficient firing frequency. In all cases we set the action potential energy ratio *r* to 100 and the minimum ISI to t_0 = 2.5 ms. Curves (a)-(c) are for exponential ISI distributions, for which we set $\mu = 1$ ms, and curves (d) and (e) are for noiseless binary and frequency codes. **a**: Exponential ISI distribution and noise, noise distribution parameter $v = 0.1$ ms, maximum (i.e., optimal mean frequency) at 18.2 Hz. **b**: Exponential ISI distribution and noise, $v = 1$ ms, maximum at 13.0 Hz. **c**: Exponential ISI distribution and noise, $v = 10$ ms, maximum at 7.5 Hz. **d**: Noiseless binary code, maximum at 13.4 Hz. **e**: Noiseless frequency code, maximum at 7.5 Hz.

Figure 3: The maximum information rates are always achieved at frequencies exceeding those which maximize energy efficiency. The range of $1/\langle T \rangle$ is double that of Figures 1 and 2 so as to show the maximum of each case. The letters denote the same distributions and codes as in Figure 2.
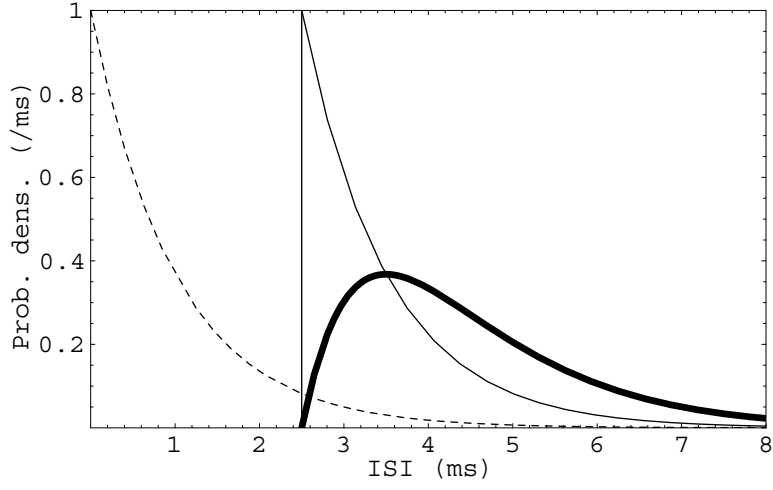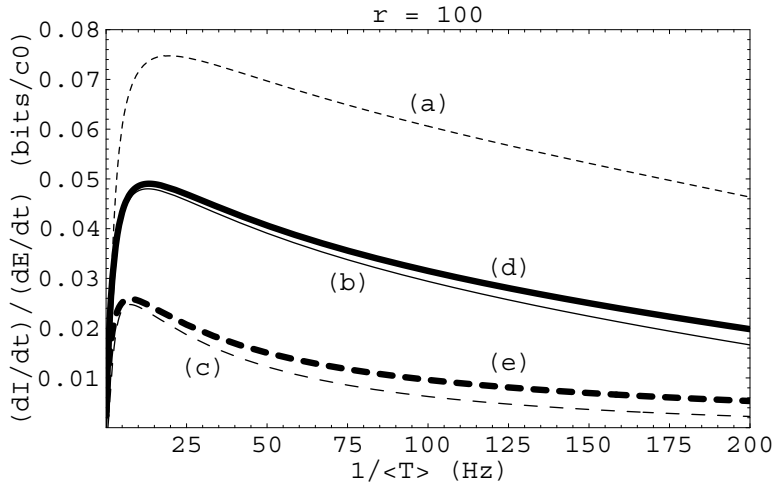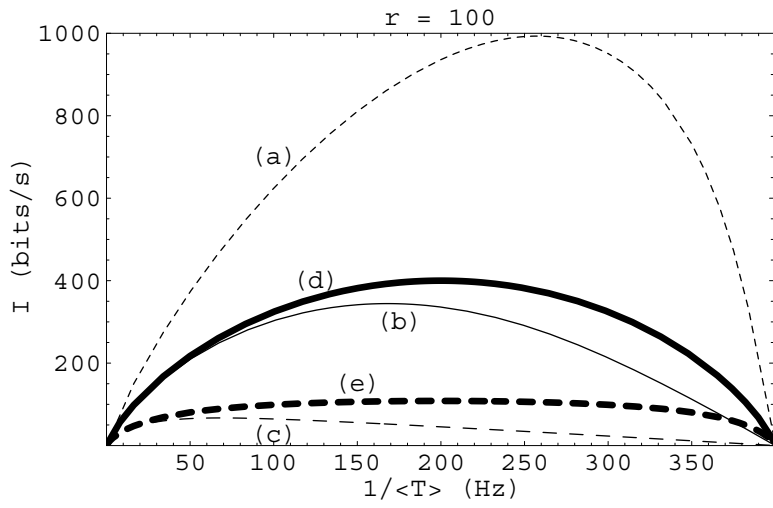
Figure 1

Figure 2

Figure 3