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

Keywords: interpulse interval, entropy, spike, axon

Abstract

We study the energy efficiency of interspike interval (ISI) neural codes. Assuming that nature maximizes the amount of information transmitted per metabolic energy cost produces a firing frequency which maximizes the energy/information ratio. Fairly simple assumptions about and parameterizations of the signal and jitter noise distributions lead to ISI codes that are at least as efficient as discrete binary and frequency codes.

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 and not false positives.

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. For signals that can vary continuously, this can be written as

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

where the differential entropies are

$$\begin{aligned} h(X) &\equiv -\int_{x_0}^{x_f} dx\, \rho_X(x) \log_2\left(\rho_X(x)\right) \\ h(Z \mid T) &\equiv -\int_{t_0}^{t_f} dt\, \rho_T(t) \int_{z_0(t)}^{z_f(t)} dz\, \rho_{Z|T}(z \mid t) \log_2\left(\rho_{Z|T}(z \mid t)\right) \end{aligned} \tag{3}$$

and where $\rho_T(t)$, $\rho_X(x)$, and $\rho_{Z|T}(z \mid t)$ are respectively the marginal probability densities for T and X and the conditional probability density for Z given $T = t$. The upper bound on T, t_f, may be infinite. The bounds on Z are considered to be functions of the value of T. The integrals in Eq. (3) are defined only over the region where the probability densities are non-zero.

The marginal p.d.f. for X is

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

where $x \geq t_0$. If Z and T are independent, we have $h(Z \mid T) = h(Z)$, so $I(T; X) = h(X) - h(Z)$. 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.

Equations (2)-(4) are the formulas we use to quantify the information transmitted with ISI codes. Although it is not explicit, the information is actually the mean information transmitted per spike. Since the mean spike rate is $1/\langle T \rangle$, the expression $\dot{I} \equiv I(T; X)/\langle T \rangle$ gives the mean information per unit time, i.e., in bits/s.

In order to determine energy efficiency, we quantify the mean amount of metabolic energy used per unit time. 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 a spike. 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.,

$$\dot{E} = c_0\left(\frac{1}{t_0}\left(1-\frac{t_0}{\langle T\rangle}\right)+\frac{r}{\langle T\rangle}\right) \tag{5a}$$

has units J/s. We can also write energy consumption as the energy used per spike:

$$E = c_0\left(\frac{\langle T\rangle}{t_0}-1+r\right) \tag{5b}$$

We note in passing that r is implicitly a function of the minimum ISI, t_0. Levy and Baxter assumed that it ranged from 10 to 200. We will assume the same range.

The energy efficiency hypothesis therefore states that the ratio

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

which has units bits/J, is at a maximum 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 with Eq. (6), 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:

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

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

where $t \geq t_0$ and $z \geq 0$. The distribution of $X = T + Z$ is

$$\begin{aligned} \rho_X(x) &= (\mu - \nu)^{-1}\left(e^{-(x-t_0)/\mu} - e^{-(x-t_0)/\nu}\right) \quad (\mu \neq \nu) \\ &= \frac{(x-t_0)}{\mu^2} e^{-(x-t_0)/\mu} \quad (\mu = \nu) \end{aligned} \tag{9}$$

where $x \geq t_0$. We plot these distributions for $\mu = \nu = 1$ ms in Figure 1. The mutual information per mean ISI, whose length is $\langle T \rangle = \mu + t_0$, is

$$\begin{aligned} I(T;X) = h(X) - h(Z) = -\int_0^\infty ds (\mu - \nu)^{-1}\left(e^{-s/\mu} - e^{-s/\nu}\right) \log_2\left[(\mu - \nu)^{-1}\left(e^{-s/\mu} - e^{-s/\nu}\right)\right] \\ - \log_2(\nu e) \end{aligned} \tag{10}$$

with $s = x - t_0$ and $\mu \neq \nu$. The integral does not have a simple closed-form solution (except, in regard to the X distribution, only when $\mu = \nu$) and must be calculated numerically.

The optimum mean firing frequencies increase as noise decreases. Setting the noise distribution parameter ν (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 $\nu = 0.1$ ms to $\nu = 1$ ms, the smaller noise increases the energy efficiency by 0.02 – 0.025 bits/c_0. For $\nu = 10$ ms versus $\nu = 1$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 ν 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), (c), and (e) correspond, respectively, to $\nu = 0.1$, 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 ν. The I/E ratio is plotted as a function of mean firing frequency, which is related to the ISI distribution parameter μ as $f =$

$(\mu + t_0)^{-1}$. 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 bit rates $\dot{I}$ alone, where we use the dot to denote that we are dividing the mutual information by $\langle T \rangle$ in order to obtain bits/sec. Curves (a), (c), and (e) 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 $\dot{I}$ by itself.

4. Comparison with noiseless binary and analog codes

In binary codes, information is conveyed by a spike or non-spike over the minimum ISI, t_0. 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

$$\frac{I(p)}{E} = \frac{H(p)}{c_0\left[1 + p(r-1)\right]} = \frac{-p\log_2 p - (1-p)\log_2(1-p)}{c_0\left[1 + p(r-1)\right]} \tag{11}$$

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

In the analog 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 the frequency is binomial. The Shannon entropy can be computed numerically using this distribution. The efficiencies for binary and analog 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 analog (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 the bit rates $\dot{I}$ 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.

4. Conclusions

We have shown that ISI codes with reasonable assumptions about the noise can be as or more efficient than discrete binary and analog codes, but their relative efficiencies depend strongly on the noise levels. 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.

Acknowledgements

This work was supported by NIH MH48161, MH63855 and RR15205 to WBL, and by the Department of Neurosurgery.

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 research associate 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: Probability density functions for ISI, noise, and observed ISI, with $\mu = \nu = 1$ ms. **dashed curve**: p.d.f. for noise, $\rho_Z(z)$. **thin curve**: p.d.f. for generated ISI, $\rho_T(t)$. **thick curve**: convolved p.d.f. for observed ISI, $\rho_X(x)$.

Figure 2: Energy efficiency curves I/E for the different distributions, all with spike energy parameter $r = 100$ and minimum ISI $t_0 = 2.5$ ms. **a**: Exponential ISI code and noise, noise distribution parameter $\nu = 0.1$ ms, optimal mean frequency $f^* = 18.2$ Hz; **b**: noiseless binary code, $f^* = 13.4$ Hz; **c**: exponential ISI code and noise, $\nu = 1$ ms, $f^* = 13.0$ Hz; **d**: noiseless frequency code, $N = 10$, $f^* = 7.5$ Hz; **e**: exponential ISI code and noise, $\nu = 10$ ms, $f^* = 7.5$ Hz.

Figure 3: Bit rate curves for the distributions, i.e., $\dot{I}$ alone is plotted. The range of f is double that in Figures 1 and 2 in order to show the maxima. The letters correspond to the same distributions as in Figure 2. Note that the frequencies at which $\dot{I}$ is at a maximum are far higher than those which maximize I/E. Conversely, the bit rates at the I/E-maximizing frequencies are far smaller than at the ones which maximize $\dot{I}$. For the f^* values given in the caption to Figure 2, these are: **a**: 165 bits/s; b: 85 bits/s; **c**: 81 bits/s; **d**: 30 bits/s; **e**: 28 bits/s.

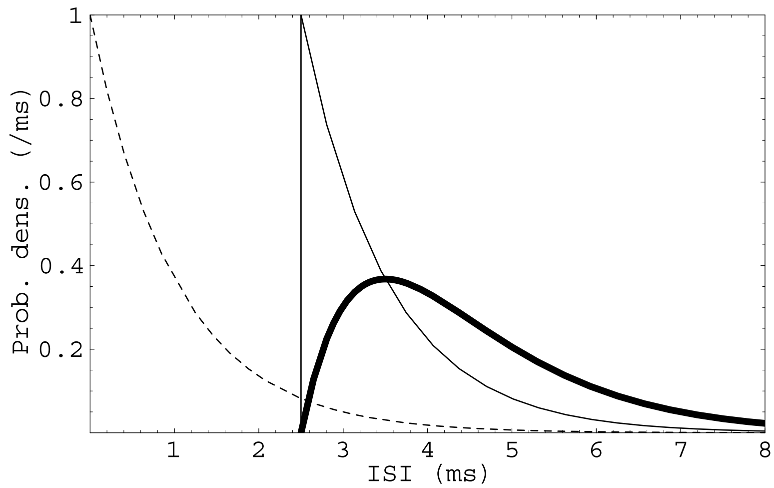

Figure 1

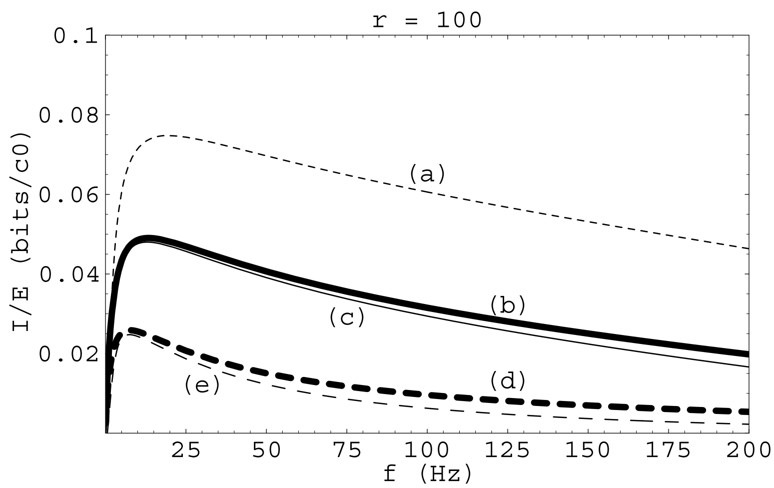

Figure 2

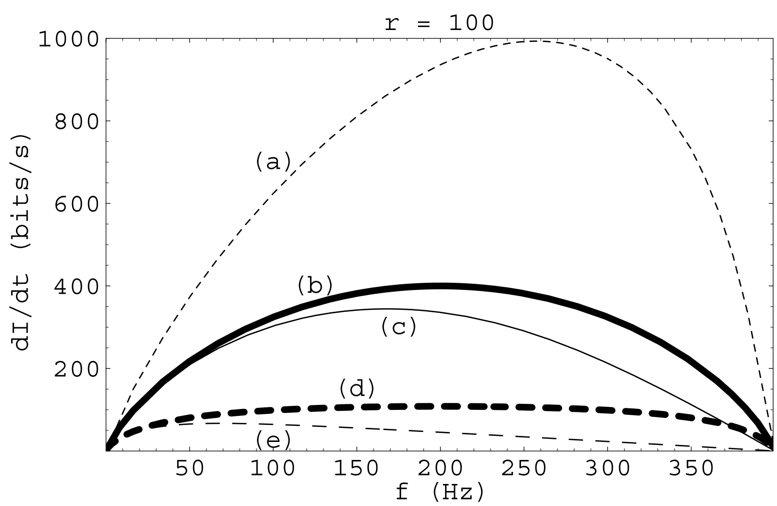

Figure 3