Fast Temporal Encoding and Decoding with Spiking Neurons

David Horn and Sharon Levanda

School of Physics and Astronomy

Raymond and Beverly Sackler Faculty of Exact Sciences

Tel Aviv University, Tel Aviv 69978, Israel

April 25, 1999

Abstract

We propose a simple theoretical structure of interacting integrate and fire neurons that can handle fast information processing, and may account for the fact that only a few neuronal spikes suffice to transmit information in the brain. Using integrate and fire neurons that are subjected to individual noise and to a common external input, we calculate their first passage time (FPT), or inter-spike interval. We suggest using a population average for evaluating the FPT that represents the desired information. Instantaneous lateral excitation among these neurons helps the analysis. By employing a second layer of neurons with variable connections to the first layer, we represent the strength of the input by the number of output neurons that fire, thus decoding the temporal information. Such a model can easily lead to a logarithmic relation as in Weber's law. The latter follows naturally from information maximization, if the input strength is statistically distributed according to an approximate inverse law.

1 Introduction

A fundamental question in Neuroscience is the understanding of the neural code (Ferster and Spruston, 1995). The standard doctrine is that information is transmitted by action potentials. Since the sizes of these signals are uniform, the question arises how do they code the information about the strength of the input that triggered them. The common belief is that the firing rate reflects this strength. However, firing rate evaluation takes a long time, whereas evidence exists that information passes quite quickly. In the hierarchical structure of the visual cortex, one can estimate (Thorpe *et al.* , 1996) that during the time that the information has passed through several synaptic connections, most of the neurons in the different layers fired only once. This can lead to the conclusion that the exact timing, or latency, of the firing is of importance. For a debate on whether latency or rate play the important role we refer to Shadlen and Newsome (1994) and Softky (1995).

In the visual system (Thorpe *et al.* , 1996) one can imagine that a single action potential carries important information because it can specify location and orientation of the stimulus. Moreover, action potentials from different neurons can specify relative strengths of the inputs according to their latency, the latter the spike the weaker is the signal. In an abstract problem we consider a neuron, or a group of neurons, subject to some input signal. To measure its strength through a single spike one needs some reference system. Hopfield (1995) suggested employing a sub-threshold periodic potential. An external constant input causes then the neuron to fire periodically with some time lag with respect to the underlying oscillation. Hence, in this model, information is carried by the relative phase of the neuron's firing with respect to the underlying periodic potential.

The fact that latency can have computational meaning is brought out clearly in the models by Maass (1997). Working with coupled integrate-and-fire ($I\&F$) neurons, and allowing the synaptic input $W_{ij}S_j$ from neuron j to neuron i to be effectively constant from the time that neuron j fired until the time neuron i fires, leads to a temporal representation of analog amplitudes. The reference frame has to be defined in some manner, e.g. by the firing pattern of one neuron.

In our study we will not employ any external reference frame. We will have, therefore, to work with two spikes per neuron, trying to obtain the information about the input signal from the inter-spike interval. We investigate systems of $I\&F$ neurons with instantaneous interactions. To imitate the biological environment we use random noise to generate a sub-threshold potential. We ask ourselves how can the timing of the neuron's firing carry the information about the input in such a system. In the following section we calculate the first passage time (FPT) for this system, i.e. the average interval between two consecutive spikes. This is the inverse of the firing rate of the neuron. It carries, therefore, the same analog information as the firing rate does. The question still remains, how can computation be carried out in view of the fluctuations that every neuron is subjected to. The answer we propose is averaging over a large number of neurons that receive the same input. This is elaborated in the third section. We look also at the effect of excitatory lateral interactions in this neuronal assembly, leading to better synchronization and lowering the FPT.

After understanding how to encode the strength of the stimulus as temporal information in the FPT, we construct a secondary neural layer that decodes this information, translating it into a number of firing neurons. Such a neural system can easily account for Weber's law that implies a logarithmic relationship between the strength of the stimulus and the reaction of the central nervous system. This follows naturally from an information maximization approach if the distribution of inputs obeys a statistical inverse law.

2 First Passage Time

Consider an $I\&F$ neuron under the influence of noisy input. Its membrane potential obeys the equation

$$\tau \frac{dV}{dt} = -V + I \tag{1}$$

where the resting potential is set to 0 and I represents the incoming current (in voltage units). When V reaches its threshold value V_{th}, a spike is emitted and V is reset to V_0. Let us now assume that the input is composed of signal and noise, $I = I_S + I_N$. The noise is assumed to be Gaussian, averaged around zero, so that, in the absence of any signal, V does not usually reach the threshold. Hence, action potentials that are caused by noise are very rare.

Once a signal is turned on, one may ask what is the time period between two consecutive action potentials. This is called the first passage time (FPT), since it is also the time it will take the neuron to fire if the membrane potential starts out at its resting value V_0. Clearly this time period is statistically distributed due to the noise. Only in the absence of noise it is deterministic. In general we ask therefore for the average value only. This value is the inverse of the firing rate of the neuron, and it depends on both the amplitude of the signal, I_S, and the characteristics of the noise, I_N.

This problem can be solved in a closed form (Ricciardi and Sacerdote, 1979). Our derivation follows Gardiner (1983). Defining a fluctuating normally distributed input for Eq. (1), with

$$E\{I(t)\} = \eta_I, \quad E\{(I(t_1) - \eta_I)(I(t_2) - \eta_I)\} = A\delta(t_1 - t_2), \tag{2}$$

one finds that V will also be normally distributed, with moments

$$\mu(t) = E\left\{V(t)\right\} = V_0 e^{-(t-t_0)/\tau} + \eta_I \left(1 - e^{-(t-t_0)/\tau}\right), \tag{3}$$

$$\sigma^2(t) = E\left\{(V(t) - \mu)^2\right\} = \frac{A}{2\tau}\left(1 - e^{-2(t-t_0)/\tau}\right). \tag{4}$$

Hence, the probability density of $V(t)$, given $V(t_0) = v_0 = V_0$ is

$$p\left(v, t \mid v_0, t_0\right) \equiv P\left(V(t) = v \mid V(t_0) = v_0\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(v-\mu)^2}{2\sigma^2}}. \tag{5}$$

The arguments of this function are limited in range. The upper bound is defined by the threshold for firing, V_{th}. We may set also a lower bound, V_{min}, that corresponds to the lowest potential value that the membrane can sustain. This probability density obeys a Fokker Planck equation, which one can use to obtain the following expression for the average FPT, T:

$$T = 2\frac{\tau^2}{A}\int_{V_0}^{V_{th}} \mathrm{d}v \cdot e^{\frac{\tau}{A}(v-\eta_I)^2} \int_{V_{min}}^{v} e^{-\frac{\tau}{A}(z-\eta_I)^2} \mathrm{d}z. \tag{6}$$

The details of this calculation are described in the Appendix. In Fig. 1 we present a comparison between the theoretical formula and numerical simulations. The average noise term vanishes, so that $\eta_I = I_S$. For large input currents, $I_S >> \sqrt{A}$, the signal term dominates the FPT, in which case one can replace it with the noiseless solution

$$T \to -\tau \log\left(1 - \frac{V_{th} - V_0}{I_S}\right). \tag{7}$$

This can be further approximated by

$$T \to \frac{\tau(V_{th} - V_0)}{I_S}, \tag{8}$$

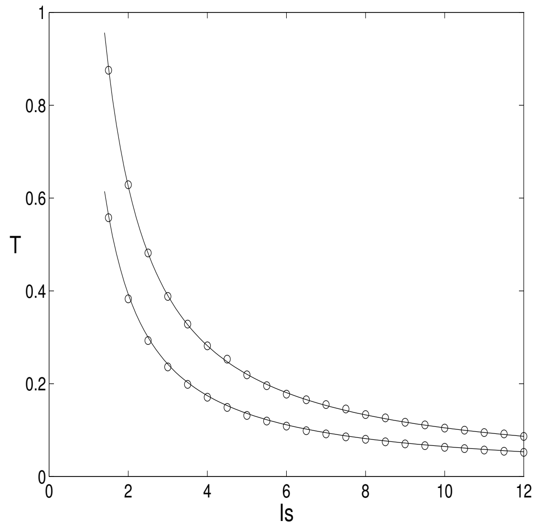

Figure 1: Comparison of the theoretical evaluation of FPT with numerical simulations as a function of input strength, I_S. The upper curve corresponds to the case of Eq. (6) with $A = 0.5$. The lower curve is a solution of Eq. (10) with $A = 0.5$ and $Nw = 0.4$. We have set $V_0 = 0 \quad V_{th} = 1 \quad V_{min} = -\infty$ and $\tau = 1$ in all figures.

for $I_S >> V_{th} - V_0$.

In this study we look at the case when the size of the stimulus is larger, or at least comparable, to that of the noise. The other extreme, when the stimulus is small compared to the noise, was recently studied by Stemmler (1996), who pointed out that a stochastic resonance behavior follows. Other works that have discussed the FPT in the context of neuronal inter spike intervals are Ricciardi and Sacerdote (1979), Tuckwell and Cope (1980), Tuckwell (1989) and Lansky *et al.* (1994).

3 Population Average in a Neural Assembly

Suppose a population of N neurons, subjected to random noise with given characteristics, starts out with random initial conditions. At a certain time a stimulus I_S is turned on, the same stimulus for all neurons. Let us denote by t_1 the time that passes from the onset of the stimulus I_S until the first firing. Can we use $< t_1 >$, the average calculated over this population, to determine I_S? Clearly not, since we still lack the information when the stimulus was turned on. If, however, we include t_2 in our computational scheme, we may do it, simply because the average value of $t_2 - t_1$ over a population of neurons is equal to the FPT

$$T =< t_2 - t_1 >\equiv \frac{1}{N}\sum_{i=1}^{N} t_2(i) - t_1(i). \tag{9}$$

Clearly this value depends on I_S. Thus, if we are interested in fast information transfer, we can resort to a population average that conveys the same information as a long temporal average needed to estimate the firing rate of any single neuron.

A simple calculational scheme for performing such averages could be to let all neurons feed some output unit that can calculate $< t_1 >$ and $< t_2 >$ separately, and train a simple network to associate the difference between the two with the relevant size of the input. The problem, however, is the separation between the probability distributions of t_1 and t_2. For weak stimuli, or high noise, some of the neurons will fire for the second time before others fired for the first time.

This problem can be alleviated by introducing synchronizing interactions between the neurons. The example that we use in our simulations is that of excitatory lateral interactions. The interactions are triggered by the occurrence of an action potential in the presynaptic neuron, and the synaptic transmission is assumed to be

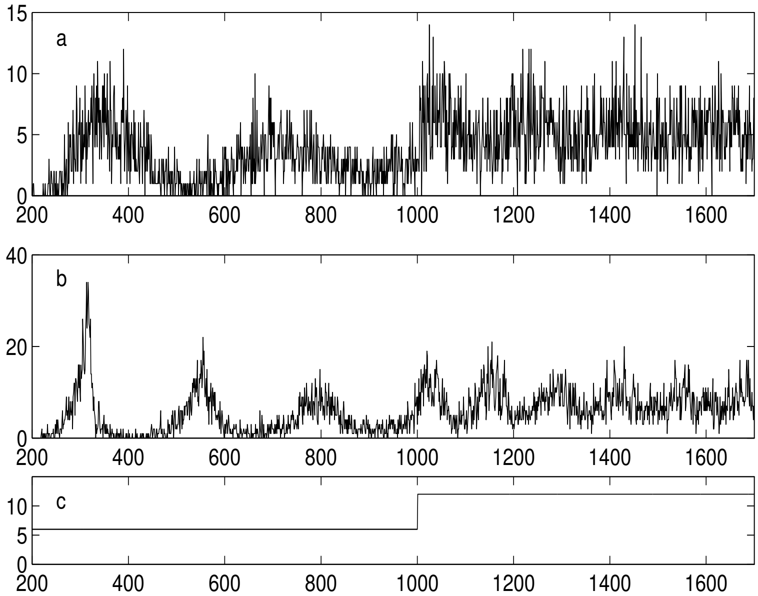

Figure 2: Firing activity of a set of $N = 1000$ integrate-and-fire neurons, as a function of time after signal onset. The input signal is $I_S = 6$ from $t = 0$ to $t = 1000$ and changes to $I_S = 12$ from $t = 1000$ onwards. The system starts out at $t = 0$ with all $V = V_0$, hence the first action potential is more synchronous than the rest. Synchrony deteriorates with time due to the random noise, whose variance is $A = 0.2$. The three frames correspond to: a. spiking activity without lateral interactions, b. spiking activity with lateral interactions of $wN = 0.33$, c. strength of the input I_S. As can be seen here, the instantaneous excitatory interactions lead to improved synchrony and shortened FPT.

instantaneous.[1] It is well known (Mirollo and Strogatz, 1990) that under these conditions neurons tend to synchronize. This effect can be seen in Fig. 2, where we compare neuronal firing activity for cases with and without interactions. The effect of the coupling is, however, twofold. In addition to improving synchrony, it shortens the resulting FPT.

The new FPT can be estimated theoretically by a mean field approximation. This approximation is valid when the accumulated effect of the noise overcomes synchronization, e.g. at the end of the time scale in Fig. 2. For simplicity let us assume that all neurons interact with all others with the same synaptic weight w. We look for a self-consistent solution of the FPT, T, in terms of an input that is changed to $I + \frac{w \cdot N}{T}$ where N is the total number of neurons :

$$T = T(I_S + \frac{wN}{T}) \tag{10}$$

The second term in the input represents the average effect of all other neurons. T remains also a function of I_N, the noise whose average vanishes. This theoretical estimate is compared in Fig. 1 with simulation results. For large inputs it corresponds to a uniform reduction of the FPT, as can be seen from the asymptotic expression that is valid in this case:

$$T \to \frac{\tau(V_{th} - V_0) - wN}{I_S} \tag{11}$$

The simulations displayed in Fig. 2 started out with $V = V_0$ at $t = 0$. As time goes on we see how the accumulated effect of the noise desynchronizes the spiking activity of the neurons, even if synchronizing interactions exist among them. The mean field approximation, used to derive Eq. 9, is valid in the limit when desynchronization takes place. However, its qualitative behaviour describes correctly even the changes

[1]Synaptic delays are known to spoil synchronization. However, it was recently found that synaptic adaptation reinstates it (van Vreeswijk and Hansel, 1997).

in the average time difference between the first few synchronized volleys of action potentials.

4 Decoding the Temporal Information

In the previous section we have discussed a way by which the intensity of the incoming stimulus can be encoded in the temporal domain. We wish to construct now a simple neuronal model that decodes this information and translates the temporal coding back into one of firing intensity of a (new) neuronal layer. This system will work as a quantizer, representing the analog input signal I_S by a number n of firing neurons. Hopfield (1995) has suggested employing delays in the decoding stage. This idea, which is backed by biological evidence, can clearly translate temporal information into a number of firing neurons by employing the latter as coincidence detectors. In the structure proposed in this section we attempt to solve the problem without employing delays, relying on appropriate distributions of synaptic weights. Clearly a general system may employ distributions of both delays and synaptic strengths in the decoding stage.

Consider the following feed-forward network of $I\&F$ neurons: The first layer is a neuronal assembly of the type described in the previous section, subjected to sub-threshold noise and acting under the influence of some external input I_S. Introduce now a second layer of N neurons that are connected by excitatory weights W_{ij} to neurons in the first layer. These weights vary within some given range. We will specify below how this variation effects the decoding task. For simplicity we assume that no synaptic delays occur. We choose the parameters of our model so that the first action potentials on the first layer are insufficient to bring the neurons on the second layer to their firing threshold. However, the second action potential of neurons on the first layer will cause some of the neurons on the second layer to fire. The number of second layer neurons that will fire after t_2, $n(I_S)$, is a function of the strength of the input. The membrane potential of neuron number i on the second layer is given by

$$V_i = W_i(e^{-T(I_S)/\tau} + 1), \tag{12}$$

where $W_i = \sum_j W_{ij}$. Clearly all neurons for which $V_i \geq 1$ will fire. This relation is valid in the limit when the signal I_S dominates the noise. We will return to the case of noisy input in section 6 below.

Let us enumerate the neurons on the second layer, $i = 1, 2, \cdots, N$, such that W_i is an increasing function of i. Each neuron can then be associated with a particular value of input $I_S = J_i$ that will be the one that is just sufficient to bring that neuron to its firing threshold. It has therefore to obey

$$1 = W_i(e^{-T(J_i)/\tau} + 1). \tag{13}$$

The values J_n can be regarded as quantizer demarcation points that discretize the input into a set of $N+1$ values, $n = 0, 1, \cdots, N$, with $J_0 = I_S^{min}$, $J_{N+1} = I_S^{max}$. The relation between W_i and J_i is depicted in Fig. 3a and 3d.

In order to determine the values of W_i we propose to use maximization of mutual information (Linsker 1988, Barlow 1989) between the input I_S and the output $n(I_S)$

$$\mathcal{I}(n; I_S) = H(n) - H(n|I_S). \tag{14}$$

This is equivalent to maximization of the output entropy $H(n)$ if $n(I_S)$ is a deterministic function. This entropy is defined as

$$H(n) = -\sum P(n)\log P(n), \tag{15}$$

where $P(n)$ is the probability of n neurons firing,

$$P(n) = \int_{J_n}^{J_{n+1}} p(I_S) dI_S. \tag{16}$$

$p(I_S)$ is the probability distribution of the input.

The output entropy is maximized by a uniform distribution,

$$P(n) = \frac{1}{N+1}. \tag{17}$$

This leads to a relation between n and J_n,

$$\frac{n+1}{N+1} = \int_{J_0}^{J_{n+1}} p(I_S) dI_S, \tag{18}$$

thus defining our (ideal) quantizer (Cover and Thomas, 1991) for the given problem . In Fig. 3a and 3b we present the quantization choices suitable for an input that follows the Gaussian distribution of Fig. 3c. We display here also the resulting relation between n and J_n, which, for high N, is well approximated by a smooth curve:

$$n(I_S) \approx N \int_{I_S^{min}}^{I_S} p(I') dI' \tag{19}$$

Such a relation between response and stimulus is a general consequence of efficient information transfer (Atick, 1992). For the Gaussian input the result is a sigmoidal relation, as shown in Fig. 3b.

This analysis provides an algebraic solution of the maximization of mutual information. Alternatively one can solve this problem iteratively by a gradient ascent method. The task is finding the quantization points J_i, or the connection strengths W_i. Starting with an initial random choice of W_i one can obtain the desired solution through the iterative process

$$\Delta W_i \sim \frac{\partial H}{\partial W_i} = \frac{\partial H}{\partial J_i} \frac{\partial J_i}{\partial W_i}. \tag{20}$$

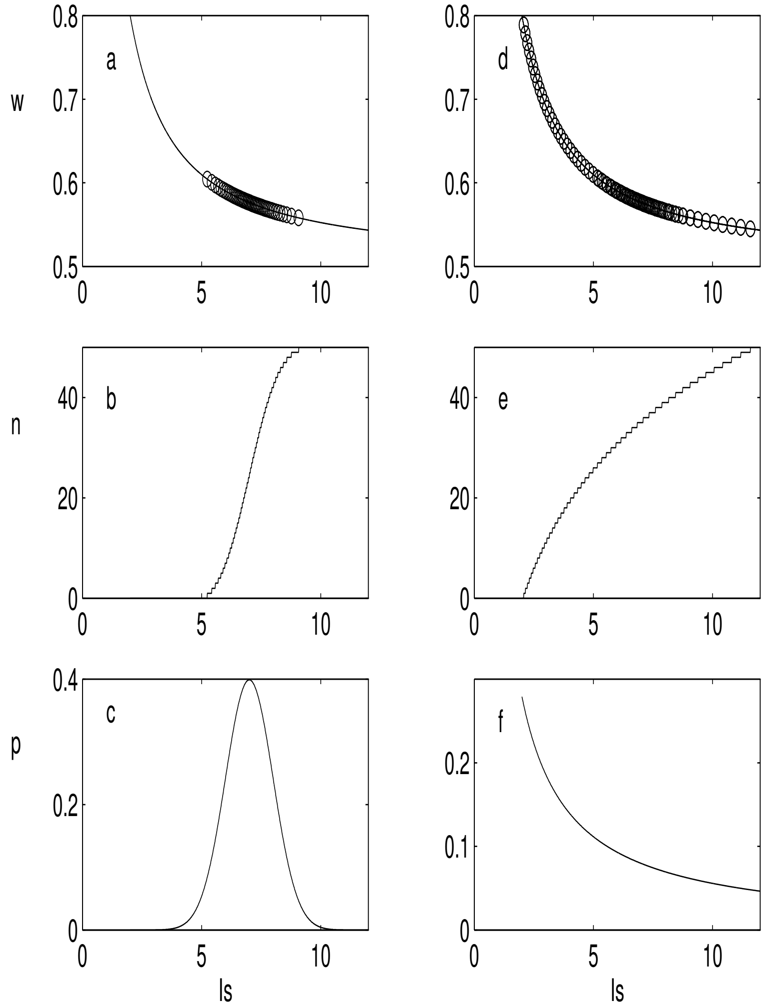

Figure 3: Decoding characteristics of a second layer of $N = 50$ neurons. Upper frames: Weights of output neurons as a function of input strength I_S. Middle frames: Number of firing neurons on the second layer as a function of input strength. Lower frames: Probability distribution of the input. (a) W_i $vs.$ J_i, and (b) n_i $vs.$ I_S are depicted for an input Gaussian distribution (c). Similarly, (d) and (e) represent the weights and the number of firing neurons for a $1/I_S$ type distribution shown in (f). (e) has the logarithmic behavior expected from Weber's law.

5 Weber's Law

Weber's law is the psychophysical observation that neuronal response is proportional to $\log(I_S)$, or

$$\Delta R = \Delta I_S / I_S, \tag{21}$$

where R denotes the response of the central nervous system. This is observed in various modalities, and analyzed best in vision (Cornsweet, 1970).

In trying to understand this general observation, one may think of it as a system's property, or look for an interpretation on the level of the single neuron. Hopfield (1995) opted for the latter. Assuming the existence of an underlying periodic potential, he aimed at obtaining delay times that are proportional to $\log(I_S)$. A different approach, still restricted to the single neuron, has been recently proposed by Tal and Schwartz (1997). They work with $I\&F$ neurons and look for a logarithmic relationship between the firing rate of the neuron and its input. The logarithm of Eq. 7 does not help, since over most of the range the approximation of Eq. 8 dominates. They show, however, that if one allows for an absolute refractory period t_0 in the range of values $0.13 < t_0/\tau < 0.23$ the neuron acts in a quasi-logarithmic fashion.

In our model it is easy to account for a logarithmic rule on the system's level, shifting the burden of proof from the first layer neuron to the analysis carried out by the second layer. To obtain $n \sim \log(I_S)$ one has to make the choice of W_i values as indicated in Fig. 3d, covering now the entire range of I_S with a varying density. This can be motivated by the information maximization principle if the distribution of the incoming intensity is $p(I_S) \sim 1/I_S$, as in Fig. 3f. It seems quite reasonable to assume that in most modalities we are subjected to input distributions that peak at low strengths. If the latter approximately resemble an inverse law distribution, the information maximization principle will lead to an approximate logarithmic relation. It should be noted that our model is a simplified construction, described here over one order of magnitude variation of I_S. The central nervous system has clearly a much more complicated architecture involved in making response decisions over ranges of many decades of input strength. Nonetheless, our model can serve as an example, implementing the principles by which this system operates.

6 Temporal Decoding in the Presence of Noise

The quantizer systems of the previous two sections were described and analyzed as coding/decoding neuronal layers in the absence of noise. Moreover, the analysis has also made the assumption that all neurons on the first layer start out with the same initial values of their membrane potential, otherwise they would not fire simultaneously. Relaxing these assumptions makes it difficult to carry out the analysis. Nonetheless, as we will show here, small changes in initial conditions, as well as low noise values, do not change the qualitative behaviour.

To understand these effects let us consider the hypothetical case that the first layer neurons fire according to a uniform distribution over a range of time Δ that is small with respect to T. This problem is still noiseless but corresponds to a (very particular) modification of the initial conditions. It leads to the following modified relation

$$1 = W_i(e^{-T(J_i)/\tau} + 1)\frac{1 - e^{-\frac{\Delta}{\tau}}}{N - Ne^{-\frac{\Delta}{N\tau}}} \tag{22}$$

which is an effective rescaling of the W_i, resulting in a shift of the curve of W $vs.$ J as shown in Fig. 4a. We consider this shift to be the important effect of any small variation due to either noise or initial conditions. Thus, in the presence of low noise we search for the best effective Δ that will give a W $vs.$ J curve that leads to a match between $P(n)$ and $p(I_S)$. This suffices to provide an adequate solution to the problem, as shown in Fig. 4b.

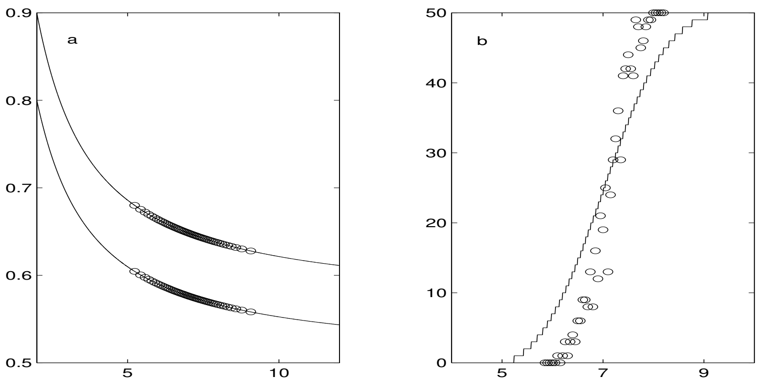

Figure 4: Left: Weights of output neurons as a function of input strength I_S for Δ =0 (upper curve) and .12τ (lower curve). Right: Number of firing neurons on the second layer for the Gaussian distribution of Fig. 3c with (circles) and without (line) noise (A=0.1). The noisy case was processed with $\Delta = 1.1\tau$.

7 Discussion

The dynamics of encoding a signal in a population of neurons is a subject of interest in the Neuroscience community for a long time. In his seminal paper, Knight (1972) has introduced the integrate and fire neuron as a tool to study such questions, and has shown its response to oscillatory input with and without noise. The $I\&F$ neuron has since become a standard simplified version of a biological neuron, and can be used to confront the problem that interests us in this paper: How can a neural system encode and decode the strength of an input signal in a very short time? We suggest a model system composed of two layers of neurons. The first one encodes the amplitude of the stimulus by a time interval, and the second one translates it into a number of firing neurons. The operation of the second layer should terminate after its first volley of action potentials. This can be brought about by a self inhibition mechanism. It is preferrable if this inhibiton acts also on the first layer in order to bring all its neurons to the baseline value before a new stimulus arrives.

One may wonder why one needs the two layers at all. Can't we connect the second layer directly to the stimulus, arranging some clever choice of weights that will lead to the same result? The problem that one encounters in such an architecture is lack of uniformity. It is difficult to construct a solution that leads to a coherent peak of activity. Allowing for neurons that react with different sensitivity, e.g. by factors of 2 and more, we will end up with outputs at different times, not to mention the problem of different numbers of spikes in the same time span. The two layer structure allows us to introduce coherence in the first layer, leading to a managable computation.

In the absence of noise, one can encode the input I_S by the inter spike interval of a single neuron. Assuming noise has to be taken into account, we have considered the case of many neurons on the first layer, each subjected to individual noise drawn from a common distribution. Clearly the average of this population serves the purpose of defining, with some accuracy, the desired time interval, or FPT. This approach replaces the insufficient time span, required for neural measurement of firing rate, by a population average of inter spike intervals, which is the inverse of the instantaneous firing rate. Thus, although temporal acuity is important in our model, it still carries information that is equivalent to rate.

The decoding stage needs many neurons for fine quantization even if one uses a single neuorn for encoding (which is possible in the noiseless case). The number of active neurons, n, is a function of T and, as such, of the input signal I_S. Such systems can be employed also for comparing latency of different neurons on the first layer. As an example, imagine taking the limit of a very narrow Gaussian distribution in Fig. 3c. This would result in a very sharp profile of $n(I_S)$ in Fig. 3b. Such a system serves as a threshold device for $I_S > I_0$ or $T < T_0$, where I_0 is the center of the Gaussian, and T_0 its corresponding FPT. For this purpose even a single neuron on the second layer would suffice. Now let us use the same architecture to measure latency. We assume that the two neurons whose latency we compare constitute the first layer, and they are connected to the second layer with weights corresponding to the threshold function described above. If one neuron emits a spike and the second follows some time later, the system will react only if the time difference between these spikes is smaller than T_0. In order to design a system that will measure a latency of just T_0, we have to split the second layer into two parts, one connected so that it reacts to $T < T_0$, while the other reacts to $T < T_0 + \Delta T$. Now one may employ an eddititional neuron, on a third layer, connected to the two parts of the second layer with opposite synaptic weights, so that it will be activated by just the correct latency within an accuracy of ΔT. The minimal number of neurons needed for this purpose is two on the second layer and one on the third.

Thus, in the general debate whether the neural code depends on rate or latency, we can take an intermediate position. We have the freedom of making use of the FPT, which is equivalent to rate, and was sufficient for the problem discussed in this paper. We can however measure, using the same apparatus, temporal correlations, which are often assumed to play an important role in problems like binding and segmentation. In our problem we have used an information maximization approach to guide us in the design of the network. This is often regarded as a natural principle for biological development of synaptic connections. The same principle can, presumably, work for general latency based structures whenever it can be appropriately defined.

Appendix: Derivation of the First Passage Time

First passage time can be evaluated by employing conventional means for treating random walk problems. The probability function for the neuron's potential, eq. (5), obeys the backward Fokker Planck equation (Kampen, 1981)

$$\frac{\partial p(v,t \mid v',t')}{\partial t'} = -a_1(v')\frac{\partial}{\partial v'}p(v,t \mid v',t') - \frac{1}{2}a_2\frac{\partial^2}{\partial v'^2}p(v,t \mid v',t') \tag{23}$$

with

$$a_1(v) = \frac{1}{\tau}(\eta_I - v) \quad a_2 = \frac{A}{\tau^2}. \tag{24}$$

This defines an Ornstein-Ulenbeck process (Kampen, 1981), with the condition

$$p(v,t \mid v',t) = \delta(v - v'). \tag{25}$$

We are interested in the time T that passes from the point that the membrane potential is v_0 until it reaches the threshold V_{th}. Following Gardiner (1983) we calculate T by erecting an absorbing barrier at V_{th} and a reflecting barrier at V_{min}, the lowest potential value that the membrane can sustain.

The probability that at time t the potential has not passed V_{th}, i.e. the probability that T$\geq t$ is given by

$$G(v_0,t) \equiv \int_{V_{min}}^{V_{th}} p(v,t \mid v_0,0)\mathrm{d}v.$$

Employing the symmetry $p(v,t \mid v_0,0) = p(v,0 \mid v_0,-t)$, the backward Fokker Planck equation can be written in the form

$$\frac{\partial p(v,t \mid v_0,0)}{\partial t} = a_1(v_0)\frac{\partial}{\partial v_0}p(v,t \mid v_0,0) + \frac{1}{2}a_2\frac{\partial^2}{\partial {v_0}^2}p(v,t \mid v_0,0). \tag{26}$$

Hence, $G(v_0,t)$ obeys

$$\frac{\partial G(v_0,t)}{\partial t} = a_1(v_0)\frac{\partial G(v_0,t)}{\partial v_0} + \frac{1}{2}a_2\frac{\partial^2 G(v_0,t)}{\partial v_0^2}. \tag{27}$$

The initial conditions are

$$
\begin{aligned}
G(v_0, 0) &= 1 \;\; V_{min} \leq v_0 \leq V_{th} \\
&= 0 \;\; elsewhere
\end{aligned}
\tag{28}
$$

and the boundary conditions are $G(V_{th}, t) = 0$, $\frac{\partial G(V_{min},t)}{\partial v} = 0$.

Since $G(v_0, t)$ is the probability that T$\geq t$, the mean of the first passage time is

$$
T = \langle \mathrm{T} \rangle = -\int_0^\infty t \cdot \frac{\partial G(v_0, t)}{\partial t} \mathrm{d}t = \int_0^\infty G(v_0, t) \mathrm{d}t. \tag{29}
$$

Integrating eq. (27) and noting that

$$
\int_0^\infty \frac{\partial G(v_0, t)}{\partial t} \mathrm{d}t = G(v_0, \infty) - G(v_0, 0) = -1,
$$

one obtains

$$
a_1(v_0) \frac{\partial T}{\partial v_0} + \frac{1}{2} a_2 \frac{\partial^2 T}{\partial v_0^2} = -1, \tag{30}
$$

with boundary condition $\frac{\partial T(v_0 = V_{min})}{\partial v_0} = 0$, $T(v_0 = V_{th}) = 0$. The solution of this equation is the desired expression for the FPT

$$
T = 2\frac{\tau^2}{A} \int_{V_0}^{V_{th}} \mathrm{d}v \cdot e^{\frac{\tau}{A}(v - \eta_I)^2} \int_{V_{min}}^{v} e^{-\frac{\tau}{A}(z - \eta_I)^2} \mathrm{d}z. \tag{31}
$$

References

[1] J. J. Atick, 1992. Ecological theory of sensory processing. Network, Computation in Neural Systems, 3, 213-251.

[2] H. Barlow, 1989. Unsupervised learning. Neural Computation, 1, 295-311.

[3] T. N. Cornsweet, 1970. *Visual Perception*. Academic Press, New York.

[4] T. M. Cover and J. A. Thomas, 1991. Elements of Information Theory. John Wiley & Sons, Inc.

[5] D. Ferster and N. Spruston, 1995. Cracking the neuronal code. Science, 270, 756-757.

[6] C. W. Gardiner, 1983.*Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences.* Springer Verlag.

[7] J. J. Hopfield, 1995. Pattern recognition computation using action potential timing for stimulus representation. Nature 376, 33-36.

[8] N. G. van Kampen, 1981. *Stochastic processes in physics and chemistry*, North Holland; Amsterdam, New York.

[9] B. W. Knight, 1972. Dynamics of encoding in a population of neurons. J. of Gen. Physiol. 59, 734-766.

[10] P. Lansky, L. Sacerdote and F. Tomasetti, 1995. On the comparison of Feller and Ornstein-Uhlenbeck models for neural activity. Biol. Cybern., 73, 457–465.

[11] R. Linsker, 1988. Self-organization in a perceptual network. IEEE Computer 21, 105-117.

[12] W. Maass, 1997. Fast sigmoidal networks via spiking neurons. Neural Computation 9, 279-304.

[13] R. Mirollo and S. Strogatz, 1990. Synchronization of pulse-coupled biological oscillators. SIAM J. in Appl. Math. 50, 1645-1662.

[14] L. M. Ricciardi and L. Sacerdote, 1979. The Ornstein-Uhlenbeck process as a model for neuronal activity. Biol. Cybern. 35, 1–9.

[15] M. N. Shadlen and W. T. Newsome, 1994. Noise, neural codes and cortical organization. Current Opinion in Neurob. 4, 569-579.

[16] W. R. Softky, 1995. Simple codes versus efficient codes. Current Opinion in Neurob. 5, 239-247.

[17] M. Stemmler, 1996. A single spike suffices: the simplest form of stochastic resonance in model neurons. Network: Comp. in Neural Systems 7, 687-716.

[18] D. Tal and E. L. Schwartz, 1997. Computing with the leaky integrate-and-fire neuron: Logarithmic computation and multiplication. Neural Computation 9, 305-318.

[19] S. Thorpe, D. Fize and C. Marlot, 1996. Speed of processing in the human visual system. Nature, 381, 520-522.

[20] H. C. Tuckwell and D. K. Cope, 1980. Accuracy of neuronal interspike times calculated from a diffusion approximation. J. Theo. Biol. 83, 377–387.

[21] H. C. Tuckwell, 1989. *Stochastic processes in the neuroscience.* Society for Industrial and Applied Mathematics.

[22] C. van Vreeswijk and D. Hansel, 1997. Rhythmic bursting in networks of adaptive spiking neurons. Presented at sixth annual computational neuroscience meeting, CNS97.