

# Synaptic Transimission, Rate Model and Recurrent Neural Network

Quan Wen

22 Oct 2019

## Quantal hypothesis of neurotransmitter release

Bernard Katz and colleagues, armed with the newly developed intracellular recording techniques, conducted experiments to study the mechanisms of neuromuscular synaptic transmission in the early 1950s. While studying the end-plate potential (EPP) by nerve stimulation in the frog neuromuscular junction, they observed the muscle fibers also exhibited small EPPs in the absence of any nerve stimulation, and these were termed miniature end-plate potentials or mEPPs. mEPPs has intriguing properties: they seemed to have a defined, unitary size or occasionally a multiple of defined size. The release of neurotransmitter also requires  $\text{Ca}^{2+}$ . When the extracellular solution contains very low concentrations of  $\text{Ca}^{2+}$ , most nerve stimulations would not evoke any EPPs. When the EPPs did occur, its amplitude is the same of mEPPs. Further reduction of the  $\text{Ca}^{2+}$  change the frequency of mEPPs occurrence, but not its amplitude. This led to the Quantal hypothesis of neurotransmitter release, that is, neurotransmitters are released in discrete quanta of uniform size.

How can we test this hypothesis? In low  $\text{Ca}^{2+}$ , the release probability is also very small. If different quanta are released independently, the probability of releasing  $k$  quanta can be calculated by the Poisson distribution:

$$P(k) = \frac{m^k}{k!} e^{-m}, \quad (1)$$

where  $m$  is the mean number of quanta that respond to an individual stimulus. By knowing the mean  $m$ , theoretically, we could immediate calculate the number of observations that have no quanta release, one quanta release, two quanta release, etc. Let us define the noise distribution of EPP is  $P(v|k=0)$ , that is no quanta is released. Let us assume that it follows a gaussian distribution with mean  $\xi_0$  and variance  $\sigma_0^2$ .

$$P(v|k=0) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left[ -\frac{(v - \xi_0)^2}{2\sigma_0^2} \right] \quad (2)$$

The amplitude of a single quanta also has a defined variance  $\sigma_1^2$  and mean  $u$ . When  $k$  quanta are independently released, the total variance and mean should be  $k\sigma_1^2$  and  $ku$ . In other words, we have

$$P(v|k \neq 0) = \frac{1}{\sqrt{2\pi(\sigma_0^2 + k\sigma_1^2)}} \exp \left[ -\frac{(v - ku)^2}{2(\sigma_0^2 + k\sigma_1^2)} \right] \quad (3)$$

Now the EPP amplitude distribution  $P(v)$  should be given by

$$P(v) = \sum_{k=0}^{\infty} P(v, k) = \sum_{k=0}^{\infty} P(v|k)P(k) \quad (4)$$

This theoretical curve has a good agreement with the experimental distribution.

## Modeling synaptic inputs

The synaptic current  $I_s$  could be described by the following equation

$$I_s = \bar{g}_s P(V - E_s), \quad (5)$$

where  $P$  is the neurotransmitter release probability. If we consider the fact that the rising time has certain delay  $\tau_r$ , one could use the following equation

$$P = \frac{P_{max}t}{\tau_r} \exp(-t/\tau_s).$$

However, when  $\tau_r \ll \tau_s$ , we may assume that the release probability approaches some constant value  $P_{max}$  immediately after the the action potential arrives at time  $t_k$ , and then decay with a time constant  $\tau_s$ :  $P = P_{max}e^{-(t-t_k)/\tau_s}$ , for  $t > t_k$ . By multiplying both sides with  $\bar{g}_s P(V - E_s)$ , and by assuming the amplitude of EPSP is much less than the reversal potential of a synapse  $V \ll E_s$ , one could treat  $\bar{g}_s P_{max}(V - E_s) = w$  as the synaptic weight, and the total synaptic current is given by summing all the spikes from the past

$$I_s = w \sum_j K(t - t_k),$$

where  $K(t) = \exp(-t/\tau_s)$ . In the form of a differential equation, this equation can be rewritten as

$$\tau_s \frac{dI_s}{dt} = -I_s + \tau_s w \sum_k \delta(t - t_k) \quad (6)$$

When we are summing synaptic inputs coming from many neurons, the total synaptic currents become

$$\tau_s \frac{dI_s^i}{dt} = -I_s^i + \tau_s \sum_j w_{ij} \sum_k \delta(t - t_{j,k}) \quad (7)$$

## Spike-Train Statistics

Before we delve deeply into Equation ??, let's first look into the statistics of spike trains, which might help us solve this equation. As we discussed in previous lectures, such as in the integrate-and-fire model, constant external current could give rise to periodic spiking pattern with the same inter-spike interval. However, is this what we have seen in a real neuron? The answer is not. The patch-clamp recording of neuronal activity *in vivo* demonstrates that the firing pattern is highly irregular. To analyze its statistics, let's introduce a critical concept in the field of computational neuroscience, the firing rate of a neuron  $r$ . Formally, we have the following mathematical definition Given that

$$\rho(t) = \sum_k \delta(t - t_k) \quad (8)$$

The firing rate is defined as

$$r(t) = \frac{1}{\Delta t} \int_t^{t+\Delta t} d\tau \langle \rho(\tau) \rangle \quad (9)$$

where  $\langle \dots \rangle$  denotes an average over different trials, or even different neurons with very similar firing patterns. In practice, we may even choose a specific weighting function  $\omega(\tau)$ , which leads to

$$r_{approx}(t) = \int_{-\infty}^{\infty} \omega(\tau) \rho(t - \tau) d\tau$$

A useful weighting function is

$$\omega(\tau) = [\alpha^2 \tau \exp(-\alpha \tau)]_+,$$

which implies that the firing rate at time  $t$  depends only on spikes fired before  $t$ . The weight function vanishes when the argument  $\tau$  is negative.

With the introduction of firing rate, we could now discuss the statistics of spike-train with some mathematical rigid. The spike-train can be viewed as a point process. In general, the probability of an event occurring at any given time could depend on the entire history of preceding events. However, if the events are themselves statistically independent, we have a Poisson process. The Poisson process provides an extremely useful approximation of irregular neuronal firing, as we will see immediately.

We denote the firing rate for a homogeneous Poisson process by  $r(t) = r$ , because it is independent of time. For a Poisson process, the probability of firing  $n$  spike in a time interval  $T$  is given by

$$P(n) = \frac{(rT)^n}{n!} \exp(-rT), \quad (10)$$

where  $rT$  is simply the mean number of spike in the time window  $T$ . Note that for large  $rT$ , the poisson distribution approaches a gaussian distribution. The variance of a Poisson distribution,

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = rT$$

Thus the variance and mean of the spike count are equal. The ratio of these two quantities,  $\sigma_n^2/\langle n \rangle$  is called the Fano factor and takes the value of 1 for homogeneous Poisson process, independent of time interval  $T$ .

Another important quantity that needs to be mentioned is the interspike interval. let's denote  $\tau < t_{i+1} - t_i < \tau + \Delta\tau$ , that is after the  $i$ th spike, the  $i + 1$  spike will occur between  $t_i + \tau$  and  $t_i + \tau + \Delta\tau$ . The probability of having no spike in an interval  $\tau$  is  $\exp(-r\tau)$ , and the probability to have one spike to occur between  $t_i + \tau$  and  $t_i + \tau + \Delta\tau$  is simply  $r\Delta\tau$ . Thus, the probability density to have an interspike interval  $\tau$  is just given by

$$p(\tau) = r \exp(-r\tau)$$

. Moreover, We can calculate the mean interspike interval

$$\langle \tau \rangle = \int_0^\infty \tau r d\tau \exp(-r\tau) \quad (11)$$

Integrating by part, we found not surprisingly,  $\langle \tau \rangle = 1/r$ . The variance of the interspike interval is given by  $\sigma^2(\tau) = \langle \tau^2 \rangle - \langle \tau \rangle^2$ .

$$\langle \tau^2 \rangle = \int_0^\infty \tau^2 r d\tau \exp(-r\tau) = \lim_{\alpha \rightarrow 1} \frac{1}{r} \frac{d^2}{d\alpha^2} \int_0^\infty d\tau \exp(-\alpha r \tau) = \frac{2}{r^2}$$

Thus,  $\sigma^2(\tau) = \frac{1}{r^2}$ . As mentioned in the last lecture, we can also define the coefficient of variation of the interspike interval  $\mathbf{CV} = \sigma/\langle \tau \rangle$ . For poisson process, we found  $\mathbf{CV}=1$ .

When the firing rate depends on time, we could also extend the homogeneous Poisson process to inhomogeneous Poisson process. When  $n$  spikes occurs in an interval  $T$  with  $0 < t_1 < t_2 < \dots < t_n < T$ , the probability density is given by

$$p[t_1, t_2, \dots, t_n] = \exp\left(-\int_0^T r(t)dt\right) \prod_{i=1}^n r(t_i) \quad (12)$$

## Firing rate model

In previous lectures, we have discussed the integrate-and-fire model, and how deterministic spiking pattern would arise given a constant external input. However, each neuron receives time-varying inputs from thousands of other neurons, and experimental data have shown that the spike timing for cortical network

is stochastic, making it difficult to treat Equation ???. Indeed, spiking model imposes a great computational and practical challenge. To ameliorate these difficulties, we shall substitute the spiking variable  $\rho(t) = \sum_k \delta(t - t_k)$  with the firing rate  $r(t)$ .

Now in a firing rate model, Equation ??? is changed to

$$\tau_s \frac{dI_s}{dt} = -I_s + \tau_s w r(t)$$

Summing up the synaptic inputs from all the neurons, we shall have

$$\tau_s \frac{dI_s}{dt} = -I_s + \tau_s \sum_j^N w_j r_j(t), \quad (13)$$

where  $j$  denotes the neuron number  $j$ .

Equation ??? shows how the synaptic current of a neuron depends on the firing rate of all presynaptic neurons. To finish formulating the firing rate model, we should also specify how the firing rate of the postsynaptic neuron  $r_i$  depends on the total synaptic inputs  $I_s$ . Recall that the membrane potential of a neuron could be modeled as a low-pass filter of the applied external inputs. Motivated by this observation, the firing rate of the postsynaptic neuron may be written as

$$\tau_r \frac{dr_i}{dt} = -r_i + F(I_s(t)) \quad (14)$$

I must emphasize that this is a phenomenological characterization. It describes how rapid the postsynaptic firing rate follows the time-varying synaptic currents.  $F(\dots)$  is a nonlinear function of the synaptic current, say  $F(I_s) = [I_s - \gamma]_+$ . Note that  $\tau_r$  is not the same as the membrane time constant  $\tau_m$ , for the latter is used to describe the membrane potential.

Equations ??? and ??? provides a complete description of the firing rate model. In the limit  $\tau_s \ll \tau_r$ , the synaptic current dynamics approaches steady state  $I_s^i = \tau_s \sum_j^N w_j r_j(t)$  much faster, and Equation ??? becomes

$$\tau_r \frac{dr_i}{dt} = -r_i + F\left(\sum_{j=1}^N w_{ij} r_j(t)\right) \quad (15)$$

Note that in practice  $\tau_s$  has been absorbed in the synaptic weight  $w_{ij}$ .

In the opposite limit  $\tau_r \ll \tau_s$ , we shall write

$$\begin{aligned} \tau_s \frac{dI_s}{dt} &= -I_s + \tau_s \sum_j^N w_j r_j(t), \\ r_i &= F(I_s(t)) \end{aligned} \quad (16)$$

## Feedforward and Recurrent Networks

There are two basic network topologies, feedforward network and recurrent network. The feedforward network has  $N_v$  output units with firing rate  $v_i$  ( $i = 1, 2, \dots, N_v$ ), denoted collectively by vector  $\mathbf{v}$ , driven by  $N_u$  input units with rates  $\mathbf{u}$ . Now we could also replace and extend the synaptic weight vectors  $w_{ij}$  with a matrix  $\mathbf{W}$ , with the matrix component  $w_{ij}$  representing the strength of the synapses from input unit  $j$  to output unit  $i$ .

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + F(\mathbf{W} \cdot \mathbf{u}) \quad (17)$$

The recurrent network also has two layers of neurons with rates  $\mathbf{u}$  and  $\mathbf{v}$ . But now the output layers are also interconnected with each other with a synaptic weight matrix  $\mathbf{M}$ . In this case, the output firing rates are determined by

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + F(\mathbf{W} \cdot \mathbf{u} + \mathbf{M} \cdot \mathbf{v}) \quad (18)$$

It is often convenient to define the total feedforward input to each neuron as a vector  $\mathbf{h} = \mathbf{W} \cdot \mathbf{u}$ . Then, the output rates are

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + F(\mathbf{h} + \mathbf{M} \cdot \mathbf{v}) \quad (19)$$

Neurons are typically classified as either excitatory or inhibitory, meaning that they have either excitatory or inhibitory effects on all of their postsynaptic targets. This property is formulated as Dale's rule, which states that a neuron cannot excite some of the postsynaptic cells and inhibit others. This rule imposes strong constraints on the weight matrix. For example,  $M_{ij}$  must have the same sign for all the postsynaptic neurons.

Now let's consider some special cases for Equation ???. For example, by taking away  $F$  and consider a linear relationship between firing rate and synaptic current. This is a drastic assumption, as it would allow  $v$  to be negative, which is impossible. Nevertheless, the linear model is extremely useful for exploring properties of recurrent circuits as we will see below. In this case, the equation will be changed to

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + \mathbf{h} + \mathbf{M} \cdot \mathbf{v} \quad (20)$$

Now let's consider the simplest case where  $M$  is symmetric. In this case, the eigenvectors  $\mathbf{e}_\mu$  for  $\mu = 1, 2, \dots, N_v$  satisfy

$$\mathbf{M} \cdot \mathbf{e}_\mu = \lambda_\mu \mathbf{e}_\mu$$

Because this is a normal matrix,  $\mathbf{e}_\mu \cdot \mathbf{e}_\nu = \delta_{\mu\nu}$ . In particular, we can use these basis to represent

$$\mathbf{v}(t) = \sum_{\mu=1}^{N_v} c_\mu(t) \mathbf{e}_\mu$$

Plugging this back to the original equation, we have

$$\tau_r \sum_{\mu=1}^{N_v} \frac{dc_\mu(t)}{dt} \mathbf{e}_\mu = - \sum_{\mu=1}^{N_v} (1 - \lambda_\mu) c_\mu(t) \mathbf{e}_\mu + \mathbf{h}.$$

The sum over  $\mu$  can be eliminated by taking the dot product on both sides of the equation, and by using  $\mathbf{e}_\mu \cdot \mathbf{e}_\nu = \delta_{\mu\nu}$ , we have

$$\tau_r \frac{dc_\mu(t)}{dt} = -(1 - \lambda_\mu) c_\mu(t) + \mathbf{e}_\mu \cdot \mathbf{h}.$$

The solution to this equation is given by

$$c_\mu(t) = \frac{\mathbf{e}_\mu \cdot \mathbf{h}}{1 - \lambda_\mu} [1 - \exp(-t(1 - \lambda_\mu)/\tau_r)] + c_\mu(0) \exp(-t(1 - \lambda_\mu)/\tau_r)$$

Thus the steady state firing rate is given by

$$\mathbf{v}_\infty = \sum_{\mu=1}^{N_v} \frac{\mathbf{e}_\mu \cdot \mathbf{h}}{1 - \lambda_\mu} \mathbf{e}_\mu$$

## Input Integration and short term memory

If the recurrent weight matrix has an eigenvalue exactly equal to 1,  $\lambda_1 = 1$ , and all the other eigenvalue satisfy  $\lambda_\mu < 1$ , a linear recurrent network can act as an integrator of its input. In this case,  $c_1$  satisfy the following equation

$$\tau_r \frac{dc_1}{dt} = \mathbf{e}_1 \cdot \mathbf{h}.$$

For arbitrary time-dependent inputs, the solution is given by

$$c_1(t) = c_1(0) + \frac{1}{\tau_r} \int_0^t dt' \mathbf{e}_1 \cdot \mathbf{h}(\mathbf{t}')$$

Now let's imagine that  $\mathbf{h}$  represents sensory stimulus represented to the network. If this input is transient, that is  $\mathbf{h}(\mathbf{t}) = \mathbf{s}_0 \delta(t - t_0)$ , then the stimulus representation can be stored persistently within the network.

$$c_1(t) \sim c_1(0) + \mathbf{e}_1 \cdot \mathbf{s}_0$$

In general, during a working memory task, we would like to hold the percept of a past memory in the neural circuit for an indefinitely period of time. Conceptually, this means that the representation of the sensory stimulus should be time invariant during that period of time. Mathematically, this means that

$$\frac{d\mathbf{s}}{dt} = 0 \tag{21}$$

In the linear coding framework, the representation of a sensory stimulus  $\mathbf{s}$ , is a linear combination of each neuron's receptive field, which can be viewed as a feature vector  $\mathbf{d}_i$ , weighted by the activity of that neuron

$$\mathbf{s} = \sum_i r_i(t) \mathbf{d}_i \quad (22)$$

Writing everything in the matrix form, where  $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2 \dots\}$ , and  $\mathbf{r} = [r_1, r_2 \dots]^T$ , we have

$$\mathbf{s} = \mathbf{D}\mathbf{r}$$

Equation ?? can now be written as

$$\sum_i \mathbf{d}_i \frac{dr_i(t)}{dt} = 0 \quad (23)$$

If each neuron encodes a different feature vector  $\mathbf{d}_i$  and all these feature vectors are linearly independent, then we must have

$$\frac{dr_i(t)}{dt} = 0$$

for all neurons. This leads to time-invariant persistent activity as discussed previously. However, if these feature vectors are *not* linearly independent, then even individual neurons have time-varying activity patterns, the representation as a *whole* can still remain constant. we now decouple the activity vector into two terms, namely

$$\mathbf{r}(t) = \mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{s} + \mathbf{r}_\perp(t) \quad (24)$$

where

$$\mathbf{D}\mathbf{r}_\perp(t) = 0, \quad (25)$$

Here  $\mathbf{r}_\perp(t)$  belongs to the null space of  $\mathbf{D}$ .

The geometrical meaning of the null space is the following. Consider the row vectors of  $\mathbf{D}$ . They expand a hyperplane and  $\mathbf{r}_\perp$  is any vector orienting perpendicular to the hyperplane.

Now let's consider a recurrent neural network

$$\tau_r \frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{M}\mathbf{r} \quad (26)$$

Plugging ?? into Equation ??, we have

$$\mathbf{D}(-\mathbf{r} + \mathbf{M}\mathbf{r}) = 0 \quad (27)$$

We impose that this should be true for all activity patterns. As a result, we found that

$$\mathbf{D} = \mathbf{D}\mathbf{M} \quad (28)$$



Again, consider the eigenvectors  $\mathbf{e}_\mu$  for  $\mu = 1, 2, \dots, N_v$  satisfy

$$\mathbf{M}\mathbf{e}_\mu = \lambda_\mu \mathbf{e}_\mu$$

Equation ?? thus requires that

$$\mathbf{D}\mathbf{e}_\mu = \lambda_\mu \mathbf{D}\mathbf{e}_\mu \quad (29)$$

This equation only has two possible solutions. First,  $\lambda_\mu = 1$ . This is very familiar, the solution of one dimensional dynamics with persistent activity. And second,

$$\mathbf{D}\mathbf{e}_\mu = 0,$$

that is, the eigenvectors of the recurrent matrix should also belong to the null space of the dictionary.

## Nonlinear Ring Model

We have talked about working memory in the context of encoding a continuous variable such as eye position in the oculomotor integrator of the goldfish, in which the firing rates of a group of neurons vary monotonically with eye positions. The neural activity can be viewed as a projection of the stimulus onto a one dimensional vector (the eigenvector of the recurrent connectivity matrix for example), and this is the so-called line attractor. In the second scheme, neural activity may represent selective features of the external world: for example, each neuron has a preferred spatial location at which it fires with highest rate. Examples are the representation of space in place cells, coding of head direction in head direction systems, and orientational selectivity in V1.

Below, I will discuss how recurrent neural dynamics can be used to implement the second scheme, and in the more abstract way, the angular variable. This variable could be the orientation of a bar in the visual field, the head direction as you move around. The ring network is a very useful and beautiful framework. Let us introduce the model

$$\tau \frac{du_i}{dt} = -u_i + F \left( \sum_{j=1..N} w_{ij} u_j + I_i^0 \right), \quad (30)$$

where  $F(x) = [x]_+$  is a rectified linear function and the weights between the neurons are determined by their angular difference and thus are translational invariant.

$$w_{ij} = \frac{1}{N} J(\theta_i - \theta_j), \quad (31)$$

where  $\theta_i = -\pi + \frac{2\pi i}{N}$  and  $J$  is a  $2\pi$  periodic function, and  $I_i^0$  is also a periodic  $2\pi$  function.

In the large  $N$  limit, we can replace  $\sum_{j=1..N} \frac{1}{N} J(\theta_i - \theta_j) u_j(t)$  by an integral, and write

$$\tau \frac{\partial u(\theta, t)}{\partial t} = -u(\theta, t) + \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} J(\theta - \theta') u(\theta', t) d\theta' + I^0(\theta) \right]_+, \quad (32)$$

where we have equated  $\frac{1}{N} = \frac{\Delta\theta}{2\pi}$  and then take the limit  $\Delta\theta \rightarrow 0$ .

Here let us consider the Fourier transform of the weight matrix .

$$J(\theta) = \mathbf{J}_0 + \mathbf{J}_1 \cos \theta + \dots \quad (33)$$

$$I^0(\theta) = I_0 + I_1 \cos(\theta - \theta_0), \quad (34)$$

and let us assume that the external stimulus is centered at  $\theta_0$ , and we shall also consider the simplest model by ignoring the higher order term in  $J$ .

Now let us consider the steady state solution of the ring model:

$$u(\theta) = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} J(\theta - \theta') u(\theta') d\theta' + I^0(\theta) \right]_+ \quad (35)$$

Since  $\cos(\theta - \theta') = \text{Re} \exp(i(\theta - \theta'))$ , we can now plug the expression for  $J$  and  $I^0$ , and obtain

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(\theta') d\theta' \quad (36)$$

$$a_1 \exp(-i\phi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} u(\theta') \exp(-i\theta') d\theta' \quad (37)$$

$$u(\theta) = [J_0 a_0 + I_0 + J_1 a_1 \cos(\theta - \phi) + I_1 \cos(\theta - \theta_0)]_+ \quad (38)$$

Now let us first consider that (1) we are at a linear region, such that the right hand side of the equation is always larger than zero. And let us make an *ansatz* that the solution has the form

$$u(\theta) = u_0 + u_1 \cos(\theta - \theta_0) \quad (39)$$

This would be the solution provided that

$$a_0 = J_0 a_0 + I_0 \quad (40)$$

$$a_1 = \frac{1}{2} (J_1 a_1 + I_1) \quad (41)$$

$$\phi = \theta_0 \quad (42)$$

In other words, we have

$$a_0 = \frac{I_0}{1 - J_0} \quad (43)$$

$$a_1 = \frac{I_1}{2 - J_1} \quad (44)$$

And the steady state neural activity is given by

$$u(\theta) = \frac{I_0}{1 - J_0} + \frac{2I_1}{2 - J_1} \cos(\theta - \theta_0) \quad (45)$$

Case 1: what happens when  $J_0 > 1$ ?

Case 2: what happens when  $J_1 > 2$ ? Here, let us consider a special case where there is no specific external inputs, so that  $I_1 = 0$ . Without losing generality, let us also consider a special case where  $\theta_0 = 0$ . We thus write  $u(\theta) = [u_0 + u_1 \cos \theta]_+ = u_1 [\cos \theta - \cos \theta_c]_+$ , where  $\cos \theta_c = -u_0/u_1$ . According to Eq. ??, we shall have

$$u_1 = \frac{J_1}{2\pi} \int_{-\theta_c}^{\theta_c} u_1 \cos \theta (\cos \theta - \cos \theta_c) d\theta \quad (46)$$

As a result, we have

$$1 = J_1 \frac{1}{2\pi} \left( \theta_c - \frac{1}{2} \sin 2\theta_c \right) \quad (47)$$

Likewise, we also have

$$-u_1 \cos \theta_c = \frac{J_0}{2\pi} \int_{-\theta_c}^{\theta_c} u_1 (\cos \theta - \cos \theta_c) d\theta + I_0 \quad (48)$$

Now define

$$f(\theta_c) = \frac{1}{\pi} (\sin \theta_c - \theta_c \cos \theta_c),$$

we obtain

$$u_1 = \frac{-I_0}{f(\theta_c)J_0 + \cos \theta_c} \quad (49)$$

## Energy landscape and Fixed point Dynamics

As discussed above, in many cases, we are interested in steady-state solutions of neural network dynamics. What are the steady solutions? By analogy of mechanics, we could consider an energy landscape. A ball is moving up and down the hills. The steady state solution is the local minimum of the energy landscape. Let's consider the following equation

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + F(\mathbf{h} + \mathbf{M} \cdot \mathbf{v}), \quad (50)$$

where  $F$ , the activation function, specifies the relationship between steady state firing rate and the total synaptic input current. In the following, we shall show that such an energy function does exist under certain conditions. To this end, let us define

$$\mathbf{s} = F(\mathbf{h} + \mathbf{M} \cdot \mathbf{v}),$$

and the inverse function for  $F$  as  $F^{-1}$ . We are using a similar notation as that for the Hopfield network. The energy function is defined as

$$E = -\frac{1}{2} \sum_{ij} s_i M_{ij} s_j - \sum_i h_i s_i + \sum_i V(s_i), \quad (51)$$

$$V = \int_0^{s_i} F^{-1}(x) dx$$

Next we shall prove that for symmetric and nonsingular weight matrix  $M$ ,  $\frac{dE}{dt} \leq 0$  and it is bounded from below for a biologically plausible activation function  $F$ . To show this, note that

$$\frac{dE}{dt} = \sum_i \frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial t} + \sum_j \frac{\partial E}{\partial s_j} \frac{\partial s_j}{\partial t}$$

Substituting the expression for  $E$ , we have

$$\frac{dE}{dt} = -\frac{1}{2} \left( \sum_i \sum_j M_{ij} s_j \frac{\partial s_i}{\partial t} + \sum_j \sum_i s_i M_{ij} \frac{\partial s_j}{\partial t} \right) - \sum_i h_i \frac{\partial s_i}{\partial t} + \sum_i F^{-1}(s_i) \frac{\partial s_i}{\partial t}$$

If the synaptic weight matrix is symmetric,  $M_{ij} = M_{ji}$ , the above equation is further simplified. Writing it down in the matrix form, we have

$$\frac{dE}{dt} = [-\mathbf{M}\mathbf{s} - \mathbf{h} + F^{-1}(\mathbf{s})] \cdot \frac{d\mathbf{s}}{dt}$$

We next derive the explicit expression for  $\frac{d\mathbf{s}}{dt}$ .

$$\frac{d\mathbf{s}}{dt} = F' \mathbf{M} \frac{d\mathbf{v}}{dt}$$

We can further express  $\mathbf{v}$  in terms of  $\mathbf{s}$  using the inverse function  $F^{-1}$ :

$$\mathbf{v} = \mathbf{M}^{-1}[F^{-1}(\mathbf{s}) - \mathbf{h}],$$

And

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + \mathbf{s} = -\mathbf{M}^{-1}[F^{-1}(\mathbf{s}) - \mathbf{h}] + \mathbf{s}.$$

Thus,

$$\tau_r \frac{d\mathbf{s}}{dt} = F' [-F^{-1}(\mathbf{s}) + \mathbf{h} + \mathbf{M}\mathbf{s}]$$

As a result, the gradient of the energy function is given by

$$\frac{dE}{dt} = -\frac{1}{\tau_r} F' \|\mathbf{h} + \mathbf{M}\mathbf{s} - F^{-1}(\mathbf{s})\|^2 \leq 0 \quad (52)$$

The energy function would be bounded from below if  $\mathbf{s}$  does not diverge. Indeed, this is a biologically plausible assumption. No neurons can have excessively high firing rate. In practice,  $F$  has been chosen as a sigmoidal function so that its value saturates for high synaptic currents.

## Balanced excitation and inhibition

Experimental data suggest that the firing pattern of a neuron *in vivo* mimic the poisson process. The *CV* of interspike interval is close to one, and the variance of spike count in a defined time window is close to the mean. Therefore, it would be interesting to understand how does such stochasticity emerge from the interaction of thousands of neurons in the neural circuit. Let us consider a toy model, in which each neuron only has two states  $\sigma_i \in [0, 1]$ . At any given time, the probability that a neuron would be active is  $p$ . Each postsynaptic neuron would receive inputs from  $N$  presynaptic neurons, and the total synaptic inputs at any given time is given by

$$I_s(t) = w \sum_{i=1}^N \sigma_i(t) \quad (53)$$

Now let us calculate the mean and variance of the synaptic inputs. The mean is simply given by  $\langle I_s(t) \rangle = wNp$ . The variance

$$\begin{aligned} \text{Var}(I_s) &= \langle I_s^2 \rangle - \langle I_s \rangle^2 \\ &= w^2 \left\langle \sum_{i=1}^N \sigma_i^2 + \sum_{j=1}^N \sum_{i=1, i \neq j}^N \sigma_i \sigma_j \right\rangle - w^2 \left\langle \sum_{i=1}^N \sigma_i \right\rangle^2 \\ &= w^2 \sum_{i=1}^N \langle \sigma_i^2 \rangle + w^2 \sum_{j=1}^N \sum_{i=1, i \neq j}^N \langle \sigma_i \sigma_j \rangle - w^2 \sum_{i=1}^N \langle \sigma_i \rangle^2 - w^2 \sum_{j=1}^N \sum_{i=1, i \neq j}^N \langle \sigma_i \rangle \langle \sigma_j \rangle \\ &= w^2 \sum_{i=1}^N [\langle \sigma_i^2 \rangle - \langle \sigma_i \rangle^2] + w^2 \sum_{j=1}^N \sum_{i=1, i \neq j}^N [\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle] \\ &= w^2 N(p - p^2) + N(N-1) \text{terms of CC} \end{aligned} \quad (54)$$

If all inputs are independent of each other, then the cross correlation terms are zero. The variance  $\text{Var}(I_s) = w^2 N(p - p^2)$ . Now let us consider the mean and variance of the total inputs as a function of the total number of inputs  $N$ . As the network is getting larger and larger,  $N$  increases, but we do not want the activity of the neurons to explode, but to have a finite mean and variance. One way to do this is to scale the synaptic weight as  $w = J/N$ . In this case,  $\langle I_s(t) \rangle = Jp$ , and  $\text{Var}(I_s) = J^2(p - p^2)/N \rightarrow 0$ . Basically, in a large network with independent synaptic inputs, the variance would go to zero, and all neurons should fire periodically!

How can we resolve this paradox between experimental data and the theoretical argument, which is basically the application of central limit theorem? Well, one solution is the so-called balanced excitation and inhibition. The total synaptic

inputs to a neuron is the sum of excitation and inhibition. In other words,

$$\begin{aligned}\langle I_s(t) \rangle &= w_e N p_e - w_i N p_i \\ \text{Var}(I_s) &= w_e^2 N (p_e - p_e^2) + w_i^2 N (p_i - p_i^2)\end{aligned}\tag{55}$$

To achieve finite variance that is independent of  $N$ , let  $w_e = J_e/\sqrt{N}$  and  $w_i = J_i/\sqrt{N}$ , and thus  $\text{Var}(I_s) = J_e^2(p_e - p_e^2) + J_i^2(p_i - p_i^2)$ , and  $\langle I_s(t) \rangle = \sqrt{N}(J_e p_e - J_i p_i)$ . Now we impose a balanced condition,  $J_e = J_i p_i/p_e$ , which means on average the inhibitory inputs completely cancel out the excitatory inputs.