

# Mathematical Modelling for Neuroscience - Lecture Notes

Fabio Pernisi

March 2023

## Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Properties of Neurons . . . . .	5
1.2	Recording Neuronal responses . . . . .	5
1.3	Neuronal membrane . . . . .	6
1.3.1	Movement of ions through the membrane . . . . .	6
1.3.2	Nernst-Planck Equation . . . . .	6
1.3.3	Goldman-Hodgkin-Katz equation . . . . .	7
<b>2</b>	<b>Model Neurons: Neuroelectronics</b>	<b>8</b>
2.1	Electrical Properties of Neurons . . . . .	8
<b>3</b>	<b>Hodgkin-Huxley model</b>	<b>10</b>
3.1	Equivalent Circuit . . . . .	10
3.2	Voltage-gated Channels . . . . .	10
3.3	The model . . . . .	11
3.4	Action Potential revisited . . . . .	12
<b>4</b>	<b>Generalization of the Hodgkin-Huxley model</b>	<b>14</b>
4.1	Diversity of electrophysiological behaviors of single neurons . . . . .	14
4.2	Diversity of Voltage-gated channels . . . . .	14
4.3	Persistent sodium: bursting . . . . .	14
4.4	M current: spike frequency adaptation . . . . .	15
4.5	Single Compartment models // HH-type to Integrate-and-fire . . . . .	16
4.6	Dynamics of Spike Initiation: HH vs LIF . . . . .	17
4.6.1	Quadratic Integrate-and-Fire . . . . .	17
4.6.2	Exponential Integrate-and-Fire . . . . .	18
4.7	Adaptive nonlinear Integrate-and-Fire models . . . . .	18
<b>5</b>	<b>Response of neurons to stochastic inputs</b>	<b>20</b>
5.1	Fluctuations: in vivo, in vitro . . . . .	20
5.2	Statistics of neuronal firing . . . . .	20
5.3	Modeling spike trains as point processes . . . . .	20
5.4	LIF model with Poisson inputs . . . . .	22
5.5	Diffusion Approximation . . . . .	22
5.6	Fokker-Planck Equation . . . . .	24
<b>6</b>	<b>Synapses</b>	<b>26</b>
6.1	In vitro recording of synapses . . . . .	27
6.2	Kinetics of synaptic transmission . . . . .	27
6.3	Synaptic Failures . . . . .	27
6.4	Classification of synapses . . . . .	28
6.5	NMDA currents . . . . .	28
6.6	A Markov model for synaptic currents . . . . .	29

6.7	Stochastic nature of synaptic transmission . . . . .	30
6.8	Quantal model . . . . .	30
6.8.1	Distribution, mean and variance of synaptic currents . . . . .	30
6.9	Summary: models for Post-Synaptic Currents . . . . .	30
<b>7</b>	<b>History dependence - short term depression</b>	<b>31</b>
7.1	Tsodyks-Markram model . . . . .	31
7.2	Average synaptic currents as a function of the frequency . . . . .	31
<b>8</b>	<b>Long-term synaptic Plasticity</b>	<b>33</b>
8.1	LTP and LTD in hippocampal slices . . . . .	33
8.2	Paired Recordings . . . . .	33
8.2.1	Spike-timing dependent plasticity (STDP) protocol . . . . .	33
8.3	Diversity of STDP curves . . . . .	34
8.4	Phenomenological learning rules - STDP . . . . .	34
8.5	STDP rule fails to reproduce data obtained in other protocols . . . . .	35
8.6	Phenomenological learning rules - firing rate . . . . .	35
8.7	Long-term plasticity: Mechanisms . . . . .	35
8.8	Calcium-based models: Shouval et al. (2002) . . . . .	35
8.9	Reproducing standard protocols with the model . . . . .	36
8.10	Synaptic strength is controlled by multiple factors . . . . .	36
<b>9</b>	<b>Networks</b>	<b>37</b>
9.1	Connectomics: Obtaining wiring diagrams . . . . .	37
9.2	The C Elegans brain network - numbers . . . . .	37
9.3	C Elegans - Connectivity matrix . . . . .	37
9.4	Feed-forward, recurrent, feed-back connections . . . . .	37
9.5	Types of networks/graphs . . . . .	38
9.6	ER Network . . . . .	38
9.6.1	Properties of ER network . . . . .	38
9.7	Stochastic Block models . . . . .	38
9.8	Spatially dependent connection probability . . . . .	38
9.9	Properties of C-Elegans network . . . . .	38
9.10	Motifs in C-Elegans network . . . . .	38
9.11	Mammalian brains - macroscopic structure . . . . .	39
9.11.1	Cortical Connectivity . . . . .	39
9.11.2	Medium-scale patchy connectivity . . . . .	39
9.12	Local cortical circuits: numbers . . . . .	39
<b>10</b>	<b>Rate models</b>	<b>40</b>
10.1	Rate model . . . . .	40
10.2	The transfer function . . . . .	40
10.2.1	From populations of individual neurons to a rate model . . . . .	40
10.3	Rate models for local networks of neurons . . . . .	41
10.4	Analysis of rate models . . . . .	41
10.4.1	Simplest case . . . . .	42
10.4.2	Oculomotor Integrator . . . . .	42
10.5	Non-linear E network - bistability . . . . .	42
10.6	E network with slow negative feedback . . . . .	43
10.6.1	Dynamics of E networks with slow negative feedback . . . . .	43
10.7	I networks with delays - oscillations . . . . .	43
10.8	E-I network - oscillations . . . . .	43
10.9	I-I networks: winner takes it all . . . . .	45

<b>11 Balanced Networks</b>	<b>46</b>
11.1 Mechanisms of spontaneous activity in cortex . . . . .	46
11.2 Visualizing Network dynamics . . . . .	46
11.3 Network models: Parts list . . . . .	47
11.4 Randomly connected E-I network model . . . . .	47
11.4.1 Orders of magnitude of parameters for local cortical networks . . . . .	47
11.5 Simulations of E-I networks . . . . .	47
11.5.1 Synaptic inputs: means and fluctuating terms . . . . .	48
11.5.2 Large $K$ limit: . . . . .	49
11.5.3 Balance conditions . . . . .	50
11.6 Evidence of synaptic scaling in culture preparations . . . . .	50
<b>12 The ring model</b>	<b>51</b>
12.1 Examples . . . . .	51
12.1.1 Primary Visual Cortex . . . . .	51
12.1.2 Motor Cortex . . . . .	51
12.1.3 Prefrontal Cortex . . . . .	51
12.1.4 Head-direction cells . . . . .	52
12.2 Rate model for spatial selectivity . . . . .	52
12.3 The ring model . . . . .	53
12.3.1 Connectivity in the $J_0 - J_1$ plane . . . . .	53
12.3.2 Analysis of the model . . . . .	53
12.4 Uniform inputs - Phase diagram of the model . . . . .	54
12.5 Response of the model to tuned inputs . . . . .	54
12.5.1 A ring attractor in the fly . . . . .	56
<b>13 Single neuron coding I: Quantifying the dependence of spike trains on external stimuli</b>	<b>57</b>
13.1 Flow of information in visual detection . . . . .	57
13.2 Response to a single stimulus: PSTH and Fano Factor . . . . .	57
13.3 Firing rates vs external stimuli: Continuous case . . . . .	57
13.4 Rat olfactory cortex - odors . . . . .	58
13.5 Single-neuron encoding of dynamic stimuli . . . . .	58
13.6 Characterizing input/output transformation: Volterra series . . . . .	58
13.7 Wiener series . . . . .	59
13.7.1 Wiener kernels in the discrete case . . . . .	60
13.8 Visual system: Spatio-temporal receptive fields . . . . .	61
13.8.1 Receptive Fields in the early visual system . . . . .	61
13.8.2 Spatial Receptive Field of a simple cell in V1 . . . . .	61
13.9 Linear-Nonlinear-Poisson (LNP) model . . . . .	61
13.9.1 Fitting a LNP model to data . . . . .	62
<b>14 Single neuron coding II: Information theory</b>	<b>63</b>
14.1 Setup . . . . .	63
14.2 Entropy . . . . .	63
14.2.1 Joint entropy . . . . .	63
14.2.2 Conditional Entropy . . . . .	63
14.3 Mutual Information . . . . .	64
14.3.1 Relationship between Entropy and Mutual Information . . . . .	64
14.3.2 Mutual Information for continuous distributions . . . . .	64
14.4 Examples . . . . .	64
14.5 Efficient coding hypothesis . . . . .	66
14.6 What to optimize? . . . . .	66
14.7 Optimizing the static non-linearity . . . . .	66
14.7.1 Histogram Equalization in the fly visual system . . . . .	66
14.8 Optimizing the temporal filter: Whitening . . . . .	67
14.8.1 Optimal receptive fields with natural images . . . . .	67

<b>15 Population coding and decoding</b>	<b>68</b>
15.1 Information measures . . . . .	68
15.2 Decoding - Estimators . . . . .	68
15.2.1 Examples . . . . .	68
15.3 Error of the estimator - bias, variance . . . . .	69
15.3.1 Cramer-Rao . . . . .	69

# 1 Introduction

Neurons are remarkable among the cells of the body in their ability to propagate signals rapidly over large distances. They do this by generating characteristic electrical pulses called action potentials or, more simply, spikes that can travel down nerve fibers. Neurons represent and transmit information by firing sequences of spikes in various temporal patterns.

The link between stimulus and response can be studied from two opposite points of view. **Neural encoding** refers to the map from stimulus to response. For example, we can catalog how neurons respond to a wide variety of stimuli, and then construct models that attempt to predict responses to other stimuli. **Neural decoding** refers to the reverse map, from response to stimulus, and the challenge is to reconstruct a stimulus, or certain aspects of that stimulus, from the spike sequences it evokes.

## 1.1 Properties of Neurons

Neurons are highly specialized for generating electrical signals in response to chemical and other inputs, and transmitting them to other cells. Some important morphological specializations are the **dendrites** that receive inputs from other neurons and the axon that carries the neuronal output to other cells.

Axons from single neurons can traverse large fractions of the brain or, in some cases, of the entire body.

Along with these morphological features, neurons have physiological specializations. Most prominent among these are a wide variety of **membrane-spanning ion channels** that allow ions, predominantly sodium ( $\text{Na}^+$ ), potassium ( $\text{K}^+$ ), calcium ( $\text{Ca}^{2+}$ ), and chloride ( $\text{Cl}^-$ ), to move into and out of the cell.

The electrical signal of relevance to the nervous system is the **difference in electrical potential** between the interior of a neuron and the surrounding extracellular medium.

Under resting conditions, the potential inside the cell membrane of a neuron is about -70 mV relative to that of the surrounding bath (which is conventionally defined to be 0 mV), and the cell is said to be polarized.

Ions flow into and out of a cell due to both voltage and concentration gradients. Current in the form of positively charged ions flowing out of the cell (or negatively charged ions flowing into the cell) through open channels makes the membrane potential more negative, a process called **hyperpolarization**. Current flowing into the cell changes the membrane potential to less negative or even positive values. This is called **depolarization**.

If a neuron is depolarized sufficiently to raise the membrane potential above a threshold level, a positive feedback process is initiated, and the neuron generates an **action potential**.

For a few milliseconds just after an action potential has been fired, it may be virtually impossible to initiate another spike. This is called the absolute refractory period. For a longer interval known as the relative refractory period, lasting up to tens of milliseconds after a spike, it is more difficult to evoke an action potential.

Action potentials are of great importance because they are the only form of membrane potential fluctuation that can propagate over large distances. Subthreshold potential fluctuations are severely attenuated over distances of 1 mm or less.

Axons terminate at synapses where the voltage transient of the action potential opens ion channels. The neurotransmitter binds to receptors at the signal-receiving or postsynaptic side of the synapse, causing ion-conducting channels to open. Depending on the nature of the ion flow, the synapses can have either an excitatory, depolarizing, or an inhibitory, typically hyperpolarizing, effect on the postsynaptic neuron.

## 1.2 Recording Neuronal responses

Membrane potentials are measured intracellularly by connecting a hollow glass electrode filled with a conducting electrolyte to a neuron, and comparing the potential it records with that of a reference electrode placed in the extracellular medium. Intracellular recordings are made either with sharp electrodes inserted through the membrane into the cell, or patch electrodes that have broader tips and are sealed tightly to the surface of the membrane.

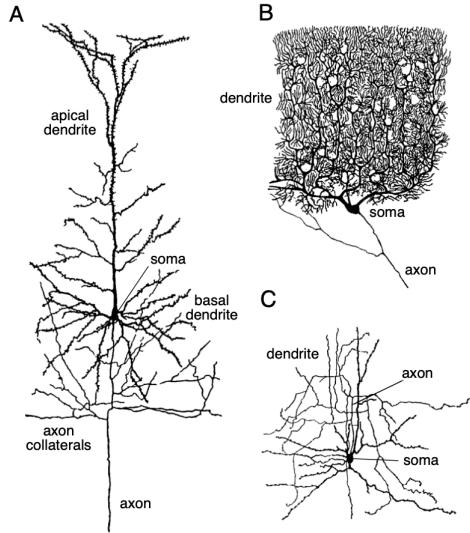


Figure 1: Diagrams of three neurons. (A) A cortical pyramidal cell. (B) A Purkinje cell of the cerebellum. (C) A stellate cell of the cerebral cortex.

### 1.3 Neuronal membrane

The neuronal membrane is rich in ion channels, which are selective to specific neurons.

When  $Na^+$  channels open,  $Na^+$  ions flow inside the cell, depolarizing it (inward current);

When  $K^+$  channels open,  $K^+$  ions flow outside the cell, hyperpolarizing it (outward current).

Ion pumps work continuously to maintain certain values in the membrane potential (difference of potential outside and inside the cell).

#### 1.3.1 Movement of ions through the membrane

When a channel is open, movement of ions is determined by two forces:

1. The **concentration gradient** tends to make  $K^+$  ions diffuse out of the cell. The diffusive flux  $J_{diff}$  is related to the concentration gradient  $\partial[C]/\partial x$  through Fick's law of diffusion

$$J_{diff} = -D \frac{\partial[C]}{\partial x} \quad (1)$$

where  $D$  is the diffusion constant

2. The difference in charge between the outside and inside of the cell leads to an **electrical field**, driving  $K^+$  ions back inside the cell. Electrical drift  $J_{drift}$  is related to  $[C]$  and voltage gradient through microscopic version of Ohm's law

$$J_{drift} = -\mu z [C] \frac{\partial V}{\partial x} \quad (2)$$

where  $\mu$  is the ion's mobility and  $z$  is ion's valence.

#### 1.3.2 Nernst-Planck Equation

The Nernst-Planck equation gives an expression for the reversal potential of a single type of neuron whose concentration is  $[C]$ .

The **Total flux** of ions is given by the sum of the two contributions given by equations 1 and 2. In particular, rewriting  $D$  as  $k_B T \mu / q$ , we have

$$J_{TOT} = -\frac{k_B T \mu}{q} \frac{\partial[C]}{\partial x} - \mu z [C] \frac{\partial V}{\partial x}$$

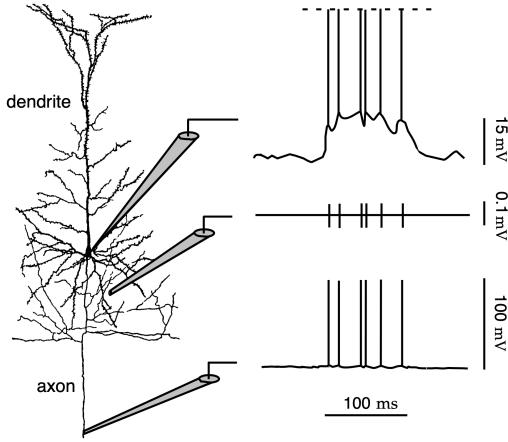


Figure 2: Three simulated recordings from a neuron. The top trace represents a recording from an intracellular electrode connected to the soma of the neuron. The bottom trace represents a recording from an intracellular electrode connected to the axon some distance away from the soma. The full height of the action potentials is indicated in this trace. The middle trace is a simulated extracellular recording. Action potentials appear as roughly equal positive and negative potential fluctuations with an amplitude of around 0.1 mV. This is roughly 1000 times smaller than the approximately 0.1 V amplitude of an intracellularly recorded action potential.

where  $k_B = 1.38 \cdot 10^{-23} J/K$  is the Boltzmann constant and  $q$  is the charge.

In particular, we are interested in the current flowing through the channels. This is given by multiplying the total flux by valence  $z$  and charge  $q$ . Hence, the **Nernst-Planck equation** gives

$$I = -uzRT \frac{\partial[C]}{\partial x} - uz^2 F[C] \frac{\partial V}{\partial x}$$

where  $u = \frac{\mu}{N_A}$  is the molar mobility,  $R = k_B N_A$  is the gas constant,  $F = q N_A$  is Faraday constant and  $N_A$  is Avogadro's number.

At equilibrium  $I = 0$  (i.e.,  $-uzRT \frac{\partial[C]}{\partial x} = uz^2 F[C] \frac{\partial V}{\partial x}$ ) we can derive a relation between the reversal potential and the ratio of the concentrations for a single ion type:

$$V_r := V_{in} - V_{out} = -\frac{RT}{zF} \log \frac{[C]_{in}}{[C]_{out}}$$

### 1.3.3 Goldman-Hodgkin-Katz equation

For the case of multiple ionic species, the Nernst equation must be replaced by the Goldman-Hodgkin-Katz equation

$$V_{eq} = \frac{RT}{F} \log \frac{P_K[K^+]_{out} + P_{Na}[Na^+]_{out} + P_{Cl}[Cl^-]_{in}}{P_K[K^+]_{in} + P_{Na}[Na^+]_{in} + P_{Cl}[Cl^-]_{out}}$$

where  $P_i \propto u_i s$  are the permeabilities of ionic species, for  $i = K, Na, Cl$ .

Plugging numbers in gives us a potential in equilibrium of about  $-70mV$ .

## 2 Model Neurons: Neuroelectronics

### 2.1 Electrical Properties of Neurons

The **cell membrane** is a lipid bilayer that is essentially impermeable to most charged molecules. This insulating feature causes the cell membrane to act as a **capacitor**.

A very useful way to describe the behavior of the membrane potential is in terms of electrical circuits. The circuit consists of three components:

1. resistors, representing the ion channels;
2. batteries, representing the concentration gradients of the ions;
3. capacitors, representing the ability of the membrane to store charge.

Consider first the case of a membrane that is only permeable to Potassium, as shown in figure 3.

Capacitors store charge and then release it in the form of currents. The relationship between the charge stored and the potential is given by  $Q = CV_M$ : the total charge  $Q$  is proportional to the membrane potential  $V_M$ , with a proportionality constant  $C$  called the membrane capacitance.

Therefore, the capacitance component of the current is given by

$$I_{cap} = \frac{dQ}{dt} = C \frac{dV_M}{dt}$$

In the equivalent circuit,  $K^+$  channels are represented as a conductor in series with a battery. Hence, if  $g_K$  is the conductance of a  $K^+$  channel ( $g_K = \frac{1}{R}$  is the inverse of the resistance), then the ionic channel through the gate is given by Ohm's law:

$$I_K = g_K(V_M - E_K) \quad (3)$$

where  $E_K$  is the potential generated by the potassium battery (i.e.,  $E_K$  is the Potassium reversal potential). Hence, if an input current  $I(t) = I$  is injected into the cell, it may add further charge on the capacitor, or leak through the channels in the cell membrane.

Therefore,  $I$  will have two components, given by

$$\begin{aligned} I_{cap} &= \frac{dQ}{dt} = C \frac{dV_M}{dt} \\ I_K &= \frac{V_M - E_K}{R} \end{aligned}$$

so that the total injected current is given by

$$I = C \frac{dV_M}{dt} + g_K(V_M - E_K) \quad (4)$$

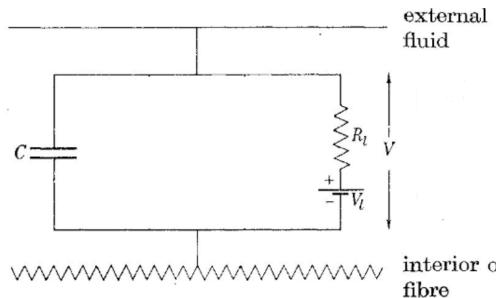


Figure 3: Electric circuit representing the structure of a single ion channel type in a neuron

Now

1. If  $I = 0$ , then  $V_M = E_K$ .
2. If  $I(t) = I \mathbb{1}_{t \geq 0}$ , then let  $\frac{1}{R} = g$  and let  $\tau = \frac{C}{g_K}$ .  
Then, equation 4 becomes

$$\begin{cases} \tau \frac{dV_M}{dt} = -(V_M - E_K) + \frac{I}{g_K} \\ V_M(-\infty) = 0 \end{cases}$$

A first guess of the solution is

$$V_M(t) = Ae^{-Bt} + C$$

where  $A$ ,  $B$  and  $C$  are to be determined.

We have

$$\tau A(-B)e^{-Bt} = -(Ae^{-Bt} + C - E_K) + \frac{I}{g_K}$$

so that we expect to have

$$\begin{aligned} \tau A(-B) &= -A \implies B = \frac{1}{\tau} \\ C - E_K &= \frac{I}{g_K} \end{aligned}$$

For  $t < 0$ , we have  $C = E_K$  and

$$V_M(t < 0) = Ae^{-\frac{t}{\tau}} + E_K$$

Now using the boundary condition  $V_M(-\infty) = 0$  we have  $A = 0$  so that

$$V_M(t < 0) = E_K$$

For  $t > 0$  we have

$$C \frac{dV_M}{dt} = -g(V_M - E_K) + I$$

that is, an exponential relaxation to a new equilibrium  $I \cdot R$ .

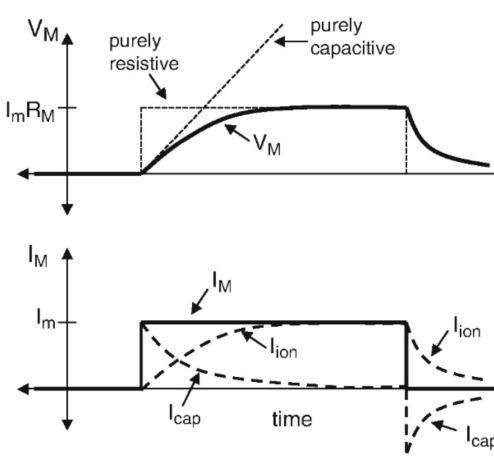


Figure 4: The graph of the membrane potential  $V_M$  as a function of the injected current  $I_M$ . Before any current is injected, the membrane potential is constant, as we derived, then we observe an exponential relaxation to a constant value. If the system was purely capacitive (so  $g_K = 0$ ),  $V_M$  would increase linearly, while if we had  $C \sim 0$  we would observe a step response.

So far we have described below-threshold behavior of the membrane potential; we have not yet discussed spiking behavior. This is the subject of the next section.

### 3 Hodgkin-Huxley model

#### 3.1 Equivalent Circuit

In the Hodgkin-Huxley model each channel is viewed as a transmembrane protein that forms a pore through which ions can diffuse down their concentration gradients.

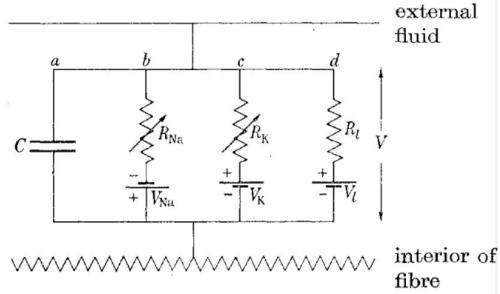


Figure 5: Equivalent circuit underlying the the Hodgkin-Huxley model

Here we assume that there are voltage-gated  $K^+$  and  $Na^+$  channels and a leak current  $I_L$ , which corresponds to passive flow of ions through non-gated channels. Consider now the circuit in figure 5. Then we have

$$\begin{cases} C \frac{dV_M}{dt} = -I_L(V_M) - I_{Na}(V_M) - I_K(V_M) \\ I_L(V_M) = g_L(V_M - E_L) \\ I_{Na}(V_M) = g_{Na}(V_M - E_{Na}) \\ I_K(V_M) = g_K(V_M - E_K) \end{cases}$$

Here the leak conductance  $g_L$  is constant, while  $g_{Na}$  and  $g_K$  may change in time, since these correspond to the opening and closing of  $Na^+$  and  $K^+$  ion channels, respectively. In particular,  $g_{Na}$  and  $g_K$  change during an action potential.

The basic **mechanisms underlying action potentials** are the following.

At rest, most of the sodium channels are closed so the membrane potential is determined primarily by the  $K^+$  Nernst potential.

If the cell is depolarized above some threshold by an input current, then sodium channels open and this further depolarizes the cell. This allows even more sodium channels to open, allowing for more sodium ions to enter the cell and forcing the cell towards the sodium Nernst potential. This is the up-stroke of the action potential. The sodium channel is transient so that even when depolarized, the  $Na^+$  channels eventually shut down. In the meantime, the depolarization opens potassium channels and potassium ions exit the cell. This hyperpolarizes the cell as the membrane potential moves toward the potassium equilibrium potential. Until the voltage-gated potassium channels close up again, the membrane is refractory.

#### 3.2 Voltage-gated Channels

The probability that a gate is open or closed depends on the membrane potential. If we let  $m$  be the fraction of open gates ( $1 - m$  is the fraction of closed gates), then  $m$  is also interpreted as the probability that a gate is open and is found by the kinetic equation

$$\frac{dm}{dt} = \alpha(V)(1 - m) - \beta(V)m$$

### 3.3 The model

Hodgkin and Huxley used two experimental methods in order to separate the ionic currents and compute how the  $K^+$  and  $Na^+$  conductances depend on voltage.

The first was a simple feedback circuit called the **voltage-clamp** that allows the experimenter to hold the membrane potential at a constant level  $V_M = V_C$ .

The voltage clamp does so by injecting a current into the axon that is equal and opposite to the current flowing through the voltage-gated channels.

**Voltage-clamp separates the total membrane current in its ionic and capacitance components.** Given that  $I_{cap} = C\partial V_M / \partial t = 0$  (as  $V_M$  is constant), then any changes in current must be due either to the leak current or to the opening and closing of voltage-gated membrane channels. This allows to determine the leak conductance  $g_L$ .

Moreover, Hodgkin and Huxley were able to determine how  $I_K$  depends on the membrane potential, by removing  $Na^+$  ions and isolating the  $K^+$  current.

Now, if  $I_K$  and  $I_L$  are known, then one computes  $I_{Na}$  simply by subtracting  $I_K$  and  $I_L$  from  $I_M$ .

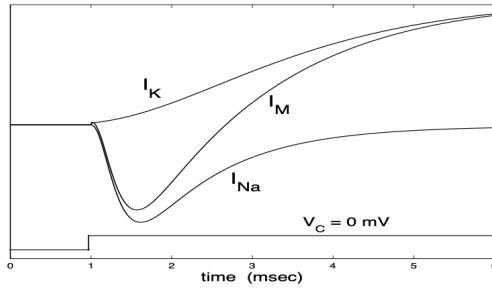


Figure 6: Numerically computed voltage-clamp experiment. The membrane potential is stepped from rest to 0 mV. This results in an inward current followed by an outward current. The separate potassium and sodium currents are also shown.

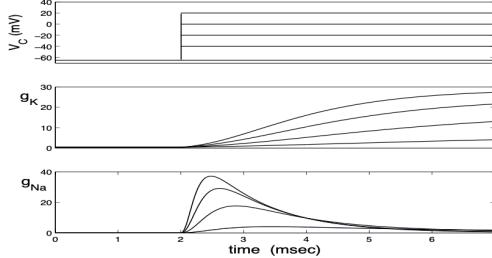


Figure 7: Numerically computed voltage-clamp experiment. The membrane potential is stepped to different values and the resulting potassium and sodium conductances are computed.

Using the voltage-clamp data, Hodgkin and Huxley derived expressions for the  $K^+$  and  $Na^+$  conductances. They proposed that

$$g_K = \bar{g}_K n^4 \quad g_{Na} = \bar{g}_{Na} m^3 h$$

where  $\bar{g}_K$  and  $\bar{g}_{Na}$  are maximum conductances and  $n$ ,  $m$  and  $h$  are gating variables that take values between 0 and 1. Hence,  $n^4$  represents the probability that a potassium channel is open.

The probability that the sodium **activation** gate is open is  $m^3$  and the probability that the sodium **inactivation** gate is open is  $1 - h$ .

Each of the gating variables satisfies a first order differential equation of the form

$$\begin{aligned}\frac{dn}{dt} &= \alpha_n(V)(1 - n) - \beta_n(V)n \\ \frac{dm}{dt} &= \alpha_m(V)(1 - m) - \beta_m(V)m \\ \frac{dh}{dt} &= \alpha_h(V)(1 - h) - \beta_h(V)h\end{aligned}$$

These equations can be rewritten in another useful way, by diving both sides by  $\alpha_i(V) + \beta_i(V)$ :

$$\begin{aligned}\tau_n \frac{dn}{dt} &= -n + n_\infty(V) \\ \tau_m \frac{dm}{dt} &= -m + m_\infty(V) \\ \tau_h \frac{dh}{dt} &= -h + h_\infty(V)\end{aligned}$$

where

$$\tau_i(V) = \frac{1}{\alpha_i(V) + \beta_i(V)}, \quad i_\infty(V) = \frac{\alpha_i(V)}{\alpha_i(V) + \beta_i(V)}$$

In Figure 8, we plot the activation curves  $n_\infty(V)$ ,  $m_\infty(V)$  and  $h_\infty(V)$  along with  $\tau_n(V)$ ,  $\tau_m(V)$  and  $\tau_h(V)$ . Note that  $n_\infty$  and  $m_\infty$  are increasing functions that approach 0 for hyperpolarizing currents and approach 1 for depolarizing currents. Hence,  $n$  and  $m$  become activated when the membrane is depolarized. On the other hand,  $h_\infty(V)$  is a decreasing function, so the sodium channels inactivate when the membrane is depolarized. It is also important to note that  $\tau_m(V)$  is considerably smaller than  $\tau_n$  or  $\tau_h$ . Hence, sodium channels activate much faster than they inactivate or potassium channels open.

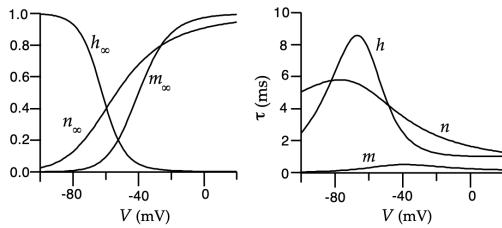


Figure 8: The voltage-dependent functions of the Hodgkin-Huxley model. The left panel shows  $m_\infty(V)$ ,  $h_\infty(V)$ , and  $n_\infty(V)$ , the steady-state levels of activation and inactivation of the  $Na^+$  conductance, and activation of the  $K^+$  conductance. The right panel shows the voltage-dependent time constants that control the rates at which these steady-state levels are approached for the three gating variables.

### 3.4 Action Potential revisited

In summary, the Hodgkin-Huxley model is a system of four differential equations; there is one equation for the membrane potential and three equations for channel gating variables.

Figure 9 shows solutions of these equations in response to different levels of steps in currents. Note that there is **“all-or-none” behavior**: When the applied current is below some threshold, the membrane potential returns quickly to rest; when the current is above some threshold, there is an action potential. If the applied current is sufficiently large and held for a sufficiently long time, then the model generates a periodic response.

Here we give a more “mathematical” explanation in terms of the behavior of the dependent variables in the differential equations.

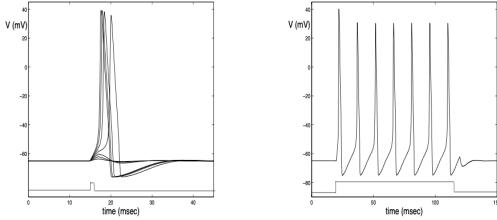


Figure 9: Responses of the HH model to applied currents. Left: transient responses showing “all-or-none” behavior; Right: Sustained periodic response.

- When we depolarize the cell, we change the values of the activation curves:  $n_\infty(V)$  and  $m_\infty(V)$  increase, while  $h_\infty(V)$  decreases. Hence, **potassium channels open**, while sodium channels both activate and inactivate.
- However,  $\tau_m$  is much smaller than both  $\tau_h$  and  $\tau_n$ . It follows that the  $Na^+$  channels activate much faster than they inactivate or  $K^+$  channels open. Therefore, the  $Na^+$  conductance,  $g_{Na} = \bar{g}_{Na}m^3$  increases faster than  $g_K = \bar{g}_n^4$ , leading to an **increase in the  $Na^+$  current**,  $I_{Na} = g_{Na}(V - E_{Na})$ .
- As long as the cell is near rest, the driving force  $V - E_{Na}$  is large. Hence, the sodium current will dominate the equation for the membrane potential and  $V$  **will increase towards the  $Na^+$  Nernst potential**. As  $V$  increases,  $m_\infty(V)$  increases further, leading to further increase in  $Na^+$  activation.
- As  $V$  increases towards  $E_{Na}$ , sodium channels inactivate. This is because  $h \rightarrow h_\infty(V) \simeq 0$ . Moreover, the sodium driving force  $V - E_{Na}$  decreases. For both reasons, the  $Na^+$  **current turns off**.
- Meanwhile, the potassium channel activates because  $n \rightarrow n_\infty(V) \simeq 0$ . Moreover, the  $K^+$  driving force  $V - E_K$  becomes very large. It follows that eventually, **the potassium current dominates and the membrane potential must fall back towards the  $K^+$  Nernst potential**. This corresponds to the **down-stroke** of the action potential.
- After the action potential, the cell is hyperpolarized with  $m_\infty \simeq 0$ ,  $n_\infty \simeq 0$  and  $h_\infty \simeq 1$ . After some time,  $m$ ,  $n$  and  $h$  approach their steady state values and the cell returns to rest.

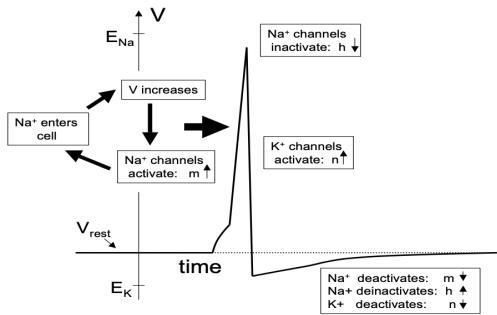


Figure 10: Mechanisms underlying the action potential.

## 4 Generalization of the Hodgkin-Huxley model

### 4.1 Diversity of electrophysiological behaviors of single neurons

There is a large diversity of neuron firing behaviors:

- Firing rate adaptation
- Bursting
- Delayed firing
- Stuttering
- Irregular firing

Importantly, these behaviors are not modelled by the HH model.

### 4.2 Diversity of Voltage-gated channels

The previously discussed version of the HH model contains only  $K^+$  and  $Na$  channels. However, there exist many other types of ionic channels that can still be modelled by Hodgkin-Huxley as follows

$$\begin{cases} I_x = g_x m_x^p h_x^q (V - V_x) \\ \tau_{m_x}(V) = \frac{dm_x}{dt} = -m_x + m_{x\infty}(V) \quad (\text{activation}) \\ \tau_{h_x}(V) = \frac{dh_x}{dt} = -h_x + h_{x\infty}(V) \quad (\text{inactivation}) \end{cases}$$

where  $g_x$  is the maximal conductance,  $m_x$  is the activation variable,  $h_x$  is the inactivation variable and  $V_x$  is the reversal potential.

**Functional classification of voltage-gated currents**, see figure 11.

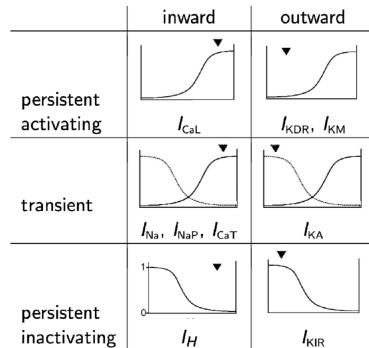


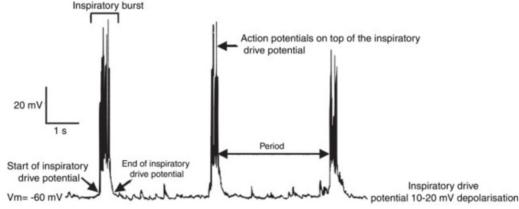
Figure 11: Depending on how the voltage-gated current changes with the potential  $V$ , it can be classified in persistent (in)activating or transient. These graphs show the activation and/or inactivation variables as a function of the membrane potential.

### 4.3 Persistent sodium: bursting

Example: neurons in the pre-Botzinger complex (brain structure controlling breathing).

This behavior cannot be modelled by the HH model, which only describes single spikes or multiple, regular ones in response to a constant input. In order to model this kind of phenomenon, we need to modify the HH model appropriately. One way to do this is to introduce an additional current (slowly varying sodium current).

$$\begin{cases} C \frac{dV_M}{dt} = -g_L(V_M - E_L) - I_{Na}(V_M) - I_K(V_M) - I_{NaP} \\ I_{NaP} = g_{NaP} m_{NaP\infty}(V_M) h_{NaP}(V_M - E_{Na}) \\ \tau_{h_{NaP}} \frac{dh_{NaP}}{dt} = -h_{NaP} + h_{NaP\infty}(V_M) \end{cases}$$



- We assume that the activation variable  $h_{NaP}$  is much slower than all the other variables.
- We assume that the activation for the  $m$  variable happens for a lower membrane potential, below the activation of the regular sodium channels.
- For an input that in the standard HH model would not produce spiking, as  $V_M$  raises, the sodium channels start to open, producing an additional input.
- This leads to an increase in the value of  $V_M$  which in turn makes the other channels (also present in HH) open, producing spiking.
- Given that the  $NaP$  current evolves (increases) slowly, we have multiple spikes. Each spike is ended by the inactivation gates, which leads to relaxation. The dynamics then repeats itself.

#### 4.4 M current: spike frequency adaptation

The M-current is a slow, low-threshold, outward potassium current which is responsible for a phenomenon known as spike frequency adaptation (SFA). It gradually reduces the firing rate of a neuron which has been depolarized sufficiently to cause repetitive firing.

The M-current and related slow potassium currents are able to stop neurons from firing if they are strong enough and thus can provide an effective brake to run-away excitation in networks.

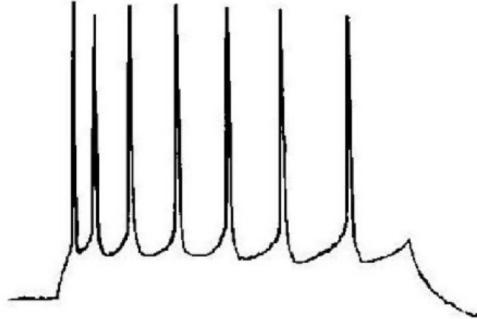


Figure 12: Example of a neuron (Cortical pyramidal cells) that, in response to a constant injected current, initially triggers many spikes which are initially close to each other and later keep taking more time to reproduce. The system adapts less in response to the same stimulus

Our goal is to model this phenomenon using the same ideas as in the previous section, this time using the potassium "M" current.

$$\begin{cases} C \frac{dV_M}{dt} = -g_L(V_M - E_L) - I_{Na}(V_M) - I_{KDR}(V_M) - I_{KM}(V_M) + I_{app} \\ I_{KM} = g_{KM}m_{KM}(V_M - E_K) \\ \tau_{m_{KM}} \frac{dm_{KM}}{dt} = -m_{KM} + m_{KM\infty}(V_M) \end{cases}$$

The potassium  $M$  current has no inactivation gates.

Assuming that the applied current is large enough to produce spiking, the effect of the  $M$  current is to produce an additional flux of Potassium, which produces an out-flux of positive charges, leading to a decreased membrane

potential  $V_M$ . This decreases the amount of current (total input) entering into the cell. Hence, the cell tends not to spike (or at least to spike less and less often). That's why we observe a gradually less intense and less concentrated spiking.

## 4.5 Single Compartment models // HH-type to Integrate-and-fire

Models that describe the membrane potential of a neuron by a single variable  $V$  are called single-compartment models, while multi-compartment models usually describe spatial variations in the membrane potential.

- Hodgkin-Huxley-type models
  - Good at capturing quantitatively diversity of dynamical features of real neurons
  - Highly non-linear, large number of variables  $\Rightarrow$  hard to analyze mathematically
  - Network simulations are computationally expensive
- Leaky integrate-and-fire (LIF) model
  - Too simple to reproduce diversity of dynamical features of real neurons
  - Can be analyzed mathematically (at both neuron and network levels)
  - Permits simulations of very large networks of such neurons
- Models in between HH and LIF: Best of both worlds?
  - Mathematically tractable
  - Capture dynamical behaviors of HH-type models (and real neurons)

Integrate-and-fire models provide simplified neuron models by stipulating that an action potential occurs whenever the membrane potential of the model neuron reaches a threshold value  $V_{th}$ . After the action potential, the potential is reset to a value  $V_{reset}$  below the threshold potential,  $V_{reset} < V_{th}$ .

The basic integrate-and-fire model was proposed by Lapicque in 1907, long before the mechanisms that generate action potentials were understood.

By avoiding a biophysical description of the action potential, integrate-and-fire models are left with the simpler task of modeling only **subthreshold membrane potential dynamics**. This can be done with various levels of rigor. In the simplest version of these models, all active membrane conductances are ignored, including, for the moment, synaptic inputs, and the entire membrane conductance is modeled as a single passive leakage term,  $I_m = \bar{g}_L(V - E_L)$ . This version is called the passive or **leaky integrate-and-fire model**.

With these approximations, the model neuron behaves like an electric circuit consisting of a resistor and a capacitor in parallel, and the membrane potential is given by

$$C_m \frac{dV}{dt} = -\bar{g}_L(V - E_L) + I_e$$

where  $I_e$  is the current injected through the electrode in an experimental setting. Multiplying this by the membrane resistance  $R_m$ , we obtain

$$\tau_m \frac{dV}{dt} = E_L - V + R_m I_e \quad (5)$$

To generate action potentials in the model, equation 5 is augmented by the rule that whenever  $V$  reaches the threshold value  $V_{th}$ , an action potential is fired and the potential is reset to  $V_{reset}$ . Equation 5 indicates that when  $I_e = 0$ , the membrane potential relaxes exponentially with time constant  $\tau_m$  to  $V = E_L$ . Thus,  $E_L$  is the resting potential of the model cell. It is possible to introduce a **refractory period** in the model by imposing that, once the membrane potential reaches the resting value, it remains there for a certain amount of time.

This model allows to gain access to some statistics such as the time needed to produce a spike and how many spikes we have in a certain time period, and so on.

**Application:** What is the frequency of activation of a cell in response to a given current?

The firing rate of an integrate-and-fire model in response to a constant injected current can be computed analytically.

To compute the inter-spike time interval  $T$ , we first observe that  $V(t)$  can be computed directly from equation 5 as follows

$$V(t) = E_L + R_m I_e + (V(0) - E_L - R_m I_e) e^{t/\tau_m}$$

Suppose that at  $t = 0$ , the neuron has just fired an action potential and is thus at the reset potential, so that  $V(0) = V_{reset}$ . The next action potential will occur when the membrane potential reaches the threshold, that is, at a time  $t = t^*$  when

$$V(t^*) = V_{th} = E_L + R_m I_e + (V_{reset} - E_L - R_m I_e) e^{t^*/\tau_m}$$

By solving this for  $t^*$ , the time of the next action potential, we can determine the interspike interval for constant  $I_e$ , or equivalently its inverse, which we call the interspike-interval firing rate of the neuron,  $r^*$

$$r^* = \frac{1}{t^*} = \left( \tau_m \ln \left( \frac{E_L + R_m I_e - V_{reset}}{E_L + R_m I_e - V_{th}} \right) \right)^{-1}$$

This expression is valid if  $R_m I_e > V_{th} - E_L$ ; otherwise  $r^* = 0$ , because if the input is below threshold there is no spiking and no response.

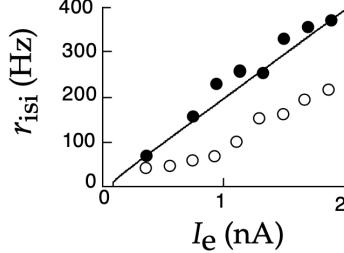


Figure 13: Comparison of interspike-interval firing rates as a function of injected current for an integrate-and-fire model and a cortical neuron measure in vivo. The line gives  $r^*$

## 4.6 Dynamics of Spike Initiation: HH vs LIF

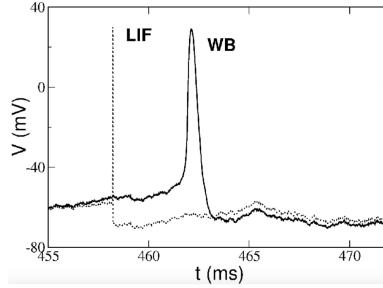


Figure 14: Comparison of Voltage predicted by the LIF and HH model (Wang and Buzsaki)

Before the spiking the potential predicted by the two models match.

The LIF model in equation 5 does not capture spike initiation properly (in this case it is too fast). One way to handle this problem is to introduce a non-linear function of  $V$ ,  $\psi(V)$ , that tries to produce a better description.

### 4.6.1 Quadratic Integrate-and-Fire

$$\begin{cases} C_m \frac{dV}{dt} = -\bar{g}_L(V - E_L) + \psi(V) + I_e \\ \psi(V) = \frac{g_L}{2\Delta_{th}}(V - V_{th})^2 + g_L(V - E_L) - I_{th} \end{cases} \quad (6)$$

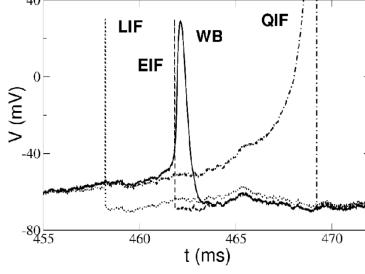


Figure 15: Comparison of Voltage values predicted by the HH, LIF and QIF models

The QIF model does not capture the spike initiation well, as it is too slow and predicts an infinite value in the potential.

Denoting the right-hand side of equation 6 by  $F(V) = -\bar{g}_L(V - E_L) + \psi(V) + I_e$ , we obtain  $F(V) = \frac{g_L}{2\Delta_{th}}(V - V_{th})^2 + I_e$ . In particular,  $F(V)$  is a parabola that has two zeros for  $I_e = 0$ . For  $I_e > 0$  the parabola moves up. When the parabola has no intercepts,  $F(V) > 0$  for every  $V$ , so that the value of  $V$  will keep increasing, as  $\frac{dV}{dt} > 0$  at all points. That's when we observe a divergence to  $+\infty$ .

For a specific input value  $I_{th}$ ,  $F(V)$  will have a single zero-point  $V^* = V_{th}$ . For all inputs  $I > I_{th}$  the RHS of equation 6 is always positive. The value  $I = I_{th}$  is the input value for which a spike generation starts (i.e., for every  $I > I_{th}$  there will be a spike).

The only parameter of this model is given by  $V_{th}$ .

#### 4.6.2 Exponential Integrate-and-Fire

$$\begin{cases} C_m \frac{dV}{dt} = -\bar{g}_L(V - E_L) + \psi(V) + I_e \\ \psi(V) = g_L \Delta_{th} \exp\left(\frac{V - V_{th}}{\Delta_{th}}\right) + g_L(V - E_L) - I_{th} \end{cases} \quad (7)$$

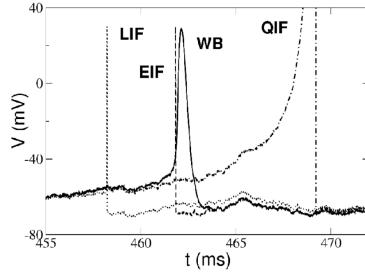


Figure 16: Comparison of Voltage values predicted by the HH, LIF, QIF and EIF models

The EIF model quantitatively captures very well the dynamics of the WB neuron, as shown in figure 16. The reason for this is given by the fact that the  $m_\infty$  curve of sodium can be fitted by an exponential curve, when close to firing threshold.

## 4.7 Adaptive nonlinear Integrate-and-Fire models

We now add a new variable  $w$  coupled with voltage:

$$\begin{cases} C_m \frac{dV}{dt} = -\bar{g}_L(V - E_L) + \psi(V) - w + I_e \\ \tau_w \frac{dw}{dt} = a(V - E_L) - w \end{cases} \quad (8)$$

After a spike occurs, a reset takes place:

$$\begin{cases} V \rightarrow V_R \\ w \rightarrow w + b \end{cases}$$

If there is no input (i.e.  $I_e = 0$ ) then  $w \rightarrow 0$ , because the voltage goes to its stationary value (which depends on the model used).

Depending on the input  $I_e$ , there will be a change in the membrane potential, triggering spikes according to the model considered. In turn, this will affect  $w$ , which will mainly change when at spike time, as shown in figure 17.

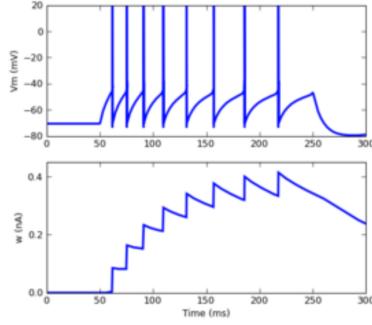


Figure 17:

The parameter  $w$  moves down the curve of  $F(V)$ , in contrast to  $I_e$ , which moves it upwards. This can lead to spiking, bursting and adaptation.

Adaptation because, each time that a spike is emitted  $w$  grows and the curve of  $F(V)$  moves downwards, which means that the amplitude of  $\frac{dV}{dt}$  is smaller, and therefore the increase of the membrane potential is smaller, so that it will take more time to reach a spike, as shown in figure 17. When  $w$  is large enough  $F(V)$  will have more than a zero (e.g. for QIF) and no spike will take place anymore.

## 5 Response of neurons to stochastic inputs

### 5.1 Fluctuations: in vivo, in vitro

With no synaptic inputs, in vitro there are small fluctuations ( $< 1mV$ ) of membrane potentials due to channel noise.

In vivo instead we observe large fluctuations, even in the absence of sensory stimulation. In this case, neurons fire at low rates, in a highly irregular fashion.

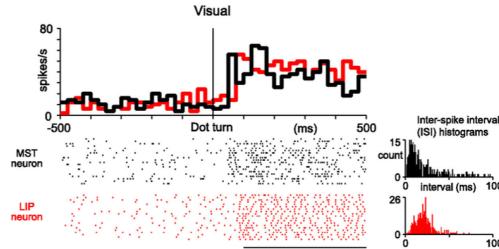


Figure 18: Graph of the number of neuronal spikes per unit of time in an animal first subject to no stimulus and then exposed to some visual stimulus. The neuronal activity is irregular regardless of a stimulus being present or not. Neurons do not fire regularly. The two histograms on the right show the inter-spike intervals: when a stimulus is present these intervals tend to be significantly shorter.

### 5.2 Statistics of neuronal firing

Starting from the experiment described in figure 18, we aim at describing the statistics of neuronal spikes. Given the distribution of the inter-spike intervals, we define the coefficient of their variation as the standard deviation divided by the mean,

$$CV = SD/\text{mean}$$

With regular spiking (as in the case of single-cell experiment) the CV coefficient will be close to 0.

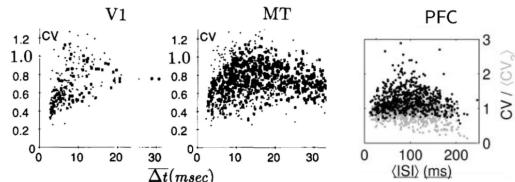


Figure 19: Plots of the CV coefficient of neurons from different parts of the brain (Visual Cortex, Higher Order Visual Area, Pre-Frontal Cortex). The x axis contains the length of the considered time intervals when measuring the number of spikes.

### 5.3 Modeling spike trains as point processes

We consider a point process to be collection of points randomly located on some underlying space, here time. We now introduce a phenomenological model that describes neuronal spike statistics.

At each point (spike time  $t_i$ ), we generate the next spike time  $t_{i+1} = t_i + T_i$  using a probability distribution  $P(T|\text{spike train history})$ .

Then, each inter-spike interval is drawn independently from previous intervals, from a p.d.f.  $p(T)$ . here are some examples of inter-spike interval distributions used in neuroscience:

- Exponential (Poisson Process)
- Gamma
- Inverse Gaussian

**Example 1** (Poisson). Let  $P(T) = \nu e^{-\nu T}$ .

Then, we define the mean firing rate to be the inverse of the average inter-spike interval, so that the mean is given by

$$\frac{1}{\langle T \rangle} \quad \text{with} \quad \langle T \rangle = \int_{-\infty}^{+\infty} T \cdot P(T) dT = \frac{1}{\nu}$$

Analogously, the standard deviation is given by

$$SD(T) = \sqrt{\langle T^2 \rangle - \langle T \rangle^2} = \frac{1}{\nu}$$

Hence, the variation coefficient is given by

$$CV = \frac{SD(T)}{\langle T \rangle} = 1$$

Moreover, the number of spikes in an interval of length  $T$  is given by the Poisson distribution,

$$P(k) = \frac{(\nu T)^k e^{-\nu T}}{k!}$$

As a side note, the number of spikes in disjoint intervals is independent.

**Example 2** (Gamma). The pdf is given by

$$P(T) = \frac{(k\nu)^k}{\Gamma(k)} T^{k-1} e^{-k\nu T}$$

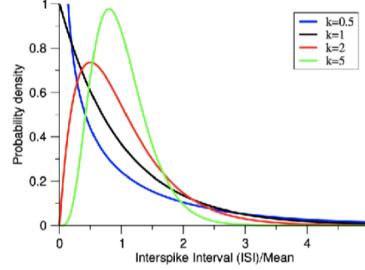


Figure 20: Gamma distribution as  $k$  varies

Mean is

$$\frac{1}{\langle T \rangle} = \nu$$

so that

$$CV = \frac{SD(T)}{\langle T \rangle} = \frac{1}{\sqrt{k}}$$

We observe that any value of CV can be obtained by an appropriate choice of  $k$ .

**Example 3** (Inverse Gaussian). Same, as before:

Mean is

$$\frac{1}{\langle T \rangle} = \nu$$

so that

$$CV = \frac{SD(T)}{\langle T \rangle} = \frac{1}{\sqrt{k}}$$

We observe that any value of CV can be obtained by an appropriate choice of  $k$ .

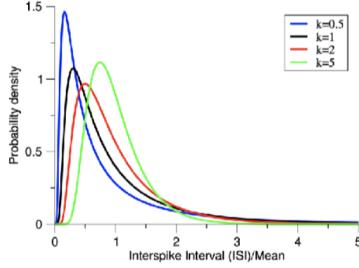


Figure 21: Inverse gaussian distribution as  $k$  varies

#### 5.4 LIF model with Poisson inputs

Our goal is to model the irregular firing behavior of neurons in the visual cortex (figure 18). So far, we have only described an approach to quantify the irregularity.

In a **single neuron**, the membrane potential obeys the LIF equation

$$\tau \frac{dV}{dt} = -V + I(t) \quad (9)$$

where  $I(t)$  is a Poisson process with rate  $\nu_{in}$

$$I(t) = J\tau \sum_k \delta(t - t_k^i)$$

where the  $\delta(\cdot)$  are Dirac functions and  $t_i^k$  is the time of the  $k^{th}$  spike of the presynaptic neuron  $i$  (here we assume there is only a single pre-synaptic cell, but for many of them, this can be easily generalized by summing over all pre-synaptic neurons).

By discretizing equation 9, we obtain

$$\tau \frac{V(t_i^{k+}) - V(t_i^{k-})}{\Delta t} = -V(t_i^{k-}) + J\tau \frac{1}{\Delta t}$$

because, in the discrete case, the Dirac  $\delta$  function takes on finite values

$$I(t_i^{k-}) = J\tau \sum_k \delta(t_i^{k-} - t_k^i) = J\tau \frac{1}{\Delta t}$$

By multiplying by  $\Delta t$  on both sides and taking the limit of  $\Delta t \rightarrow 0$  we have

$$V(t_i^{k+}) = V(t_i^{k-}) + J$$

Hence a single spike in the pre-synaptic neuron produces a jump in the post-synaptic membrane potential of magnitude  $J$ .

#### 5.5 Diffusion Approximation

We assume that the firing of each pre-synaptic neuron is a Poisson process with rate  $\nu_{in}$ .

Let  $n_i(t)$  be defined as follows

$$I(t) = J\tau \sum_{i=1}^N \underbrace{\sum_k \delta(t - t_k^i)}_{n_i(t)}$$

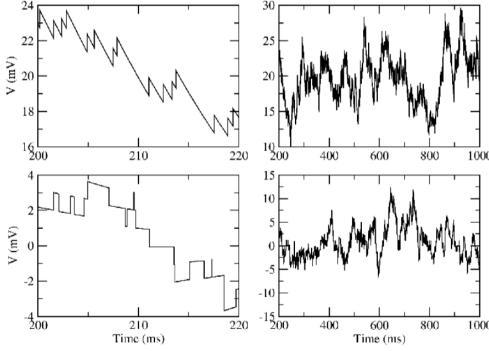


Figure 22: Membrane potential as function of time for different parameters (20 milliseconds vs 800 milliseconds)

so that  $n_i(t)$  is a Poisson random variable that counts the number of spikes between  $t$  and  $t + \Delta t$ . In particular  $n_i(t)$  has mean  $\nu_{in}$ . Indeed:

$$n_i(t) = \begin{cases} \frac{1}{\Delta t} & \text{w. p. } \nu_{in}\Delta t \\ 0 & \text{w. p. } 1 - \nu_{in}\Delta t \end{cases}$$

It holds

$$\langle I \rangle = J\tau N\nu_{in}$$

where  $N$  is the number of pre-synaptic neurons.

Noting that

$$I(t) = J\tau \sum_{i=1}^N n_i(t)$$

the covariance of  $I(t)$  and  $I(t')$  is given by

$$\begin{aligned} \text{Cov}(I(t), I(t')) &= \langle I(t')I(t) \rangle - \langle I(t) \rangle \langle I(t') \rangle \\ &= J^2\tau^2 \sum_{i,i'} \langle n_i(t)n_{i'}(t') \rangle - J^2\tau^2 N^2 \nu_{in}^2 \\ &= J^2\tau^2 \left[ \underbrace{N(N-1)\langle n_i(t) \rangle \langle n_{i'}(t') \rangle}_{i \neq i'} + \underbrace{N\langle n_i(t)n_i(t') \rangle}_{i=i'} \right] - J^2\tau^2 N^2 \nu_{in}^2 \\ &= J^2\tau^2 \left[ (N^2 - N)\nu_{in}^2 + N\langle n_i(t)n_i(t') \rangle \right] - J^2\tau^2 N^2 \nu_{in}^2 \\ &= J^2\tau^2 N \left[ \langle n_i(t)n_i(t') \rangle - \langle n_i(t) \rangle \langle n_i(t') \rangle \right] \\ &= J^2\tau^2 N \left[ \langle n_i(t)^2 \rangle - \langle n_i(t) \rangle^2 \right] \mathbb{1}(t \neq t') \quad \text{as } t \neq t' \implies \langle n_i(t)n_i(t') \rangle = \langle n_i(t) \rangle \langle n_i(t') \rangle \\ &= J^2\tau^2 N \left( \frac{\nu_{in}}{\Delta t} - \nu_{in}^2 \right) \mathbb{1}(t = t') \\ &= J^2\tau^2 N \nu_{in} \delta(t - t') \end{aligned}$$

where the last equality holds in the limit of  $\Delta t \rightarrow 0$  for a discrete version of the Dirac delta function.

In particular, we can now write the current as a random variable

$$I(t) = J\tau \sum_k \delta(t - t_k) \simeq \mu + \sigma\sqrt{\tau}\eta(t)$$

where  $\mu = JN\nu_{in}\tau$  and  $\sigma^2 = NJ^2\nu_{in}\tau$  and  $\eta(t)$  represents noise, satisfying  $\langle \eta(t) \rangle = 0$  and  $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$ . This choice is motivated by the empirical observation of post-synaptic potential, that has a Brownian behavior. Given that Brownian motion is given by the integral of white noise (by definition), then it is sensible

to assume a white-noise structure in the Input current (whose integral is part of the expression for the potential).

Moreover, in the limit of  $N\nu_{in}\tau \gg 1$ , i.e., for a large number of pre-synaptic neurons, we can apply the **Central Limit Theorem**, so that  $I(t)$  is well-approximated by a Gaussian random variable and satisfying  $I(t) \perp\!\!\!\perp I(t') \forall t \neq t'$ .

The other underlying assumption is that  $J/V \ll 1$ , i.e., each jump in the membrane potential should be much smaller than the value of the membrane potential. This is because, as shown in the upper left quadrant in figure 22, the jump induced by the current in the membrane potential are never large enough to make the potential reach the threshold value (because that would cause a reset of  $V$  to its rest value). This allows to make a large number of jumps without reaching threshold.

As a consequence, the membrane potential obeys the following Langevin equation

$$\tau \frac{dV}{dt} = -V + \mu + \sigma\sqrt{\tau}\eta(t)$$

along with its discretized version

$$V(t + \Delta t) = V(t) + \frac{\Delta t}{\tau} \left( \mu - V(t) + \sqrt{\frac{\Delta t}{\tau}} \sigma Z(t) \right) \quad (10)$$

where  $Z(t)$  is a random Gaussian variable of zero mean and unit variance.

## 5.6 Fokker-Planck Equation

The goal now is to understand the statistics of the membrane potential (relevant to understanding the firing probability of the post-synaptic neuron) from the statistics of the input.

Hence, we are interested in computing  $P(V, t)$ , the probability that the membrane potential takes a value  $V$  at time  $t$ .

Specifically we want to do so starting from equation 10.

To derive the expression of the probability distribution of the membrane potential starting from its differential counterpart, we apply the Fokker-Planck equation.

Consider a general function

$$\begin{aligned} f(V(t + \Delta t)) &= f\left(V(t) + \frac{\Delta t}{\tau} \left( \mu - V(t) + \sqrt{\frac{\Delta t}{\tau}} \sigma Z(t) \right)\right) \\ &= f(V(t)) + f'|_{V(t)} \left[ (\mu - V) \frac{\Delta t}{\tau} + \sigma Z(t) + \sqrt{\frac{\Delta t}{\tau}} \right] + \frac{f''}{2} \Big|_{V(t)} \left[ \sigma^2 Z^2(t) \frac{\Delta t}{\tau} \right] \end{aligned}$$

where  $f$  has been expanded using the Tailor method,  $f(V + \epsilon) = f(V) + \epsilon f'(V) + \frac{\epsilon^2}{2} f''(V)$  with

$$\epsilon = \frac{\Delta t}{\tau} \left( \mu - V(t) + \sqrt{\frac{\Delta t}{\tau}} \sigma Z(t) \right)$$

The goal is now to compute  $\frac{d\langle f \rangle}{dt}$ .

We have

$$\begin{aligned} d\langle f \rangle &= \langle f(V(t + \Delta t)) \rangle - \langle f(V(t)) \rangle \\ &= \langle f(V(t)) + f'|_{V(t)} \left[ (\mu - V) \frac{\Delta t}{\tau} + \sigma Z(t) + \sqrt{\frac{\Delta t}{\tau}} \right] + \frac{f''}{2} \Big|_{V(t)} \left[ \sigma^2 Z^2(t) \frac{\Delta t}{\tau} \right] \rangle - \langle f(V(t)) \rangle \\ &= \langle f'(V) \frac{\mu - V}{\tau} \Delta t \rangle + \langle \frac{f''}{2} \rangle \frac{\sigma^2}{\tau} \Delta t \end{aligned}$$

so that

$$\begin{aligned}
\frac{d\langle f \rangle}{dt} &= \langle f'(V) \frac{\mu - V}{\tau} \rangle + \langle \frac{f''}{2} \rangle \frac{\sigma^2}{\tau} \\
&= \int P(V, t) \left[ f'(V) \frac{\mu - V}{\tau} f''(V) \frac{\sigma^2}{2\tau} \right] dV \\
&= \int f(V) \left[ \frac{\partial}{\partial V} \left[ \frac{V - \mu}{\tau} P(V, t) \right] + \frac{\sigma^2}{2\tau} \frac{\partial^2 P(V, t)}{\partial V^2} \right] dV
\end{aligned} \tag{11}$$

Moreover, we know that

$$\begin{aligned}
\frac{d\langle f \rangle}{dt} &= \frac{d}{dt} \int P(V, t) f(V) dV \\
&= \int \left[ \frac{\partial}{\partial t} P(V, t) \right] f(V) dV
\end{aligned} \tag{12}$$

Equations 11 and 12 are true for any function  $f$ , so that what is in the squared brackets must always hold. Hence  $P(V; t)$  is described by

$$\tau \frac{\partial P(V, t)}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 P(V, t)}{\partial V^2} + \frac{\partial}{\partial V} \left[ (V - \mu) P(V, t) \right]$$

$P(V, t)$  can also be rewritten as a continuity equation:

$$\frac{\partial P(V, T)}{\partial t} = - \frac{\partial S(V, T)}{\partial V}$$

with  $S(V, T) = -\frac{\sigma^2}{2} \frac{\partial P(V, t)}{\partial V} + (\mu - V) P(V, t)$  being interpreted as a flux of probability. In particular, it represents how many neurons per unit time are subject to an increase in the value of their membrane potential. Hence, by definition, the number of neurons per unit time whose membrane potential reaches threshold value is the firing rate, i.e.

$$S(V_{th}, t) = \nu_{out} \tag{13}$$

We need to set boundary conditions to solve the differential equation:

a) At threshold  $V_{th}$  we impose

$$P(V_{th}, t) = 0, \quad \frac{\partial P}{\partial V}(V_{th}, t) = -\frac{2\nu_{out}\tau}{\sigma^2}$$

This is given by substituting the expression of  $S(V_{th}, t)$  in equation 13.

b) At reset potential  $V_R$  we have

$$P(V_R^-, t) = P(V_R^+, t), \quad \frac{\partial}{\partial V} P(V_R^-, t) - \frac{\partial}{\partial V} P(V_R^+, t) = -\frac{2\nu_{out}\tau}{\sigma^2}$$

because all neurons that emit a spike then go to reset value  $V_R$ .

Finally, the firing rate of the post-synaptic neuron  $\nu_{out}$  can be found by first finding the average time it takes for the membrane potential of the neuron to reach the threshold value, and then by taking the inverse of this quantity.

For a constant mean  $\mu = \mu_0$  and constant variance  $\sigma = \sigma_0$ , the Fokker-Planck equation converges to a stationary distribution that is very close to a Gaussian if the mean is below-threshold ( $\mu < V_{th}$ ), while it strongly deviates from Gaussianity for  $\mu > V_{th}$ , as shown in figure 23. Also, for a sub-threshold mean, the firing is only due to random fluctuations in the membrane potential. The firing statistics turns out to be very close to a Poisson process. For above-threshold mean instead, we have regular spiking.

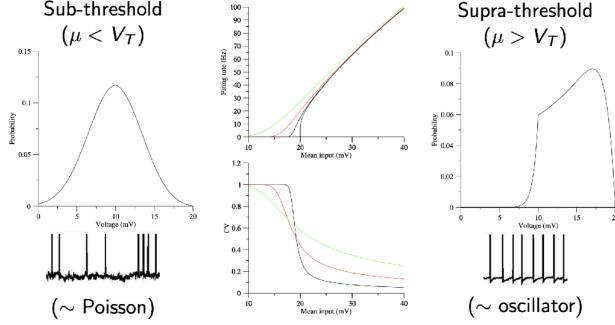


Figure 23: graphs of  $P(V, t)$  for different means. The central graphs represent the firing rate of the neuron as a function of the mean input  $I$ , with different lines representing different noise levels. (the non-noise curve was already computed in the case of a LIF model with 0 noise and a step-input-function). With noise we have responses even for inputs that are below threshold. This theory has been experimentally verified by injecting a current with a fixed mean and variance into a single neuronal cell.

## 6 Synapses

Interactions between neurons are mediated by **spines**. A spine is the point of interaction between two neurons, and its structure is shown in figure 24.

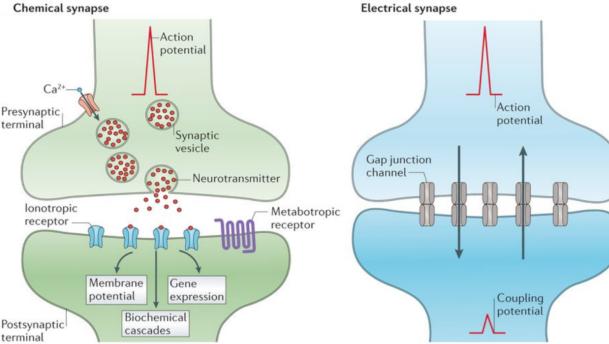


Figure 24: Structure of a pre-synaptic and post-synaptic spine

There are two main types of synapses: **chemical** and **mechanical**.

For chemical synapses: when there is an action potential generated in the pre-synaptic neuron, the spike travels along the axon and reaches the spine. Here the vesicles reach the surface of the spine and release the neurotransmitters in the extracellular medium. The neurotransmitters then reach the detectors in the post-synaptic neuron, where ionic channels open to allow ions to go through.

Regarding electrical synapses: a Gap junction channel directly allows a flux of ion to flow from one cell to the other.

### Mathematically:

For chemical synapses (which represent the vast majority), the input current is represented as

$$I_{ij} = g_{ij}(V_i - V_R) \sum_k s(t - t_j^k) \quad (14)$$

where  $g_{ij}$  is the maximal synaptic conductance,  $V_i$  is the post-synaptic potential, and  $V_R$  is the reversal potential (characteristic of the specific ionic channel in the synapse). The function  $s()$  describes the dynamics of the opening and closing of the ionic channels in response to the receptors.

Electrical synapses, instead tend to connect neurons of the same type. The current here is simply given by

$$I = g(V_i - V_R)$$

where  $g$  is the gap junction conductance.

## 6.1 In vitro recording of synapses

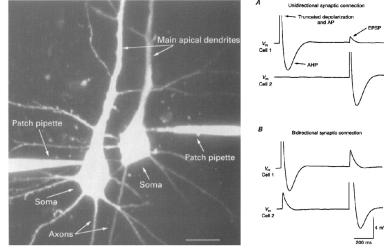


Figure 25: Experiment on synapses with two neurons. In the upper graph there is an unidirectional synaptic connection, as the spike in Cell 2 happens only after a spike in Cell 1. In the lower graph is shown a bidirectional synaptic connection, as a spike in either cell causes a change in potential in the other one (the change in potential is given by the equation  $\frac{dV}{dt} = -V + I$ , where  $I$  is found through equation 14).

## 6.2 Kinetics of synaptic transmission

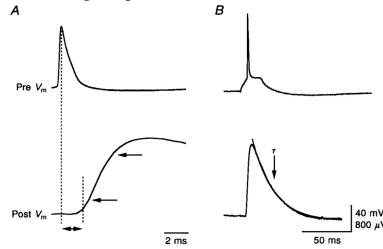


Figure 26: Graphs of pre-synaptic and post-synaptic membrane potential in different time scales. The post-synaptic increase in potential is not immediate after the pre-synaptic spike: there is some latency in the order of milliseconds. These graphs are obtained by averaging across many measurements. Synaptic transmission is highly stochastic, in the sense that there can be fairly strong fluctuations and deviations across measurements.

## 6.3 Synaptic Failures

A synaptic failure occurs when there is no significant change in the post-synaptic membrane potential in response to a pre-synaptic spike.

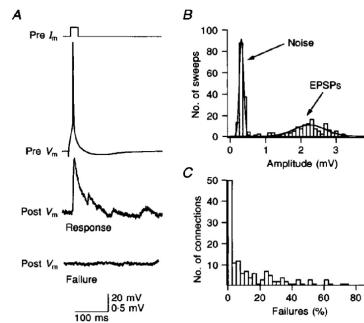


Figure 27: Synaptic failures (left), number of connections and failure probability: high connections means low failure probability (bottom right).

## 6.4 Classification of synapses

For chemical synapses we can make a further distinction based on the neurotransmitters and their neuro-receptors.

The main distinction is between **excitatory** and **inhibitory** synapses, depending on the sign of the produced current (leading to an increase or decrease in the membrane potential).

- **Excitatory synapses:** neurotransmitter = glutamate

- **Ionotropic receptors:** a neurotransmitter binds to ionic channels, controlling the gating of the channels
  - \* AMPA (fast, not voltage-dependent)
  - \* NMDA (slow, voltage-dependent)
- **Metabotropic receptors:** a receptor is not associated to any channel. The binding of the neurotransmitter to the receptor leads to some chemical processes in the post-synaptic neuron that then modulate the opening of ionic channels ("second-order interaction")

- **Inhibitory synapses:** neurotransmitter = GABA

- **Ionotropic receptors**
  - \*  $GABA_A$ , similar to AMPA (fast)
- **Metabotropic receptors**
  - \*  $GABA_B$ , similar to NMDA (slower)

- Other: glycinergic, aminergic, peptidergic

Receptor	E/I	Ion	$V_R$	Does g depend on postsyn V?	Kinetics
AMPA	E	Na, K, Ca	0	No	Fast (1-5ms)
NMDA	E	Na, K, Ca	0	Yes	Slow (20-200ms)
GABA	I	Cl	-80	No	Fast (3-10ms)

Figure 28: Summary of properties of most common synaptic currents. A main difference w.r.t. previous models (especially HH) is that, here, each type channel allows the flux of different ions, e.g. AMPA is associated with channels that allows flux of Sodium, Calcium and Potassium. The  $V_R$  column indicates the reversal potential, i.e. the value of potential that causes a change in sign in the expression of the current  $I = g(V_i - V_R)s(t - t^k)$ .

**Remark.** How are excitatory and inhibitory synaptic potential? By moving the potential  $V_i$  to the reversal potential of the opposite channel.

For example, if a post-synaptic neuron has an Excitatory and an Inhibitory connection and the goal is to measure only the Excitatory current, we can use voltage clamps in the soma of the post-synaptic neuron, setting a voltage equal to the Reversal potential of the Inhibitory connection, which will then produce zero current. This allows to measure the properties of just the excitatory current.

**Remark.** According to Dale's Law, a neuron performs the same chemical action at all of its synaptic connections to other cells, regardless of the identity of the target cell.

## 6.5 NMDA currents

The NMDA conductance is voltage-dependent, as shown in figure 29.

NMDA receptors have glutamate and glycine as neurotransmitters. These bind to receptors in ionic channels every time there is a pre-synaptic spike. However, the presence of magnesium,  $Mg^{2+}$ , will prevent a flux of ions to go through the channels.

At the same time, as the post-synaptic membrane potential increases, magnesium detaches from the channel. In this situation, (detachment of magnesium and binding of glutamate and glycine), there is a flux of ions through the channel. This is all shown in figure 30.

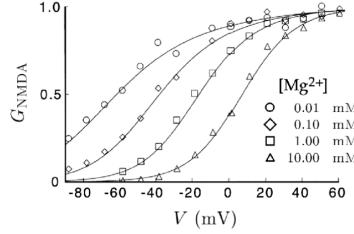


Figure 29: NMDA conductance as a function of voltage. For low values of  $V$ , the conductance is also very small, which means that even if there is a spike, the current produced will be small. For higher concentration of magnesium, the NMDA conductance becomes small, as magnesium tends to block the flux of ions through the ionic channels in NMDA receptors.

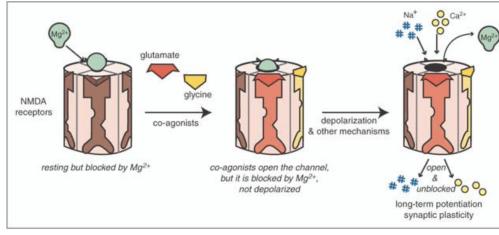


Figure 30: Magnesium detachment, NMDA neurotransmitter and flux of ion for NMDA receptors

**Remark.** NMDA receptor can be considered as a "coincidence" detector because it's active just if there's an high voltage both in the pre-synaptic neuron (the neurotransmitters are released) and in the post-synaptic neuron (the magnesium detached from its binding site)

## 6.6 A Markov model for synaptic currents

Recall equation 14, that states that the current produced is given by  $I = g(V_i - V_R)s(t - t^k)$ . We now assume that  $s$  is the fraction of open channels. In particular, it satisfies

$$\frac{ds}{dt} = \alpha[N](1 - s) - \beta s$$

where  $[N]$  is the concentration of neurotransmitters.

This means that the Poisson processes regulating the fraction of open and closed channels has a rate of going from "Closed" to "Open" of  $\alpha[N]$ , meaning that, as  $[N]$  increases, channels will tend to open. On the other hand, we have a fixed relaxation rate  $\beta$  (rate of going from "Open" to "Closed").

Consider  $[N] = N_0\delta(t)$ . Then,

$$\frac{ds}{dt} = \alpha N_0 \delta(t) - [\beta + \alpha N_0 \delta(t)]s$$

so that if  $t \neq 0$  then

$$\frac{ds}{dt} = -\beta s \implies s(t) = s_0 e^{-\beta t}$$

In order to determine  $s_0$  we integrate around  $t = 0$ , the time of the spike:

$$\int_{-\Delta t}^{+\Delta t} \frac{ds}{dt} dt = \alpha N_0 - \alpha N_0 s(0) - \beta s(0) 2\Delta t$$

which implies, as  $\Delta t \rightarrow 0$

$$s(0^+) = \alpha N_0 - \alpha N_0 s(0)$$

because we assume  $s(t) = 0$  for every  $t < 0$ .

For  $[N] = N_0 e^{-t/\tau_r} \mathbb{1}_{(t \geq 0)}$  we have that  $s$  rises with a time constant  $\tau_r$ , then decays with time constant  $1/\beta$ . In general, the neurotransmitter concentration can be described as a difference between exponentials for positive

$t$  and this will lead to an exponential rise (with rate  $\tau_r$ ) and an exponential decay (with rate  $\tau_s$ ) in the variable  $s$ .

This allows to model the experimental observations of exponential rises and relaxations in the synaptic currents.

## 6.7 Stochastic nature of synaptic transmission

Katz et al. discovered in the '50s discovered that **neurotransmitter release is a stochastic event** and that **Neurotransmitters are released in discrete quantities called quanta**.

They studied the neuro-muscular junction (neuronal system controlling muscles), creating an action potential in the pre-synaptic neuron and measuring the potential generated in the muscle fiber.

In particular, they observed a stochastic behavior across different observations of the phenomenon and that the distribution of the potential is not continuous.

## 6.8 Quantal model

We assume that each synapse has  $N$  vesicles. When a spike arises, each one of them is released with a fixed probability  $p$ . If a single vesicle is released, it produces in the post-synaptic neuron a current  $q$  (quantum).

### 6.8.1 Distribution, mean and variance of synaptic currents

Assuming that  $n$  vesicles are released out of the  $N$  total ones, the total current will be given by

$$I = nq$$

Then

$$P(I = nq) = \binom{N}{n} p^n (1-p)^{N-n}$$

We can compute the failure probability as

$$P(fail) = (1-p)^N$$

The mean and variance as

$$\begin{aligned} \langle I \rangle &= \sum qnP(n) = pqN \\ Var(I) &= q^2p(1-p)N = q\langle I \rangle - \frac{\langle I \rangle^2}{N} \end{aligned}$$

Therefore,  $Var(I)$  as a function of  $\langle I \rangle$  is a parabola with zeros at  $\langle I \rangle = 0, NQ$ . Hence, by moving  $\langle I \rangle$  (e.g. by changing the values of  $p$ ), we can produce points along this curve. By fitting the data on the curve, we can infer the parameters  $q$  and  $N$  of the model.

## 6.9 Summary: models for Post-Synaptic Currents

- Standard Biophysical, conductance-based

- The post-synaptic current is given by

$$I = g(V_i - V_r) \sum_k s(t) \tag{15}$$

where  $s(t)$  is the difference of two exponentials

- $g$  can be deterministic or stochastic
- For NMDA currents  $g$  is voltage-dependent

- Simplified current-based models

- The current has no dependance on voltage

$$I(t) = Is(t)$$

- $s(t)$  is sometimes also further simplified to a single exponential or delta function

## 7 History dependence - short term depression

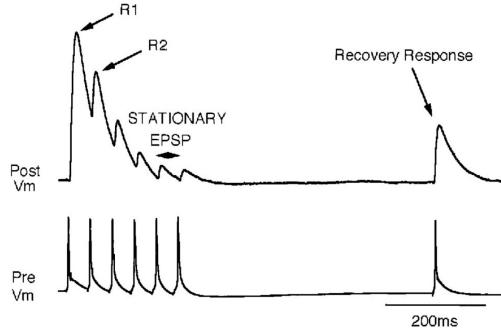


Figure 31: A typical example of short-term plasticity. In the upper panel is a post-synaptic neuron, while in the lower part is shown a pre-synaptic neuron. The response to the first spikes is much larger than the ones for later spikes. Waiting long enough allows a recovery of the response to a new spike in the pre-synaptic neuron.

### 7.1 Tsodyks-Markram model

The weaker response for later spikes is given by the finiteness of vesicles: if we keep on generating spikes, we will run out of them eventually.

Vesicles can be in two possible states, ‘available for release’ (A) or ‘refractory’ (R).

Refractory vesicles capture neurotransmitters in the extra-cellular medium and form new vesicles.

When there is no spike, we have an accumulation (up to a certain value) of available-for-release vesicles. When a spike occurs, with probability  $u$  the vesicle is released ( $u$  can be seen as the probability that a vesicle is used by the Pre-synaptic neuron) and shifts to the Refractory state, while with probability  $1 - u$  the vesicle is not released.

Let  $x$  be the fraction of available vesicles. Then:

$$\frac{dx}{dt} = \frac{1 - x}{\tau_D} - ux \sum_k \delta(t - t_k) \quad (16)$$

where  $u$  is the fraction of used vesicle by a pre-synaptic action potential.

If there is no spike,  $x = 1$ . When a spike occurs, the value of  $x$  drops by  $u$ , then the  $\delta$  term in equation 16 is zero and  $x$  tends to 1 again. Depending on the recovery time constant  $\tau_D$  and on the time between spikes, the value of  $x$  may or may not go back to 1 before the next spike occurs.

The parameters of the model are  $\tau_D$  and  $u$ .

### 7.2 Average synaptic currents as a function of the frequency

If there are no spikes, equation 16 becomes

$$\frac{dx}{dt} = \frac{1 - x}{\tau_D}$$

whose general solution in a right neighbourhood of  $t_k$  is

$$x(t) = Ae^{\frac{-t+k^+}{\tau_D}} + 1$$

where  $x(k^+)$  is the value of  $x$  right after the  $k^{th}$  spike, with

$$x(k^+) = x(k^-)(1 - u) \quad (17)$$

because the fraction of available vesicles at  $t = k^+$  is the fraction of available vesicles at  $t = k^-$  times the probability/proportion of vesicles that are not released,  $1 - u$ .

Thus, plugging  $t = k^+$  in equation 17, allows to find the value of  $A$ ,

$$A = x(k^-)(1 - u) - 1$$

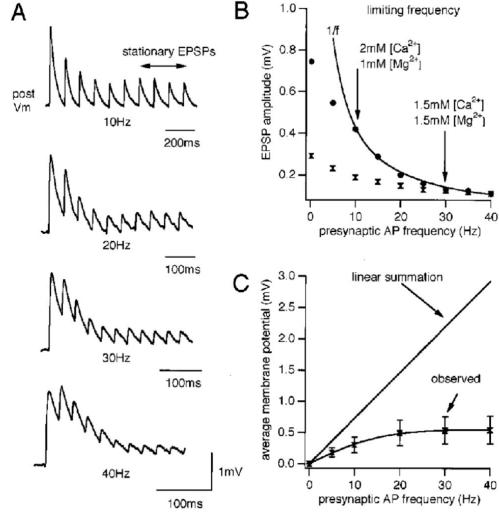


Figure 32: Post-synaptic potential generated based on the temporal frequency of pre-synaptic spikes. The important observation is that the amplitude of the post-synaptic potential decreases as the inverse of the frequency (upper-right graph)

Hence, given two spikes occurring at times  $t_k$  and  $t_{k+1}$

$$x(t_{k+1}^-) = \left[ x(t_k^-)(1 - u) - 1 \right] \cdot \exp \left[ \frac{-t_{k+1}^- + t_k}{\tau_D} \right] + 1$$

We are especially interested in stationary solutions, i.e.,  $x(t_k^-) = x(t_{k+1}^-)$ .

If we impose this condition in the last equation we obtain the kind of dependence of post-synaptic potential on temporal frequency of pre-synaptic spikes shown in figure 32.

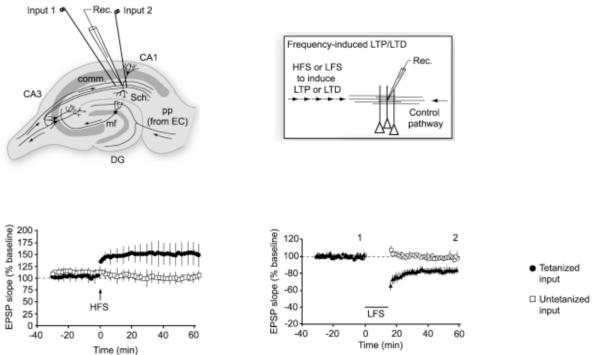
# 8 Long-term synaptic Plasticity

## 8.1 LTP and LTD in hippocampal slices

It is an experimental setup with two cells in the hippocampus, from which the recording is taken; two axons coming from other cells project into the two considered hippocampus neurons.

There are two kind of stimulation: the first one involves an axons stimulating the considered cell, and the second one involves an axon of a cell that has nothing to do with that cell (as control).

They measure the post-synaptic potential in response to the first stimulation (black dots) with a high-frequency (many pre-synaptic spikes). After this protocol, the postsynaptic potential is increased by about 50%, where it remains in the long term (about an hour). The white dots instead represent the control stimulation (which remains constant), as shown in the bottom left of figure 8.1. In the bottom-right instead is shown a low-frequency stimulation. After this the excitatory post-synaptic potential (EPSP) is decreased for several minutes (long-term depression).



## 8.2 Paired Recordings

Intra-cellular recordings are performed with both cells being stimulated.

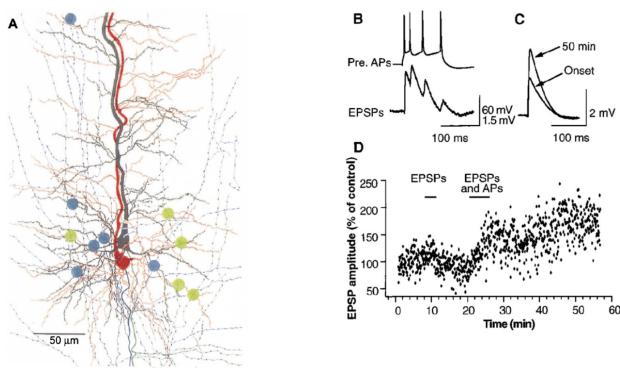


Figure 33: In the bottom-right graph is shown the Excitatory Post-Synaptic Potential (EPSP) of the post-synaptic neuron. Up to the first 20 minutes, only the pre-synaptic spike is stimulated. After that both neurons are subject to stimuli, leading to an overall increase in EPSP in the long term.

### 8.2.1 Spike-timing dependent plasticity (STDP) protocol

- For small values of  $\Delta t > 0$  (i.e. pre-synaptic spike occurs before the post-synaptic spike) we observe **suppression**

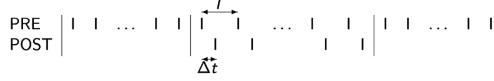


Figure 34: Related to Figure 33, the first chunk represents the first 20 minutes in which only the pre-synaptic neuron is stimulated. The second section represents spikes in both cells, where  $\Delta t$  represents the interval of time between the two cells. In the third chunk we still have only pre-synaptic stimulations.

- for small values of  $\Delta t < 0$  (i.e. pre-synaptic spike occurs after the post-synaptic spike) we observe **potentiation**.
- For larger time intervals no significant changes in the post-synaptic potential are observed.

### 8.3 Diversity of STDP curves

Many types of Synaptic plasticity have been observed (other than STDP). Examples are shown in figure 35.

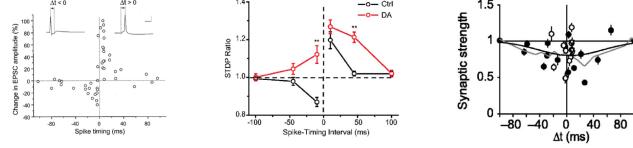


Figure 35: Neuromodulators like dopamine can affect the shape of the STDP curve (central graph).

### 8.4 Phenomenological learning rules - STDP

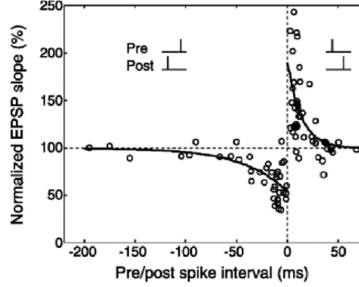


Figure 36: Illustration of spike timing dependent plasticity time windows. Depending on the precise time difference between a post- and a pre-synaptic spike, the synaptic weight can be either depressed or potentiated.

In general we have an STDP rule in which each neuron has a weight which changes with every spike according to the following equation

$$\begin{cases} w_{ij} \rightarrow w_{ij} + \Delta w_{ij} \\ \Delta w_{ij} = f(\{t_{ik}\}, \{t_{jk}\}, w_{ij}) \end{cases}$$

where  $\{t_{ik}\}$  is the time of the  $k^{th}$  spike of the pre-synaptic neuron  $i$  and  $\{t_{jk}\}$  is the time of the  $k^{th}$  spike of the post-synaptic neuron  $j$ . With this notation we consider the  $k^{th}$  spikes to be the closest in time between the two neurons.

We can fit the curve shown in figure 36 with a functional form of the following type

$$f(\{t_{ik}\}, \{t_{jk}\}, w_{ij}) = \sum_{k,k'} F(t_{ik} - t_{jk'}, w_{ij})$$

with

$$F(\Delta t, w_{ij}) = \begin{cases} A_+(w_{ij})e^{-\Delta t/\tau_+}, & \Delta t > 0 \\ -A_-(w_{ij})e^{-\Delta t/\tau_-}, & \Delta t < d0 \end{cases}$$

## 8.5 STDP rule fails to reproduce data obtained in other protocols

We can establish a protocol in which only three spikes occur. We can have a pre-, post-, pre-synaptic spikes or post-, pre-, post-synaptic spikes.

According to the rule described in 8.2, we expect to observe no significant changes in the post-synaptic potential, because we would have either a depression first and a potentiation afterwards or vice-versa; in both cases the two phenomena should cancel each other out (the STDP rule is symmetric).

However, in the post-pre-post case we observe potentiation.

## 8.6 Phenomenological learning rules - firing rate

Often, the change in the weights of a pair of neurons is considered to be determined by a function of the firing rate (number of emitted spikes per second) of both neurons

$$\Delta w_{ij} = f(r_i, r_j)$$

Most common models are

- Pure Hebbian

$$f(r_i, r_j) = r_i r_j$$

- Covariance

$$f(r_i, r_j) = (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle)$$

- BCM

## 8.7 Long-term plasticity: Mechanisms

Typical models of Long-term plasticity are induced by calcium in-flow through NMDA channels in the spine. This leads to a cascade of bio-physical processes in protein interaction networks. The final consequences of these processes lead to

- A change in properties of AMPA receptors
- The addition or removal of AMPA receptors
- Structural changes on the spine
- Changes at pre-synaptic level, modifying the probability of release.

## 8.8 Calcium-based models: Shouval et al. (2002)

Calcium enters the spine through NMDA channel, when there is a coincidence of pre-synaptic (glutamate binding - corresponding to a pre-synaptic spike) and post-synaptic (depolarization due to back-propagating action potential) activity.

Synaptic efficacy (weights  $w_{ij}$ ) changes as a function of the concentration of Calcium

$$\frac{dw_{ij}}{dt} = \eta([Ca]_{ij})(\Omega([Ca]_{ij}) - w_{ij}) \quad (18)$$

where  $\Omega$  has the behavior shown in figure 37.

In particular, if the Calcium concentration changes from a value below  $\theta_d$  to a value in  $(\theta_d, \theta_p)$  i.e., below baseline, then (ignoring the value of  $\eta([Ca])$ ) we have depression, while if the calcium concentration increases above  $\theta_p$  we will have potentiation.

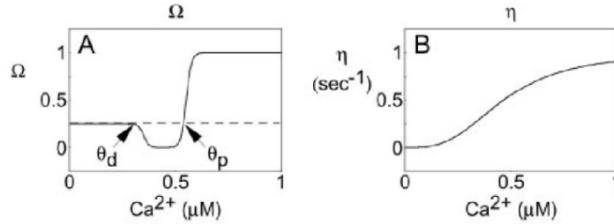


Figure 37: Graphic representation of the  $\eta$  and  $\Omega$  functions

## 8.9 Reproducing standard protocols with the model

If there is just a pre-synaptic spike (first column of figure 38) the calcium concentration is such that the value of  $\Omega([Ca])$  remains in  $[0, \theta_d]$ , so that, according to equation 18, nothing changes in the synaptic weights  $w_{ij}$ . If the post-synaptic spike occurs before the pre-synaptic one (with  $\Delta t = -10ms$ ), then the concentration of calcium increases to a value in  $[\theta_d, \theta_p]$ , so that the value of  $\Omega$  leads to depression. Lastly, for  $\Delta t = +10ms$  (third column), then the concentration of calcium exceeds  $\theta_p$  and we have potentiation.

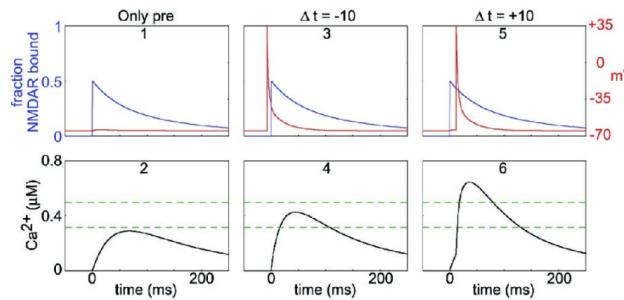


Figure 38: The upper graphs represent the fraction of NMDA receptors in which there is bounding of neurotransmitters (in blue) and a post-synaptic spikes (in red). The binding of NMDA receptors induces a flux of calcium (lower graphs). In particular, the values of the dashed lines represent the two values of  $\theta_d$  and  $\theta_p$  in figure 37

## 8.10 Synaptic strength is controlled by multiple factors

The extent to which synaptic weights change in time is determined by many factors:

- Spike timing, figure 36
- Firing rate
- Neuromodulators (such as dopamine)
- Post-synaptic membrane potential. In particular, at low potential we generally have depression, while at higher potential we observe excitation, as shown in figure 39. This can be explained in the model because when the post-synaptic potential is high, the NMDA conductance will in turn be large, causing the influx of calcium to be sustained enough to produce potentiation.

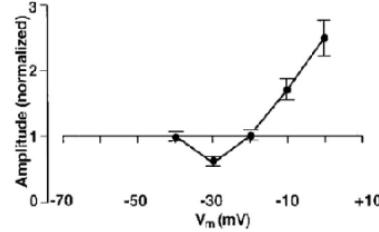


Figure 39: The upper graphs represent the fraction of NMDA receptors in which there is binding of neurotransmitters (in blue) and a post-synaptic spikes (in red). The binding of NMDA receptors induces a flux of calcium (lower graphs). In particular, the values of the dashed lines represent the two values of  $\theta_d$  and  $\theta_p$  in figure 37

## 9 Networks

### 9.1 Connectomics: Obtaining wiring diagrams

There are ways to reconstruct the wiring diagrams from brain tissue.

Usually a cube of brain structure of an animal is solidified, then sliced thinly. Each slice is then photographed in order to reconstruct axons and dendrites and ultimately the neural network.

### 9.2 The C Elegans brain network - numbers

- 302 neurons
- 279 out of 302 neurons form a connected network
- 6393 chemical synapses, for 2194 directed connections (average 7.9/ neuron, connection probability 0.03)
- 890 gap junctions, for 514 connected pairs (average 1.8/neuron, connection probability 0.007)
- 1410 neuromuscular junctions

### 9.3 C Elegans - Connectivity matrix

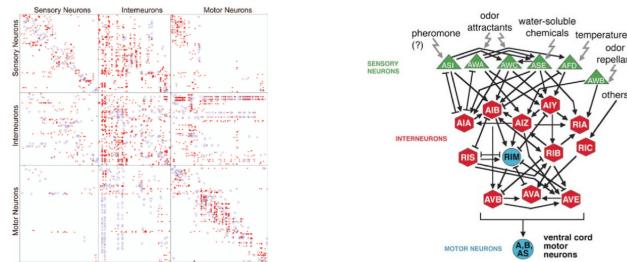


Figure 40: 302x302 matrix with all the neurons classified in Sensory neurons, motor neurons and interneurons. Sensory neurons collect data from the environment, interneurons process them and motor neurons generate output.

### 9.4 Feed-forward, recurrent, feed-back connections

A feed-forward connection connects a neuron to another one in a different part of the macro-structure of the brain. So it can be of the type  $S \rightarrow I$ ,  $S \rightarrow M$  or  $I \rightarrow M$ .

Recurrent connection are between neurons of the same type,  $S \rightarrow S$ ,  $M \rightarrow M$  or  $I \rightarrow I$ .

Lastly, feed-back connections are of the type  $I \rightarrow S$ ,  $M \rightarrow S$  or  $M \rightarrow I$ , opposed to the general direction of information processing in the network on the right of figure 40.

## 9.5 Types of networks/graphs

Electrical synapses are represented by non-directed graphs, chemical synapses are encoded through directed graphs (as we talk about pre-synaptic and post-synaptic neurons), while weighted graphs represent both.

## 9.6 ER Network

It is a (directed or non-directed) network with  $N$  neurons with uniform connection probability  $p$ , where each connection is drawn randomly and independently of the other.

### 9.6.1 Properties of ER network

The indegree or outdegree are given by a binomial distribution

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Where  $P(k)$  is the probability that a cell has indegree (or outdegree) equal to  $k$ .

## 9.7 Stochastic Block models

Consider a network of  $N$  nodes, divided into  $n$  modules or communities.

We use a  $n \times n$  matrix of connection probabilities, where  $p_{ij}$  is the probability that a node in module  $j$  is connected to a node in module  $i$ .

Each connection is drawn randomly and independently of all others, with probability  $p_{ij}$ .

Clearly, if  $p_{ij} = p$  for all  $i, j$ , then we obtain the ER network.

This type of graph is often used to describe networks with multiple cell types.

## 9.8 Spatially dependent connection probability

### 9.9 Properties of C-Elegans network

In the C Elegans network, the degree distribution (in, out degree) evolves with a power-law tail, i.e., there are many neurons with few connections and few neurons with many connections, and this histograms decays with a power law. This means that, in the C-elegans network the indegree or outdegree of a cell is described as

$$P(K) \propto K^{-\alpha}$$

for some  $\alpha$ .

In an ER network, this decays exponentially, because the indegree and outdegree distributions are binomial. Moreover, the mean shortest path length in C Elegans is 3.5 (while it is 2.9 in an ER network) and the Clustering coefficient is 0.22 (0.08 in ER).

## 9.10 Motifs in C-Elegans network

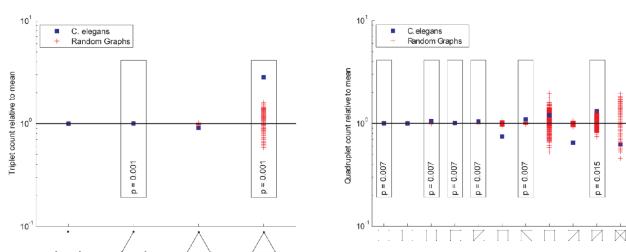


Figure 41: Comparison of the motifs in the C Elegans neural network (in blue) and an ER network (in red) in the motifs for three neurons (left) and four neurons (right)

## 9.11 Mammalian brains - macroscopic structure

### 9.11.1 Cortical Connectivity

It is possible to analyze the connectivity between macroscopic areas of the brain. Most of the relevant neural activity, however, happens within a certain area of the brain.

In the cortex there are multiple interconnected networks (so networks on a smaller scale than the whole neural network of the brain).

Cortex has several layers (figure 43). Inside each layer there are typical connectivity patterns.

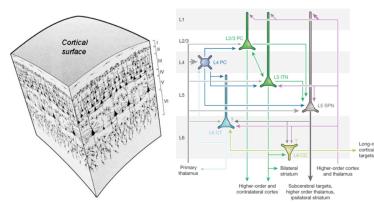


Figure 42: A cube of cortical tissue (left) and its layer structure (left)

### 9.11.2 Medium-scale patchy connectivity

Specific layers of the visual cortex present groups of neurons that behave similarly and are strongly connected. This is seen in experiments in the neurons of a specific layer of the visual cortex of an animal which is subject to visual stimuli consisting of a bar set at different angles. Plotting the firing rate of each neuron as a function of the angle of the bar, we obtain a sort of map that describes graphically the arrangement of these groups of neurons.

## 9.12 Local cortical circuits: numbers

In a cubic millimeter of cortical tissue there are about 100 000 cells. Usually 80% of cells are pyramidal (i.e. excitatory), while the rest are interneurons (i.e. inhibitory).

There are about  $10^9$  synapses, 10 000 for each neuron.

Moreover, cells are potentially connected to all other cells, with a connection probability of around 0.1.

## 10 Rate models

Our approach to analyze neuronal spiking activity so far has been based on single-compartment models, i.e., using a differential equation to describe the evolution of the membrane potential of a cell. In a network with  $N$  neurons this leads to  $O(N)$  differential equations, coupled through network connectivity matrix.

Rate models (firing rate model, neural mass model) describe the activity of a whole population of neurons by a single ‘average firing rate’ variable  $r(t)$ .

This is especially useful if we are studying a network with random connectivity in which we know whether the cells are Excitatory or Inhibitory. In that case we can consider the two groups of neurons as single entities, greatly simplifying the network, because the number of parameters is reduced from  $O(N^2)$  to 2: the average firing rate of Excitatory cells and of Inhibitory cells.

This is motivated by experimental observations showing that there are groups of neurons that behave similarly and are strongly connected.

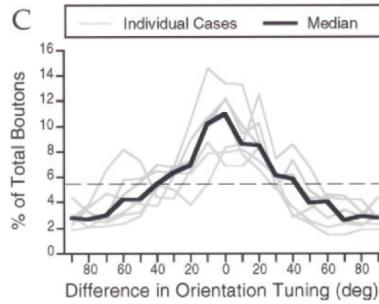


Figure 43: Number of cell connections as a function of the difference in the preferred orientation: connectivity is higher between neurons whose preferred orientation is closer.

### 10.1 Rate model

Consider a set of neurons ordered according to a variable  $x$  (e.g. the preferred orientation of a bar representing the visual stimulus). Then we define  $r(x, t)$  to be the average firing rate of all the neurons at  $x$  at time  $t$ . In a standard rate model  $r(x, t)$  evolves in time according to the following equation

$$\tau \frac{dr(x, t)}{dt} = -r(x, t) + \phi\left(I(x, t) + \int J(x, y)r(y, t)dy\right)$$

where  $\tau$  is the time constant of firing rate dynamics;  $\phi()$  is the static transfer function;  $I(x, t)$  is the external input and  $J(x, y)$  is the strength of synaptic connections between neurons at values  $x$  and  $y$ .

For example, in the model of Excitatory/Inhibitory cells in a randomly connected graph previously described,  $x$  represents whether a neuron is Excitatory or Inhibitory and  $\phi(I)$  gives the average firing rate of a neuron when it receives an input  $I$ .

### 10.2 The transfer function

Threshold linear (ReLU)

$$\phi(x) = \max\{0, x - T\}$$

Sigmoidal

$$\phi(x) = \frac{1}{1 + e^{-\beta(x-T)}}$$

#### 10.2.1 From populations of individual neurons to a rate model

In LIF models with noise inputs the firing rate in response to the mean input was zero if the input is low, linear and positive if it is bigger, and then superlinear for a larger mean input. This motivates the ReLU function, as well as the sigmoidal.

The population activity of homogeneous populations of Stochastic binary neurons and of Stochastic spiking neurons (EIF) can sometimes be shown to be well approximated by firing rate equations

$$\tau \frac{dr}{dt} = -r(t) + \phi(I(t))$$

Hence, rate models can also be used to effectively describe the average behavior of a single neuron in response to a stimulus.

### 10.3 Rate models for local networks of neurons

In all brain networks, neurons are heterogeneous and can be divided in classes. Assuming there are  $n$  sub-populations, they can be described by their average firing rate  $r_i$ , for  $i = 1, \dots, n$ .

$$\tau_i \frac{dr_i(x, t)}{dt} = -r_i(x, t) + \phi_i\left(I_{iX} + \sum_j J_{ij} r_j\right)$$

In particular, the E-I network (Excitatory-Inhibitory) falls under this category

$$\begin{cases} \tau_E \frac{dr_E}{dt} = -r_E + \phi_E\left(I_{EX} + J_{EE}r_E - J_{EI}r_I\right) \\ \tau_I \frac{dr_I}{dt} = -r_I + \phi_I\left(I_{IX} + J_{EI}r_E - J_{II}r_I\right) \end{cases}$$

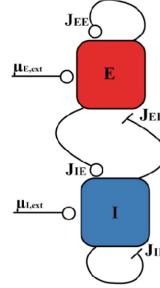


Figure 44: Scheme of the rate model for the E-I network. Here  $\mu_{E,ext}$  corresponds to  $I_{EX}$ , while  $\mu_{I,ext}$  corresponds to  $I_{IX}$ . These are the total inputs coming from the rest of the network for the  $E$  and  $I$  populations.

### 10.4 Analysis of rate models

$$\tau \frac{dr}{dt} = -r + \phi(I + Jr)$$

where  $r$  is a vector containing all the possible rates.

Let  $r_0$  be the point in which the time-derivative vanishes (fixed point). Then

$$r_0 = \phi(I + Jr_0)$$

We now want to analyze the stability of each fixed point. We write  $r = r_0 + \delta_r$  and we expand around  $r_0$

$$\tau \frac{dr}{dt} = \tau \frac{d}{dt}(r_0 + \delta_r) = \tau \frac{d\delta_r}{dt} = -r_0 - \delta_r + \phi(I + Jr_0) + \Delta\phi J \delta_r = -\delta_r + \Delta\phi J \delta_r$$

where the expansion is carried out as  $\phi(x + \epsilon) = \phi(x) + \epsilon\phi'(x)$ , with  $x = I + Jr_0$  and  $\epsilon = J\delta_r$ . Hence, we have obtained an equation that is linear in  $\delta_r$  and we can write it as

$$\frac{d\delta_r}{dt} = M\delta_r$$

for a matrix  $M = \Delta\phi J - \mathbb{1}$ . In one dimension, this equation will be  $\frac{d\delta_r}{dt} = \lambda\delta_r$ .

### 10.4.1 Simplest case

As an example, we consider a self-interacting cell (which could be a set of just Excitatory neurons, for example) which receives external inputs. Here the average firing rate is  $r(t)$  and, with the assumption that  $\phi$  is the identity, we have that  $r(t)$  evolves in time as

$$\begin{aligned}\tau \frac{dr}{dt} &= -r + \phi(I + Jr) \\ &= (J - 1)r + I\end{aligned}$$

There is only one fixed point,  $r_0 = \frac{I}{1-J}$ .  
The eigenvalue is  $\lambda = J - 1$ . So that

- If  $J > 1$ , the system is unstable, as  $r$  diverges to infinity
- If  $J < 1$ , then
  - If the network is Excitatory ( $0 < J < 1$ ), there is an amplification of inputs and a slow response.
  - If the network is inhibitory ( $J < 0$ ), there is attenuation of inputs, fast response.
- If  $J = 1$  we can integrate the differential equation and obtain that

$$r(t) = \frac{1}{\tau} \int I(t') dt'$$

### 10.4.2 Oculomotor Integrator

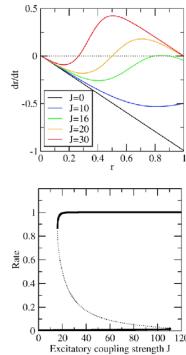
Is an example of perfect integrator ( $J = 1$ ) neuronal system. Cells like these are found in the brain; they activate with eye movement.

## 10.5 Non-linear E network - bistability

For certain models and protocols, the transfer function is sigmoidal. For an Excitatory network with non-linear (sigmoidal) transfer function, we can solve graphically the equation

$$\tau \frac{dr}{dt} = -r + \phi(I + Jr)$$

As shown in figure 10.5, as  $J$  increases, there are more solutions. If  $J$  is large enough the system becomes bistable (multiple stable points). The point in the middle is unstable, while the two on the sides are stable. The lower part of the figure represents values of  $r$  as a function of  $J$ , and the dashed line represents unstable solutions.



## 10.6 E network with slow negative feedback

Our goal is to obtain a generalization of the previous model, in order to introduce sources of slow negative feedback which can model phenomena as adaptation currents and synaptic short-term depression.

We introduce a simplified model for adaptation in which there is a negative-feedback term  $a$  ("negative" because the larger the firing rate  $r$ , the larger in absolute value  $a$ )

$$\begin{cases} \tau \frac{dr}{dt} = -r + \phi(I + Jr - a) \\ \frac{da}{dt} = -a + ra \end{cases}$$

This model can then be studied by finding the fixed points and by analyzing their stability, as done before.

### 10.6.1 Dynamics of E networks with slow negative feedback

Slow negative feedback leads to slow oscillations, in the form of alternations between an 'up state' and a 'down state'.

Slow negative feedback builds up during up state → leads to its destabilization.

Slow variable recovers/decays during down state → system goes back to up state.

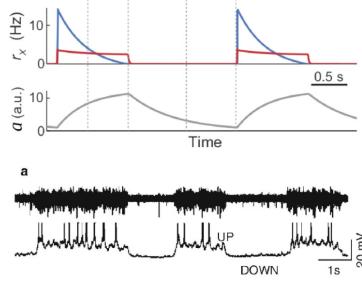


Figure 45: Graphs of  $a$ ,  $r$  and  $I$  with slow negative feedback. When the fixed point  $r$  (red) increases, so does  $a$ , gradually decreasing the net input (given by  $I - a$ ) into the cell, shifting down the curve of the graph of  $dr/dt$ , leading to a new fixed point  $r = 0$ . Now,  $a$  becomes small again, shifting up the graph and increasing the fixed point  $r$ . This repeats periodically leading to oscillations.

## 10.7 I networks with delays - oscillations

Another way to produce oscillations is by introducing a delay in the self-interaction of a cell in an I network

$$\tau_I \frac{dr_I}{dt} = -r_I + \phi_I \left( I_{IX} - J_{II} r_I(t - D) \right)$$

where  $D$  is a constant called the average synaptic delay.

## 10.8 E-I network - oscillations

We have an E-I network with parameters  $r_E$  and  $r_I$  (average firing rate of the Excitatory and Inhibitory cells, respectively). Both variables evolve in time according to the following equations

$$\begin{cases} \tau_E \frac{dr_E}{dt} = -r_E + \phi_E \left( I_{EX} - J_{EI} r_I + J_{EE} r_E \right) \\ \tau_I \frac{dr_I}{dt} = -r_I + \phi_I \left( I_{IX} - J_{II} r_I + J_{IE} r_E \right) \end{cases} \quad (19)$$

Fixed points are found through **nullclines**, i.e., by finding the zeroes of the RHS of each of the two above equations

$$\begin{aligned} r_I &= \phi_I \left( I_{IX} - J_{II}r_I + J_{IE}r_E \right) \\ \phi_I^{-1}(r_I) &= I_{IX} - J_{II}r_I + J_{IE}r_E \\ r_E &= \frac{\phi_I^{-1}(r_I) + J_{II}r_I - I_{IX}}{J_{IE}} \end{aligned}$$

where we usually take  $\phi$  to be sigmoidal. The graph of the function  $r_E = h(r_I) = \frac{\phi_I^{-1}(r_I) + J_{II}r_I - I_{IX}}{J_{IE}}$  defines the Inhibitory nullclines, which looks like the graph of a tangent with vertical asymptote at zero (and the function goes to  $-\infty$ ), so that each point lying on that curve will be a zero to the RHS of the second equation in 19.

Now, we also need the RHS of the first equation to be zero. To do so, we need to express  $r_I$  as a function of  $r_E$

$$\begin{aligned} r_E &= \phi_E \left( I_{EX} - J_{EI}r_I + J_{EE}r_E \right) \\ \phi_E^{-1}(r_E) &= I_{EX} - J_{EI}r_I + J_{EE}r_E \\ r_I &= \frac{J_{EE}r_E + I_{EX} - \phi_E^{-1}(r_E)}{J_{EI}} \end{aligned}$$

the Excitatory nullcline looks like a flipped graph of the tangent function, with vertical asymptote at zero (where the function goes to  $+\infty$ ). This curve however depends on the value of  $J_{EE}$ .

Now, we can find the point at which both  $\tau_E dr_E/dt$  and  $\tau_I dr_I/dt$  are zero (i.e. the points that belong to both nullclines). Depending on the Excitatory nullcurve there can be one or more of them.

If  $J_{EE}$  is small then we always have a single fixed point, if it is large we have multiple fixed points, where the middle one is typically unstable and the ones at the side are usually stable.

We now discuss the stability in the case of a single fixed point  $(r_E^0, r_I^0)$ .

$$\begin{aligned} r_E(t) &= r_E^0 + \delta r_E(t) \\ r_I(t) &= r_I^0 + \delta r_I(t) \end{aligned}$$

with  $\delta r_E(t)$  and  $\delta r_I(t)$  small, so that we can expand

$$\begin{aligned} \tau_E \frac{dr_E}{dt} &= \tau_E \frac{d}{dt} (r_E^0 + \delta r_E(t)) = \tau_E \frac{d\delta r_E}{dt} = - (r_E^0 + \delta_E) + \phi_E \left( I_{EX} - J_{EI}(r_I^0 + \delta_I) + J_{EE}(r_E^0 + \delta_E) \right) \\ &= - (r_E^0 + \delta_E) + \phi_E \left( I_{EX} - J_{EI}r_I^0 + J_{EE}r_E^0 \right) + \phi'_E \left( -J_{EI}\delta_I + J_{EE}\delta_E \right) \end{aligned}$$

so that

$$\begin{pmatrix} \tau_E d\delta_E/dt \\ \tau_I d\delta_I/dt \end{pmatrix} = \begin{pmatrix} \phi' J_{EE} - 1 & -\phi' J_{EI} \\ \phi' J_{IE} & -\phi' J_{II} - 1 \end{pmatrix} \begin{pmatrix} \delta_E \\ \delta_I \end{pmatrix} =: M \begin{pmatrix} \delta_E \\ \delta_I \end{pmatrix}$$

The goal is now to obtain

$$\frac{d}{dt} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (20)$$

where  $(x, y)$  is a linear combinations of  $(\delta_E, \delta_I)$ .

$$\begin{aligned} \det(M - \lambda \mathbb{1}) &= \det \begin{pmatrix} \frac{\phi' J_{EE} - 1}{\tau_E} - \lambda & -\frac{\phi' J_{EI}}{\tau_E} \\ \frac{\phi' J_{IE}}{\tau_I} & -\frac{\phi' J_{II} + 1}{\tau_I} - \lambda \end{pmatrix} \\ &= \left( \lambda - \frac{\phi' J_{EE} + 1}{\tau_E} \right) \left( \lambda + \frac{\phi' J_{II} + 1}{\tau_I} \right) + \frac{\phi' J_{EI}}{\tau_E} \frac{\phi' J_{IE}}{\tau_I} \end{aligned}$$

Setting this equation to zero allows to obtain the values of the eigenvalues  $\lambda_{1,2}$

$$\lambda_{1,2} = -b \pm \sqrt{b^2 - c}$$

so that the real part of  $\lambda_{1,2}$  only depends on the sign of  $b$ . For large  $J_{EE}$  and  $J_{II}$  we have

$$b \simeq \frac{1}{2} \left( -\frac{\phi'_E J_{EE}}{\tau_E} + \frac{\phi'_I J_{II}}{\tau_I} \right)$$

The fixed point is then stable if

$$\operatorname{Re}[\lambda_1] < 0, \quad \operatorname{Re}[\lambda_2] < 0$$

In this case both  $x$  and  $y$  as defined in 20 decay exponentially if the Imaginary part is zero, and decay exponentially with oscillations if the Imaginary part is non-zero.

The fixed point is unstable if

$$\operatorname{Re}[\lambda_1] > 0, \quad \operatorname{Re}[\lambda_2] > 0$$

If the fixed point is unstable, the dynamics of the firing rate will oscillate on a unit circle around the fixed point in the plane with axis  $r_E$  and  $r_I$ . We know that the dynamics will not be stationary because in every point different from the fixed one, at least one of the time derivatives of the firing rates is non-zero.

In the case of an unstable fixed point (we are still in the case in which there is only one fixed point) we have the unit circle because the transfer function  $\phi$  is bounded, and so it is not possible for the firing rate to diverge to  $+\infty$ .

Under instability, if the Imaginary part is zero we have large oscillations which are bounded (just like a cosine) by the transfer function  $\phi$ .

If the Imaginary part is non-zero, we obtain an imaginary cycle that is more localized around the fixed point. A sufficient condition for instability is that the largest eigenvalue has positive real part. It is possible to show that this happens when both  $J_{EI} \cdot J_{IE}$  and  $J_{EE}$  are sufficiently large.

## 10.9 I-I networks: winner takes it all

We consider the case in which there are two Inhibitory populations that mutually inhibit each other. Both populations receive external inputs.

We observe a winner-takes-it-all dynamics: if  $I_1 < I_2$ , due to the recurrent inhibition,  $r_1 < r_2$ , so that population 2 completely suppresses population 1.

# 11 Balanced Networks

## 11.1 Mechanisms of spontaneous activity in cortex

We have seen that the membrane potential as a function of time in neurons in the Cortex looks like a random process, with random inter-spike interval lengths.

We previously built a model of a single LIF neuron with noisy inputs coming from outside (we didn't question the source of these inputs). Our goal is to now provide a model that can describe the source of these noisy inputs.

In the brain we observe a wide distribution of firing rates, with the mean firing frequency being around 1 spike per second.

Firing is due to fluctuations in membrane potential, whose average value remains below firing threshold.

Moreover, we can observe that correlations across neurons are weak.

Our goal is to understand the origin of these properties.

## 11.2 Visualizing Network dynamics

A common graphical tool is the Raster plot, in which the vertical axes represents neurons, and we plot the activity of each cell as a function of time. Each dot represents the time at which one cell emitted a spike.

With the Raster plot we can compute the average population activity by averaging the number of spikes that each cell emitted in a give time interval.

Another graphical tool is given by the graph of the membrane potential and the current as a function of time. The current can usually be divided into two components. The membane potential indeed evolves in time according to

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

where the input can be divided, for example for an E-I network, into

$$I = I_E + I_I + I_X$$

where  $I_X$  is the external input. We can then plot these components.

We can also plot in an histogram the number of neurons spiking with a given rate, the inter-spike intervals (by computing the average distance between spikes in each cell) and the Coefficient of Variation (CV) of the Inter-spike intervals (ISI) which is the standard deviation over the mean of the ISI. In real neurons the CV is close to 1 (quite large).

All these graphical tools are shown in figure 46.

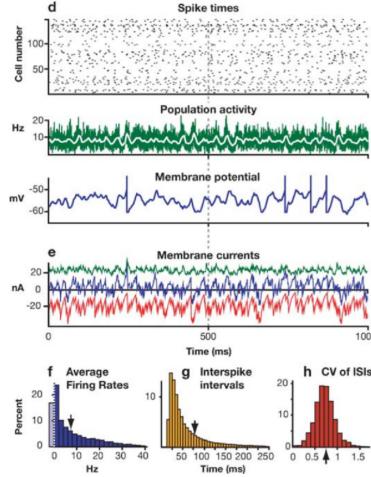


Figure 46: Graphical tools

### 11.3 Network models: Parts list

How many types of neurons?

How many neurons?

How are neurons connected? Which external inputs?

Which neuron model?

Which synapse model?

### 11.4 Randomly connected E-I network model

ER networks with two types of neurons (E and I) are popular models for local cortical networks.

Thousands of neurons of each type (80% E, 20% I)

Connectivity = ER (connection probability  $\sim 10\%$ )

External inputs = Uncorrelated Poisson processes

Neurons = Leaky Integrate-and-Fire

Synapses = Current based, delayed instantaneous.

There are  $N = N_E + N_I$  neurons, with  $N_E = 0.8N$  and  $N_I = 0.2N$ .

For the single-neuron model we use LIF, so that the membrane potential evolves according to the following equation

$$\tau_i \frac{dV_i}{dt} = -V_i + \sum_i I_{iX} + \sum_j J_{ij} S_j(t)$$

where  $V_i$  is the membrane potential of neuron  $i$ ,  $I_{iX}$  is the external input to neuron  $i$ ,  $S_j(t) = \sum_k \delta(t_j^k - D)\tau_i$  ( $D$  being the latency/delay) is the spiking activity of pre-synaptic neuron  $j$  as a function of time and  $J_{ij}$  is the connectivity strength between neuron  $j$  and neuron  $i$ .

Additionally, we have parameters  $V_T$ ,  $V_R$  and  $\tau_{rp}$  so that, each time the membrane potential exceeds  $V_T$  the neuron emits a spike, after which the potential is reset to  $V_R$ , where it stays for  $\tau_{rp}$  units of time.

We also assume that the strength of connection of group  $A$  and  $B$  (with  $A, B \in [\{E, I\} \times \{E, I\}]$ ) is fixed. The variability in the connection strength between a specific pair of neurons  $i, j$  is given by an entry of the adjacency matrix  $C_{ij}$ , so that

$$J_{ij} = J_{AB} c_{ij} \quad \text{with } A, B \in [\{E, I\} \times \{E, I\}] \text{ and for all } i, j \in \{1, \dots, N\}$$

where  $c_{ij} \in [0, 1]$  (neurons are either connected or not) and there is random connectivity between cells (ER network) so that  $c_{ij} = 1$  with probability  $p$ .

We can use the same approach to rewrite the external input of each neuron as

$$I_{iX} = J_{aX} \sum_{j=1}^{K_{aX}} \sum_k \tau_i \delta(t - t_j^k)$$

where  $J_{aX}$  is the amplitude of Post-synaptic potential triggered by spikes from  $K_{aX}$  ‘external’ neurons (not part of the E-I network). Spike times  $t_j^k$  for each of the inputs are generated as independent Poisson processes with rate  $r_X$ .

#### 11.4.1 Orders of magnitude of parameters for local cortical networks

Here  $N \sim 10^4 - 10^5$ . Regarding single-neuron parameters we have  $\tau_i \sim 10 - 20\text{ms}$ ,  $V_T = 20\text{mV}$ ,  $V_R = 0 - 18\text{mV}$ . Connectivity parameters:  $p \sim 0.1$ ,  $K_{EE} = pN_E$  is the number of connections between pre-synaptic excitatory cells and post-synaptic Excitatory cells while  $K_{IE} = pN_E$  is the number of connection between pre-synaptic Excitatory cells that an Inhibitory neuron has. Also  $J_{AB} \sim 0.1 - 1\text{mV}$ .

### 11.5 Simulations of E-I networks

With realistic parameters and sufficiently low inhibition, we can reproduce many of the patterns observed in the empirical data (low mean firing rates, broad distribution of firing rates across the network, irregular spiking trains).

### 11.5.1 Synaptic inputs: means and fluctuating terms

Goal: computing the currents that each neuron receives, by decomposing it in a common term across neurons along with a neuron-specific fluctuation term. We'll argue that, if the fluctuation is of a given size, then we can reproduce the features that we observe in experiments.

In particular, we are interested in finding the average membrane potential across cells and across time. This will be determined by the average input (later denoted as  $\mu_E$ ). When the average membrane potential is below the LIF threshold, we need temporal fluctuations (later denoted as  $\sigma$ ) that are big enough to generate spikes. Another property we will investigate is the neuron-to-neuron fluctuations in the average input. If these are large enough, than we can observe significant fluctuations in the activity in the network, producing a broad distribution of firing rates.

Consider Excitatory neuron  $i$ . Then the input it receives is given by

$$\begin{aligned} I_i(t) &= J_{EX}\tau_{mE} \sum_{j \in X_E} c_{ij}S_j(t) + J_{EE}\tau_{mE} \sum_{j \in E} c_{ij}S_j(t) - J_{EI}\tau_{mE} \sum_{j \in I} c_{ij}S_j(t) \\ &= \mu_E + \Delta\mu_i + \sigma\tau_{mE}\eta_i(t) \end{aligned}$$

where

- $\mu_E$  is the common term across neurons, given by the current averaged over all cells,

$$\begin{aligned} \mu_E &= \mu_{EE} + \mu_{EI} + \mu_{EX} \\ &= k\tau_{mE}(J_{EX}r_X + J_{EE}r_E - J_{EI}r_I) \end{aligned}$$

(here  $K$  being the average number of input per cells)

- $\Delta\mu_i$  is a time-independent term that accounts for the deviation of the current of each neuron from the average  $\mu_E$ , (i.e.,  $\Delta\mu_i$  accounts for the difference between  $\mu_E$  and the time-averaged current of neuron  $i$ )
- the third component accounts for the fluctuation in time of the current.

We will compute one of these sums, the other are done in a similar way. Let  $S$  be defined as

$$S := J_{EE}\tau_{mE} \sum_{j \in E} c_{ij}s_j(t)$$

with  $s_j(t) = S_j(t)/\tau_{mE}$ . Then, for a fixed time  $t$ ,  $S \sim \mathcal{N}(\mu_{EE}, \sigma_{EE})$ , by the central limit theorem. We have

$$\begin{aligned} \mu_{EE} &= J_{EE}\tau_{mE}N_E \cdot p \cdot r_E \\ \sigma_{EE}^2 &= J_{EE}^2\tau_{mE}^2N_Ep\left[(1-p)r_E^2 + \Delta r_E^2\right] \end{aligned}$$

that is because  $Var(XY) = Var(X)Var(Y) + Var(X)Mean(Y)^2 + Var(Y)Mean(X)^2$  for 2 independent random variable  $X$  and  $Y$ , so that  $Mean(c_{ij}) = p$ ,  $Var(c_{ij}) = p(1-p)$ ,  $Mean(s_j(t)) = r_E$ ,  $Var(s_j(t)) = \Delta r_E^2$  is the standard deviation of the distribution of rates for a given unknown parameter  $\Delta r_E$ .

The excitatory input that excitatory neuron  $i$  receives is given by

$$\begin{aligned} \mu_{EE}^i(t) &= J_{EE}\tau_{mE} \sum_{j=1}^N c_{ij}s_j(t) \\ &= \mu_{EE}^i - \mu_{EE}^i + \mu_{EE}^i(t) \\ &= \mu_{EE} + (\mu_{EE}^i - \mu_{EE}) + (\mu_{EE}^i(t) - \mu_{EE}^i) \end{aligned} \tag{21}$$

where  $\mu_{EE}^i$  is the average input across time for neuron  $i$ ,

$$\mu_{EE}^i = \langle \mu_{EE}^i(t) \rangle_t = J_{EE}\tau_{mE}p \cdot r_e$$

while  $\mu_{EE}$  is the average input across neuron **and** across time.

Let  $\Delta\mu_{EE}^i = \mu_{EE}^i - \mu_{EE}$ . Then

$$\begin{aligned}\langle \Delta\mu_{EE}^i \rangle_i &= 0 \\ \langle (\Delta\mu_{EE}^i)^2 \rangle_i &= \sigma_{EE}^2 = J_{EE}^2 \tau_{mE}^2 N_E p [(1-p)r_E^2 + \Delta r_E^2]\end{aligned}$$

Hence, we have computed the first two out of the three terms in equation 21. To compute the last term, we need to observe that

$$\mu_{EE}^i(t) - \mu_{EE}^i \sim \mu_{EE}(t) - \mu_{EE}$$

meaning that the fluctuations across cells are negligible. This means that

$$\begin{aligned}\langle \mu_{EE}(t) - \mu_{EE} \rangle_t &= 0 \\ \langle (\mu_{EE}(t) - \mu_{EE})^2 \rangle_t &= \sigma_{EE}^2 \tau_E \eta_i(t)\end{aligned}$$

where  $\eta(t)$  is whit noise with zero mean and unit variance for a fixed time  $t$ . Otherwise it holds  $\langle \eta(t)\eta(t') \rangle = \delta_{t,t'}$ .

**Finally**, we can compute mean and variances of the input  $I_i(t)$  by considering all of the sums in the equation for  $I_i(t)$  (at the beginning of the section) and sum their means and variances,

$$\begin{aligned}\mu_E &= \mu_{EE} + \mu_{EX} + \mu_{IE} = \tau_{mE} N_E \cdot p (J_{EE} r_E + m j_{EX} r_X - J_{EI} r_I) \\ \langle \Delta\mu_i \rangle^2 &= \tau_{mE}^2 N_E \cdot p [J_{EE}(r_E^2 + \Delta r_E^2) + J_{EX}(r_X^2 + \Delta r_X^2) + J_{EI}(r_I^2 + \Delta r_I^2)]\end{aligned}$$

### 11.5.2 Large $K$ limit:

We are interested in the limit of large  $K = N_E \cdot p$  (average number of input per cell). What we want to investigate is how big  $J$  needs to be, in order to have a model that explains empirical observations in cortex (irregular firing and broad distribution of rates, shown in figure 46). The magnitude of  $J$  for large  $K$  is relevant because in the equation for  $\mu_E$  (beginning of section 11.5.1  $K$  and  $J$  are multiplied).

We'll express  $J$  as a function of  $K$ .

Consider  $J$  to be proportional to  $1/K$

$$J = \frac{J_0}{K}$$

Then

$$\begin{aligned}\mu_E &= K \tau_{mE} J \left( \frac{J_{EX}}{J} r_X + \frac{J_{EE}}{J} r_E - \frac{J_{EI}}{J} r_I \right) \\ &= \tau_{mE} J_0 \left( \frac{J_{EX}}{J} r_X + \frac{J_{EE}}{J} r_E - \frac{J_{EI}}{J} r_I \right)\end{aligned}$$

is finite, as it does not depend on  $K$ . Regarding the variance

$$\begin{aligned}\langle \Delta\mu_i^2 \rangle &= K \tau_{mE}^2 J^2 \left( \frac{J_{EX}^2}{J^2} (r_X^2 + \Delta r_X^2) + \frac{J_{EE}^2}{J^2} (r_E^2 + \Delta r_E^2) + \frac{J_{EI}^2}{J^2} (r_I^2 + \Delta r_I^2) \right) \\ &= \tau_{mE}^2 \frac{J_0^2}{K} \left( \frac{J_{EX}^2}{J^2} (r_X^2 + \Delta r_X^2) + \frac{J_{EE}^2}{J^2} (r_E^2 + \Delta r_E^2) + \frac{J_{EI}^2}{J^2} (r_I^2 + \Delta r_I^2) \right) \\ &<< 1\end{aligned}$$

Analogously,

$$\sigma_E^2 \propto \frac{J_0^2}{K}$$

Therefore, fluctuations are really small, which is a problem because we both need the mean and variance to be of order one to guarantee an heterogeneity in the neuronal firing.

We thus need a different scaling. Let

$$J = \frac{J_0}{\sqrt{K}}$$

Then

$$\mu_E = \tau_{mE} \sqrt{K} J_0 \left( \frac{J_{EX}}{J} r_X + \frac{J_{EE}}{J} r_E - \frac{J_{EI}}{J} r_I \right)$$

and

$$\begin{aligned} \langle \Delta \mu_i^2 \rangle &= \tau_{mE}^2 J_0 \left( \frac{J_{EX}^2}{J^2} (r_X^2 + \Delta r_X^2) + \frac{J_{EE}^2}{J^2} (r_E^2 + \Delta r_E^2) + \frac{J_{EI}^2}{J^2} (r_I^2 + \Delta r_I^2) \right) \\ \sigma_E^2 &\propto \tau_{mE} J_0 \end{aligned}$$

We thus obtain that fluctuations do not vanish, but the mean input could potentially explode. To prevent that from happening we need to set the factor that multiplies  $\sqrt{K}$  in the expression for  $\mu_E$  to be of order  $1/\sqrt{K}$ , so that the mean will be of order 1

$$\left( \frac{J_{EX}}{J} r_X + \frac{J_{EE}}{J} r_E - \frac{J_{EI}}{J} r_I \right) \propto \frac{1}{\sqrt{K}} \quad (22)$$

Importantly, this does **not** make the expression for the variances ( $\langle \Delta \mu_i^2 \rangle$  and  $\sigma_E^2$ ) converge to 0, as the factors there only contain + signs.

### 11.5.3 Balance conditions

Equation 22 represents the condition for Excitatory populations under which we have a broad distribution of firing rates with non-zero oscillations. We can give the same condition for Inhibitory populations

$$\begin{cases} j_{EX} r_X + j_{EE} r_E - j_{EI} r_I = O\left(\frac{1}{\sqrt{K}}\right) = 0 \\ j_{IX} r_X + j_{IE} r_E - j_{II} r_I = O\left(\frac{1}{\sqrt{K}}\right) = 0 \end{cases}$$

where  $j_i := \frac{J_i}{J}$ . These conditions provide a linear dependence of rates ( $r_E$  and  $r_I$ ) on external inputs ( $r_X$ ). We only need to have  $j$ 's such that the system is solvable. As long as they exist, the dynamics will drive the network towards this state. This means, that the rates can be directly found by the Balance condition system, without the need to solve differential equations.

## 11.6 Evidence of synaptic scaling in culture preparations

Researchers focused on growing various in-culture populations of neurons. They observed that, having low density (meaning few neurons in the culture) for a culture is equivalent to having a small  $K$ , while high density is equivalent to a large value of  $K$ . This allows to experimentally produce networks with different values of  $K$ . Measurements for the Post-synaptic Excitatory Potential and Post-synaptic Inhibitory potential are shown in figure 47. We can observe that the network has adjusted itself in such a way to have a larger post-synaptic potential for smaller  $K$ . The amplitude of the upper graph in the left column is what we denoted with  $J$ , showing that  $J$  decreases the larger  $K$  is.

This is also confirmed by the plot on the right part of the figure, in which measurements were performed for many cultures. The result is that the Post-Synaptic Potential (i.e.  $J$ ) decreases with  $K$  as about  $J \sim K^{-0.6}$ , so very close to the inverse of the square root.

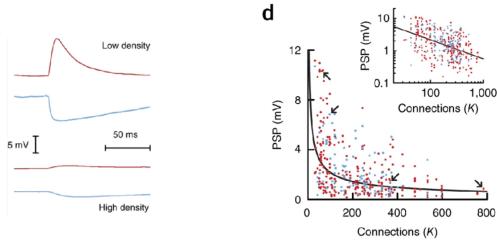


Figure 47: Experiments with cultures

## 12 The ring model

The ring model uses rate models to model the fact that neurons are tuned for specific features.

### 12.1 Examples

#### 12.1.1 Primary Visual Cortex

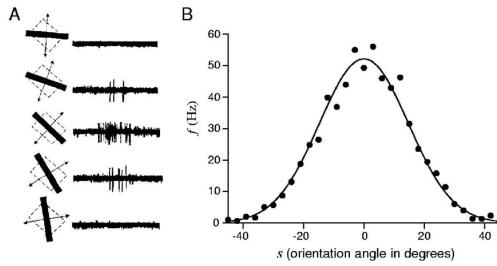


Figure 48: Spiking frequency of a neuron as a function of the orientation of the visual stimulus (bar) to which the animal is subject

We can classify each neuron based on the angle at which it has a peak in spiking activity (in the picture it is 0 degrees for the considered neuron). Then, with a fixed stimulus, it is also possible to plot, for each angle  $\theta \in [0, 2\pi]$ , what is the average firing frequency computed over all neurons whose peak is at  $\theta$ .

#### 12.1.2 Motor Cortex

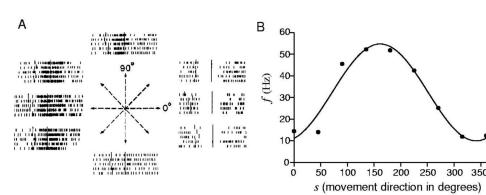


Figure 49: Experiment with a monkey. At different times one of the bulb positioned at the end of the lines crossing the circle lights up, and the monkey has to reach for it. Measuring the response of a single neuron to each of the movements, we find a tuning for a specific orientation.

#### 12.1.3 Prefrontal Cortex

See figure 50.

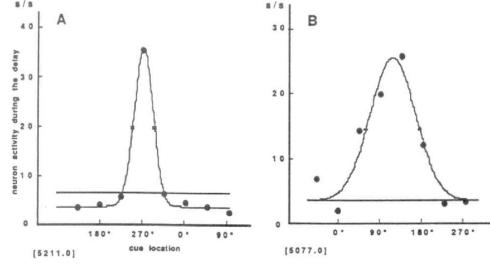
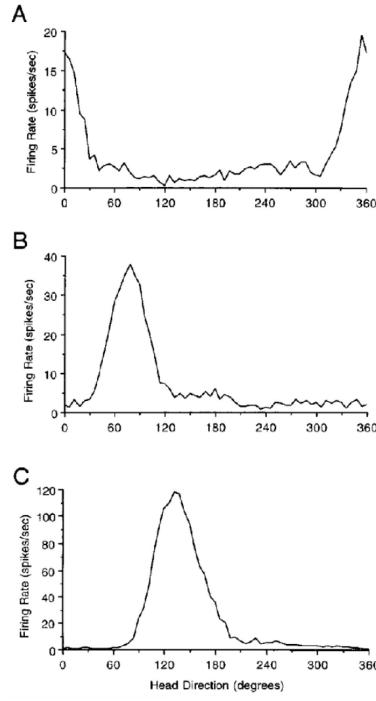


Figure 50: Experiment with a monkey. At different times one of the bulb positioned at the end of the lines crossing the circle lights up, and the monkey has to follow it with its eyes. Measuring the response to a single neuron to each of the eyes' movements, we again find a tuning for a specific orientation.

#### 12.1.4 Head-direction cells

Head-direction are mostly found in the Hippocampus and related areas. The experiment involves an animal freely moving in an arena; the movement of the head is recorded with respect to some external, fixed landmark. We can plot the firing rate of neurons as a function of the heading direction of the animal (shown in figure 12.1.4) and we can observe how specific neurons are tuned for head directions.



## 12.2 Rate model for spatial selectivity

We have a variable  $x$  which orders the neurons. For example,  $x$  could be the bar orientation; in that case we define  $r(0, t)$  to be the average firing rate of all the neurons which are classified with  $x = 0$ . In general the firing rate for neurons classified with  $x$  evolves in time according to

$$\tau \frac{dr(x, t)}{dt} = -r(x, t) + \phi(I_X(x, t) + \int J(|x - y|)r(y, t)dy)$$

where  $\phi()$  is the input-output transfer function,  $I_X(x, t)$  is the external drive/input to the network. The integral term is the total input that neuron  $x$  receives from all other neurons in the network. In particular, for a specific neuron  $y$ , the input that  $x$  receives from it is given by the firing rate of  $y$  weighted by the synaptic

strength between them (i.e.,  $J(|x - y|)$ ).

Importantly, the variable  $x$  is not necessarily representing an angle; for example, there is a model for neurons called "place-cells" which are tuned for the position of the arena in which the animal can move. Here  $x$  represents 2d spatial coordinates.

### 12.3 The ring model

In the ring model  $x$  represents an angle, so that  $x \in [-\pi, \pi]$ . Moreover, we take  $\phi$  to be threshold-linear

$$\phi(x) = \max\{0, x\}$$

Moreover, we take  $J()$  to be

$$J(|x - y|) = J_0 + J_1 \cos(x - y) \quad (23)$$

#### 12.3.1 Connectivity in the $J_0 - J_1$ plane

We can plot equation 23 as a function of  $J_0$  and  $J_1$ . In all cases, it is a sinusoidal function. The only difference is given by whether it crosses zero or not.

If  $|J_1| < |J_0|$ , then the connectivity function never crosses zero, and it is always positive.

For  $J_1 > J_0 > 0$  the  $J()$  function crosses zero infinitely many times. This is shown in figure 51.

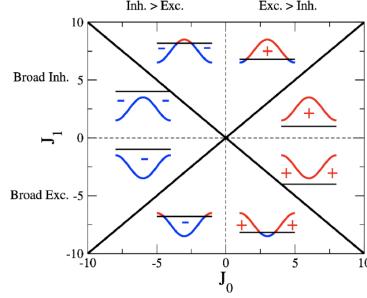


Figure 51:

#### 12.3.2 Analysis of the model

Our goal is now to study the behavior of the model as a function of the connectivity.

The equation now becomes

$$\tau \frac{dr(x, t)}{dt} = -r(x, t) + \max\left\{0; I_X(x, t) + \frac{1}{2\pi} \int_{-\pi}^{+\pi} (J_0 + J_1 \cos(x - y)) r(y, t) dy\right\} \quad (24)$$

which we solve using a Fourier transform

$$r(x, t) = \sum_{n=0}^{+\infty} r_n(t) \cos(nx - \psi_n(t)) \quad (25)$$

Plugging equation 25 into equation 24 implies

$$\begin{aligned} \tau \frac{dr(x, t)}{dt} &= -r(x, t) + \max\left\{0; I_X(x, t) + \frac{1}{2\pi} \int_{-\pi}^{+\pi} (J_0 + J_1 \cos(x - y)) \sum_{n=0}^{+\infty} r_n(t) \cos(ny - \psi_n(t)) dy\right\} \\ &= -r(x, t) + \max\left\{0; I_X(x, t) + J_0 \sum_n r_n(t) \frac{1}{2\pi} \int_{-\pi}^{+\pi} \cos(x - y) \cos(ny - \psi_n(t)) dy + \right. \\ &\quad \left. + J_1 \sum_n r_n(t) \frac{1}{2\pi} \int_{-\pi}^{+\pi} \cos(x - y) \cos(ny - \psi_n(t)) dy\right\} \\ &= -r(x, t) + \max\left\{0; I_X(x, t) + J_0 r_0(t) + J_1 r_1(t) \pi \cos(x - \psi_1(t))\right\} \end{aligned} \quad (26)$$

because  $\int_{-\pi}^{+\pi} \cos(ny - \psi_n(t)) dy \neq 0$  only if  $m = 0$  (and we incorporated  $\psi_0(t)$  into  $r_0(t)$ ) and because  $\int_{-\pi}^{+\pi} \cos(x - y) \cos(ny - \psi_n(t)) dy \neq 0$  only if  $n = 1$ .

Now, we can write

$$\begin{aligned} r_0(t) &= \int_{-\pi}^{+\pi} \frac{r(x, t)}{2\pi} dx \\ r_1(t) &= \int_{-\pi}^{+\pi} \frac{r(x, t)}{2\pi} \cos(x - \psi_1(t)) dx \\ 0 &= \int_{-\pi}^{+\pi} \frac{r(x, t)}{2\pi} \sin(x - \psi_1(t)) dx \end{aligned}$$

Now, denote by (1) equation 26. Then integrating it yields

$$\begin{aligned} \int_{-\pi}^{+\pi} \frac{(1)}{2\pi} dx &\implies \tau \frac{dr_0(t)}{dt} = -r_0(t) + \int_{-\pi}^{+\pi} \frac{1}{2\pi} \max\{0, \dots\} dx \\ \int_{-\pi}^{+\pi} \frac{(1)}{2\pi} \cos(x - \psi_1(t)) dx &\implies \tau \frac{dr_1(t)}{dt} = -r_1(t) + \int_{-\pi}^{+\pi} \cos(x - \psi_1(t)) \frac{1}{2\pi} \max\{0, \dots\} dx \\ ????? &\implies \tau \frac{d\psi(t)}{dt} r_1 = -r_1 + \int_{-\pi}^{+\pi} \sin(x - \psi_1(t)) \frac{1}{2\pi} \max\{0, \dots\} dx \end{aligned} \quad (27)$$

Importantly, every higher order terms (e.g.,  $r_2, r_3$ ) are uniquely determined by  $r_0, r_1, \psi(t)$ . This is because the term inside the transfer function  $\max\{0, \dots\}$  only depends on  $r_0, r_1, \psi_1$  even for higher order terms. These quantities also have a simple interpretation:

- $r_0$  is the average firing across the population
- $r_1$  is how modulated w.r.t. a constant the population is. So for example if the function  $r(x)$  is constant,  $r_0 \neq 0$  and  $r_1 = 0$ , while if the function is modulated, e.g.  $r(x) = e^{-x^2}$ , then  $r_0 \neq 0$  and  $r_1 \neq 0$ ;  $r_1$  represents the magnitude of the deviation from a constant value.
- $\psi_1(t)$  represents the location of the peak of the population (for  $r(x) = e^{-x^2}$  the peak is at 0).

Often, we will have that the rate function  $r(x)$  will be zero except for a bump of amplitude  $r_1$  at  $\psi_1$ .

## 12.4 Uniform inputs - Phase diagram of the model

As for rate models and E networks, we can analyze fixed points and their stability.

We can show that there is a uniform stationary state in which  $r_0 = R_0$  is a constant and  $r_1 = \psi_1 = 0$ , so a network in which the neurons are uniformly active with firing rate  $R_0$ .

If  $J_0 > 1$  the uniform state is unstable (runaway activity in the network), but if  $J_0 < 1$  the network is stable. If  $J_1 < 2$  the only stable state in the network is the state in which  $r(x)$  is constant. If  $J_1 > 2$  instead, the network spontaneously generates a bump of activity (so  $r(x)$  looks something like  $e^{-x^2}$ ).

This is all shown in figure 52.

The fixed points are, as usual, found by setting to zero the right-hand side of equations 27. Let, for example,  $\hat{r}_0$  be the fixed point for  $r_0(t)$ . Stability analysis is then performed by setting  $r_0(t) = \hat{r}_0 + \delta_{r_0(t)}$  and expanding in  $\delta_{r_0(t)}$  in the equation for  $r_0(t)$ .

## 12.5 Response of the model to tuned inputs

Consider a time-independent input having the form of the connectivity function  $J()$

$$I(x) = I_0 + I_1 \cos(x)$$

Our goal is to compute  $r(x)$ , which we assume to be positive for all  $x$  (this will allow us to have the identity as a transfer function). Since we are looking for the stationary state, we set the time derivative of  $r$  to zero

$$\begin{aligned} 0 &= -r(x) + (I_0 + I_1 \cos(x) + J_0 r_0 + J_1 r_1 \cos(x)) \\ r(x) &= I_0 + I_1 \cos(x) + J_0 r_0 + J_1 r_1 \cos(x) \end{aligned} \quad (28)$$

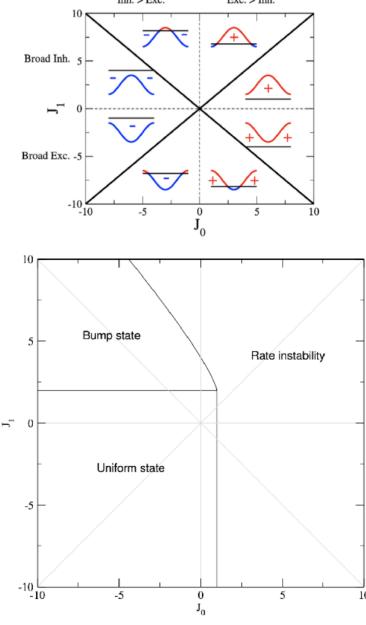


Figure 52:

Now, given the Fourier transform of  $r(x)$  -equation 25- we obtain that  $r_n(x) = 0 \ \forall n > 1$ , because in the expression for  $r(x)$  we just derived, the only dependency from  $\cos(nx)$  is with  $n = 0$  and  $n = 1$ .

Now, substituting the transform of  $r(x)$  in equation 28 implies

$$\sum_n r_n \cos(nx - \psi_n) = I_0 + I_1 \cos(x) + J_0 r_0 + J_1 r_1 \cos(x)$$

which we can integrate to obtain the values of  $r_0$  and  $r_1$

$$\begin{aligned} r_0 &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} \sum_n r_n \cos(nx - \psi_n) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} I_0 + I_1 \cos(x) + J_0 r_0 + J_1 r_1 \cos(x) dx \\ &= I_0 + J_0 r_0 \end{aligned}$$

so that

$$r_0 = \frac{I_0}{1 - J_0}$$

and  $r_1$

$$\begin{aligned} r_1 &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} \cos(x) \sum_n r_n \cos(nx - \psi_n) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} (I_1 + J_1 r_1) \cos^2(x) dx + \frac{1}{2\pi} \int_{-\pi}^{+\pi} (I_0 + J_0 r_0) \cos(x) dx \\ &= \frac{I_1 + J_1 r_1}{2\pi} \pi \\ &= \frac{I_1 + J_1 r_1}{2} \end{aligned}$$

Hence

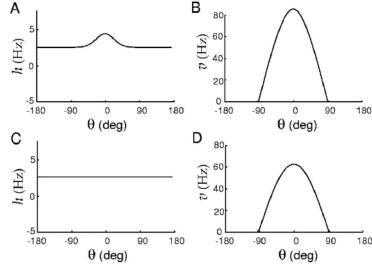
$$r_1 = \frac{I_1 + J_1 r_1}{2} \quad \Rightarrow \quad r_1 = \frac{I_1}{2 - J_1}$$

Hence, every time  $J_1 < 1$  we have that  $r_1 < I_1$ , which means that the tuned component in the input is suppressed by the recurrent connectivity.

If  $1 < J_1 < 2$ , then the tuned component in the input is amplified by the recurrent connectivity.

If  $J_1 > 2$ , then the tuned component is amplified, and maintained in short-term memory after the stimulus is removed.

One of the reasons for which this model was created to explain what happens in the Primary Visual cortex. The output of the model (right column in figure 12.5) is indeed the same regardless of the amplitude in modulation of the input (left column in figure 12.5). This model, however, turns out to be inaccurate for that part of the brain.



### 12.5.1 A ring attractor in the fly

Another feature in which the ring model has found success regards the ellipsoid body (EB) structure in the fly brain with a ring structure.

In particular, the head orientation of the fly is represented as an EB bump of activity.

Moreover, in the experiment, optogenetics (stimulating neurons with light) was used to provide an input tuned to a different angle from the one corresponding to the activation bump peak for a constant input.

Researchers found that the activation bump peak changed accordingly to where the input was tuned through optogenetics, and the bump peak remained there even after the stimulation was removed.

# 13 Single neuron coding I: Quantifying the dependence of spike trains on external stimuli

We here tackle a different perspective of neuronal activity. We are interested, in particular, in investigating what kind of stimuli drive activity in specific neurons, and investigating why neurons behave in a specific way.

## 13.1 Flow of information in visual detection

It is possible to define a hierarchical structure in the brain on how information is processed and transferred through various parts of the brain. This argument is corroborated by the latency observed in the spiking activity caused by the stimulus to the various parts of the brain, schematized in figure 53.

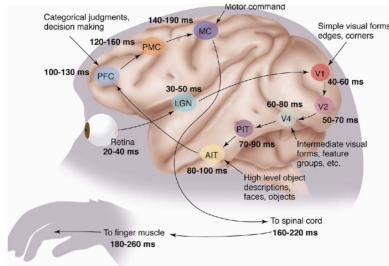


Figure 53: Schematic of information flow in the brain during a standard reaction-detection task in a monkey. The retina transfers encoded visual information to the thalamus (LGN), from which it flows to the Primary Visual Cortex (V1). Here we have a series of visual-information-processing areas that ends with the AIT. This sends a signal to the Pre-Frontal Cortex where categorical judgment takes place. From there a signal is generated to the Motor Command which is then used to drive the muscle.

## 13.2 Response to a single stimulus: PSTH and Fano Factor

We can plot in an histogram the activity of a single neuron in response to a stimulus, as captured in figure 54.

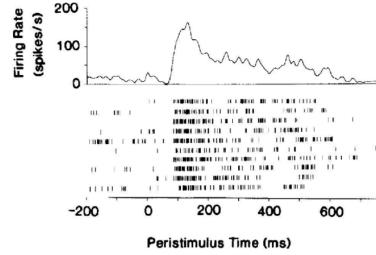


Figure 54:

There is a trial-to-trial variability, i.e., if we present the stimulus multiple times, the activity will slightly change. Hence, we can measure the Fano Factor ( $FF$ ), that measure the variance in the spike count over the mean. In particular it measures the spike variability in a time bin from trial to trial. For a pure Poisson Process  $FF = 1$ . However, that's not what is observed in the brain.

A good model that can reproduce the Fano Factors observed in figure 55 (so  $FF$  larger than 1) is a Poisson process whose parameter changes in time  $r \sim \text{Poisson}(\lambda(t))$ .

The modulation of the firing rate  $\lambda(t)$  in time can be justified by, for example, how the movement modulates activity in Visual Cortex.

## 13.3 Firing rates vs external stimuli: Continuous case

If we now consider the mean firing rate of a neuron, we can compute tuning curves, which quantify how the firing rate of a neuron depends on a continuous parameter characterizing the stimulus.

Tuning curves are often bell-shaped (as we have seen for the Visual Ocrtex, Motor Cortex and other examples,

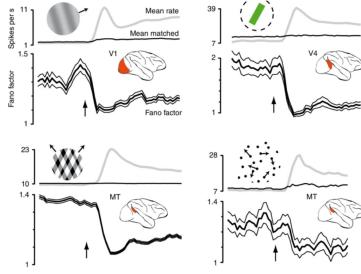


Figure 55: Fano Factor of various experiments in different parts of the brain. Before a stimulus is shown, there is a high neuron-to-neuron variability in the firing rate, which are generally low. When a stimulus is shown, some neurons increase their rates, while others decrease their rates with the average overall rate increasing. Fano Factors decreases, but remains above 1.

section 12.1).

Often, however, the stimulus is not well-described by a continuous value.

### 13.4 Rat olfactory cortex - odors

Odors represent a discrete set.

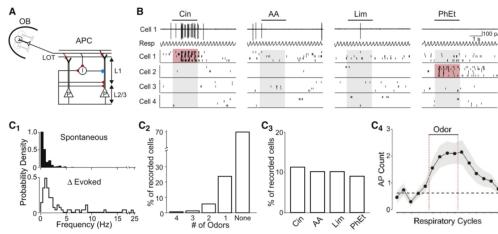


Figure 56: Many neurons do not respond to the odor, while there is a small fraction of cells that activates in response to a specific odor.

### 13.5 Single-neuron encoding of dynamic stimuli

We consider an experimental setup in which an animal is subject to a stimulus that changes with time (e.g. the contrast of an image, the intensity of an odor). We repeat many times the experiment in order to record the average over trials of the instantaneous firing rate of the response of the specific neuron in that instant of time. Our goal is to **predict the firing rate based on the stimulus**.

### 13.6 Characterizing input/output transformation: Volterra series

We represent the firing rate as a function of the stimulus  $S(t)$

$$r(t) = F(S(t))$$

which we write as a series

$$\begin{aligned} r(t) &= F(s(t)) \\ &= h_0 + \int_0^{+\infty} h_1(t_1)S(t-t_1)dt_1 + \int_0^{+\infty} h_2(t_1, t_2)S(t-t_1)S(t-t_2)dt_1dt_2 + \dots \end{aligned}$$

where the functions  $h_i()$  are Volterra currents. Our goal is to find "good" functions  $h()$  that give an accurate representation of the neuron.

**Example 4** (Artificial neuron). We have  $h_0 = 0$ ,  $h_1(t) = \frac{1}{T}e^{-\frac{t}{T}}$ . We also take  $h_n = 0$  if  $n \geq 2$ . We define  $s(t) = A\mathbb{1}_{t \geq 0}$ . In this case

$$r(t) = \int_0^{+\infty} h_1(x)s(t-x)dx = \int_0^t \frac{1}{T}e^{-\frac{x}{T}}Adx = A[1 - e^{-\frac{t}{T}}]$$

The integral stops at  $t$  because  $s(t-x) \neq 0 \iff t-x \geq 0 \iff t \geq x$ .

Our goal, in general, is to infer the values of the functions  $h$ , which are usually not available in experiments. Indeed, if we know the  $h()$  function, we can compute the firing rate.

### 13.7 Wiener series

We take the stimulus to be white noise, i.e.

$$\begin{aligned}\langle S(t) \rangle_t &= 0 \\ \langle S(t)S(t') \rangle_t &= S\delta(t-t')\end{aligned}$$

Our goal is to compute the  $h()$  functions. We can now write the firing rate as

$$r(t) = H_0 + H_1 + H_2 + \dots$$

where  $H_0 = h_0$ ,  $H_1 = \int_0^{+\infty} h_1(x)S(t-x)dx$  and so on. Now, since  $S()$  are random variable, so are the  $H_i$ 's. We want to compute the average firing rate over time

$$\begin{aligned}\langle r \rangle_t &= \frac{1}{T} \int_0^T r(t)dt \\ &= h_0 + \frac{1}{T} \int_{x=0}^{x=+\infty} h_1(x)dx \int_0^T S(t-x)dt + \langle H_2 \rangle_t + 0 + \langle H_4 \rangle_t + 0 + \dots \\ &= h_0 + \langle H_2 \rangle_t + \langle H_4 \rangle_t + \dots\end{aligned}$$

because  $\int_0^T S(t-x)dt = \langle S \rangle_t = 0$ . So all the even terms will be non-zero. This is not ideal, because we have an infinite number of unknowns.

We can also measure the following

$$\begin{aligned}\langle r(t)S(t-\tau) \rangle &= \langle H_0S(t-\tau) \rangle_t + \langle H_1S(t-\tau) \rangle_t + \langle H_2S(t-\tau) \rangle_t + \langle H_3S(t-\tau) \rangle_t \\ &= 0 + h_1(\tau)S + 0 + S^2 \int h_3(\cdot)\end{aligned}$$

However, we would need to compute an infinite number of such averages.

To simplify, we can assume that  $r(t)$  is given by  $H_0$  and  $H_1$ , while all the higher order terms are negligible. We thus assume the following

$$\begin{aligned}\langle r(t) \rangle_t &= H_0 \\ \langle r(t)S(t-\tau) \rangle_t &= Sh_1(\tau)\end{aligned}$$

Adding an extra term, i.e., assuming  $r(t) = H_0 + H_1 + H_2$ , we obtain

$$\begin{aligned}\langle r(t) \rangle_t &= H_0 + \langle H_2 \rangle \\ \langle r(t)S(t-\tau) \rangle_t &= Sh_1(\tau) \\ \langle r(t)S^2(t-\tau) \rangle_t &\propto h_2()\end{aligned}$$

So, every time we add a term in the expression of  $r$ , we need to re-adjust all the other terms, and this is not ideal. Our representation is not stable.

The idea of the Wiener kernel is to define a different series in which the terms are statistically independent

$$r(t) = G_0 + G_1 + G_2 + \dots$$

$$G_0 = g_0$$

$$G_1 = \int_0^{+\infty} g_1(x)S(t-x)dx$$

$$G_2 = \int_0^{+\infty} \int_0^{+\infty} g_2(x,y)S(t-x)S(t-y)dxdy - S \int_0^{+\infty} g_2(x,x)dx$$

$$G_3 = \int_0^{+\infty} \int_0^{+\infty} \int_0^{+\infty} g_3(x,y,z)S(t-x)S(t-y)S(t-z)dxdydz - 3S \int_0^{+\infty} g_3(x,x,y)S(t-y)dxdy$$

where in the last two terms we subtracted the average over  $t$  (this is the difference w.r.t. the Volterra series). If we do the same for all the other terms we obtain

$$\begin{aligned} \langle r \rangle_t &= g_0 + 0 + 0 + \dots \\ \langle r(t)S(t-\tau) \rangle_t &= Sg_1(\tau) \end{aligned}$$

For example

$$\langle G_2 \rangle_t = \left\langle S \int_0^{+\infty} g_2(x,x)dx - S \int_0^{+\infty} g_2(x,x)dx \right\rangle = 0$$

because the average over time of the term  $S(t-x)S(t-y)$  in the expression for  $G_2$  reduces to  $\delta(x-y)$ , by definition of  $S()$ , while the second term in  $G_2$  is time-independent.

Now, both  $\langle r \rangle_t$  and  $\langle r(t)S(t-\tau) \rangle_t$  are quantities that we can measure. We also know  $S$ , as it is fixed in the experiment. This means that, starting from the experiment, we can measure the kernels

$$\begin{aligned} g_0 &= \langle r(t) \rangle_t \\ g_1 &= \frac{1}{S} \langle r(t)S(t-\tau) \rangle_t \end{aligned}$$

The wiener representation is stable, because  $g_0$  only depends on the value of  $r$  and not on higher-order terms. Importantly, this holds regardless of how many terms we use to express  $r(t)$ . In turn, this allows to uniquely define the kernels.

### 13.7.1 Wiener kernels in the discrete case

In the previous section, we represented neuronal activity with a continuous firing rate. Neuronal activity, however, is usually recorded as a sequence of spikes and denoted  $y(t)$ . Then

$$\begin{aligned} y(t) &= \sum_{t_i} \delta(t - t_i) \\ g_0 &= \langle y(t) \rangle_t \\ g_1 &= \frac{1}{S} \langle y(t)S(t-\tau) \rangle_t \end{aligned}$$

Now, let

$$C(\tau) := \frac{1}{n} \sum_{i=1}^n S(t_i - \tau)$$

which is the convolution of  $S()$  and  $y()$ :

$$\begin{aligned} \frac{1}{n} \int S(t-\tau)y(t)dt &= \frac{1}{n} \sum_{i=1}^n \int S(t-\tau)\delta(t_i - t)dt \\ &= \frac{1}{n} \sum_{i=1}^n S(t_i - \tau) \\ &= C(\tau) \end{aligned}$$

The function  $C(\tau)$  is given by the average of the stimulus that happen  $\tau$  units of time before each spike. A graphical representation is in figure 57.

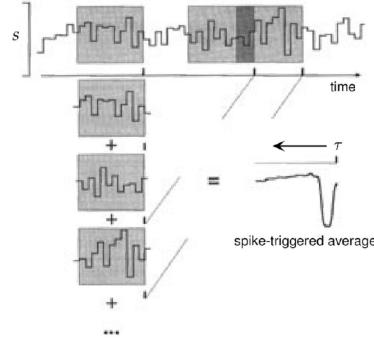


Figure 57:

### 13.8 Visual system: Spatio-temporal receptive fields

We can generalize the previous discussion to stimuli that depend also on space.

For example, we can have as stimuli, images that change in time (black and white pixels whose contrast changes as a white noise process). Hence we can obtain a Wiener kernel  $g(x, y, t)$  that depends on time and space. So we can compute the structure of the visual stimulus that maximally activates a neuron.

We can have a separable receptive field

$$g(x, y, t) = g_s(x, y)g_t(t)$$

or a non-separable one

$$g(x, y, t) \neq g_s(x, y)g_t(t)$$

The receptive field is given by all points in space for which the kernel is non-zero  $\{(x, y)\} : g(x, y, t) \neq 0$ .

#### 13.8.1 Receptive Fields in the early visual system

It is possible to measure the receptive field in Visual cortex using the aforementioned images. The spatial component (with respect to the input) of the receptive fields is shown in figure 58.

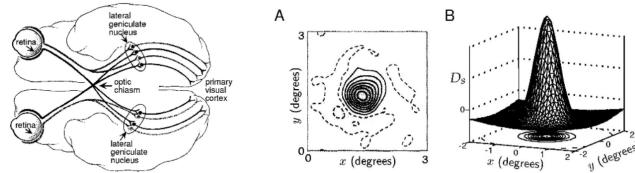


Figure 58: Receptive fields for neurons in the retina (LGN) and in the Thalamus. Here the spatial coordinates are the ones of the visual stimulus (images). The neurons taken in consideration respond mostly to the central part of the images, while there is suppression in the outer area of the image (as  $g$  is negative there).

#### 13.8.2 Spatial Receptive Field of a simple cell in V1

See figure 59.

### 13.9 Linear-Nonlinear-Poisson (LNP) model

In general, the firing rate of a neuron is not a linear function of the stimulus, as we assumed when we stopped at  $g_1$  in the expression for the firing rate  $r$ .

The approach to include non-linearity is the following: perform Spike-Triggered Average to compute

$$L(t) = \int_0^{+\infty} g_1(x)S(t - x)dx$$

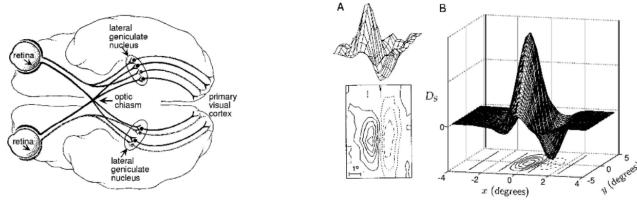


Figure 59: Spatial receptive field in V1 (Visual CORte). The structure is typical of the Gabor function, a Gaussian composed with a cosine along a direction

and then we set the firing rate to be a general non-linear function of  $L(t)$ :

$$r(t) = \phi(L(t))$$

We then generate spikes by assuming that the neuron generates spikes as a Poisson process with rate  $r(t)$ .

### 13.9.1 Fitting a LNP model to data

We plot the firing rate as a function of  $L$ , which shows generally a non-linear relation. In particular, this allows to measure the non-linear function  $\phi$ .

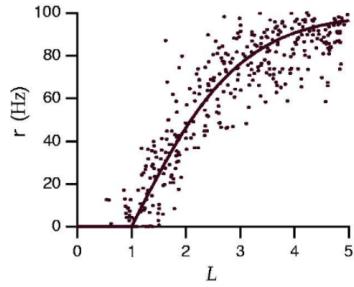


Figure 60: Non-linearity in LNP model

## 14 Single neuron coding II: Information theory

### 14.1 Setup

We consider an experimental setup with a stochastic stimulus  $s$  (and its probability distribution  $P(s)$ ) and the neuronal response  $r$  to the stimulus (and its related conditional probability on the stimulus  $P(r|s)$ ).

Our goal is to quantify the information we can recover about the stimulus, given the neuronal response.

### 14.2 Entropy

Given a discrete random variable  $X$ , we define its entropy as

$$H(x) = - \sum_i p(x_i) \log_2(x_i)$$

which is non-negative and measures the uncertainty in the random variables. If the random variable is not random (always takes the same value) then the entropy is zero.

Entropy is maximized if the variable is uniformly distributed.

#### 14.2.1 Joint entropy

The joint entropy of two random variables  $X$  and  $Y$  is defined as

$$H(X, Y) = - \sum_{x,y} p(x, y) \log_2(p(x, y))$$

if the variables are independent, then

$$H(X, Y) = H(X) + H(Y)$$

while if the variables are correlated

$$H(X, Y) < H(X) + H(Y)$$

#### 14.2.2 Conditional Entropy

The conditional entropy of  $X$  given  $Y$  is defined as

$$H(X|Y) = - \sum_y p(y) \sum_x p(x|y) \log_2(p(x|y))$$

If the variables are independent then  $p(x|y) = p(x)$  which implies

$$H(X|Y) = H(X)$$

In general

$$H(X|Y) = H(X, Y) - H(Y)$$

This is because of Bayes theorem

$$\begin{aligned} H(X, Y) &= - \sum_{x,y} p(x, y) \log_2(p(x, y)) \\ &= - \sum_{x,y} p(x|y)p(y) \log_2(p(x|y)p(y)) \\ &= - \sum_{x,y} p(x|y)p(y) [\log_2(p(x|y)) + \log_2(p(y))] \\ &= - \sum_y p(y) \sum_x p(x|y) \log_2(p(x|y)) - \sum_x p(x|y) \sum_y p(y) \log_2(p(y)) \\ &= H(X|Y) + H(Y) \end{aligned} \tag{29}$$

because  $\sum_x p(x|y) = 1$ .

### 14.3 Mutual Information

The mutual information between  $X$  and  $Y$  is defined as

$$I(X, Y) = \sum_{x,y} p(x, y) \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right)$$

which is a non-negative quantity.

The mutual information quantify the decrease in uncertainty in  $X$  once  $Y$  is known. Importantly, mutual information is symmetric, i.e.,  $I(X, Y) = I(Y, X)$  and it is zero for independent variables.

#### 14.3.1 Relationship between Entropy and Mutual Information

Importantly

$$\begin{aligned} I(X, Y) &= \sum_{x,y} p(x, y) \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right) \\ &= \sum_{x,y} p(x, y) \log_2 (p(x, y)) - \sum_{x,y} p(x, y) \log_2 (p(x)) - \sum_{x,y} p(x, y) \log_2 (p(y)) \\ &= -H(X, Y) - \sum_x p(x) \log_2 (x) - \sum_y p(y) \log_2 (y) \\ &= H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X|Y) \end{aligned}$$

where the last equality is a direct consequence of equation 29.

#### 14.3.2 Mutual Information for continuous distributions

For continuous random variables, Mutual Information is defined as

$$I(X, Y) = \int p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) dx dy$$

which is always non-negative, as in the discrete case.

### 14.4 Examples

**Example 5** (Binary stimulus). Let the time-dependent stimulus be uniformly distributed,  $S = \{0, 1\}$ . Let the spiking activity of the neuron be defined as

$$\begin{aligned} \text{if } S = 0 &\implies \begin{cases} \text{spike with prob. } p \\ \text{no spike with prob. } 1 - p \end{cases} \\ \text{if } S = 1 &\implies \begin{cases} \text{spike with prob. } 1 - p \\ \text{no spike with prob. } 1 \end{cases} \end{aligned}$$

The mutual information between the stimulus and the spike can be easily found by

$$I(Y, S) = H(Y) - H(Y|S)$$

the probability distribution of the spiking variable  $Y$  is

$$Y = \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}$$

Indeed  $p(Y = 0) = p(S = 0)(1 - p) + p(S = 1)p = \frac{1}{2}$ . Hence

$$H(Y) = - \sum_{y \in \{0,1\}} p(y) \log_2 (p(y)) = -2 \left( \frac{1}{2} \log_2 \frac{1}{2} \right) = 1$$

Now, to compute  $H(Y|S)$  we need the conditional distribution

$$Y|(S=0) = \begin{cases} 0 & \text{with probability } p \\ 1 & \text{with probability } 1-p \end{cases}$$

and the same for  $S=1$ . Therefore,

$$H(Y|S) = (1-p)\log(1-p) + p\log(p)$$

Next, we discuss an example that is more realistic and closer to real-life applications.

**Example 6** (Noisy Gaussian Channel). We have a continuous stimulus  $S$ , corrupted by noise  $z$ . The output  $Y$  is given by the sum of the two

$$Y = S + z$$

We assume that both the noise and the stimulus are Gaussian

$$\begin{aligned} z &\sim \mathcal{N}(0, \sigma_N) \\ S &\sim \mathcal{N}(\mu, \sigma_S) \end{aligned}$$

Hence, we can directly compute the probability distribution of the input

$$Y \sim \mathcal{N}(\mu, \sigma_y) \implies p(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y-\mu)^2}{2\sigma_y^2}}$$

where  $\sigma_y = \sqrt{\sigma_N^2 + \sigma_S^2}$ .

The mutual information between stimulus and output is given by

$$\begin{aligned} I(Y, S) &= H(Y) - H(Y|S) \\ &= - \int p(y) \log_2 p(y) dy - H(S+z|S) \\ &= - \int p(y) \left[ -\frac{(y-\mu)^2}{2\sigma_y^2} \log_2 e - \frac{1}{2} \log_2 (2\pi\sigma_y^2) \right] dy - H(z) \\ &= \frac{\log_2 e}{2\sigma_y^2} \int p(y)(y-\mu)^2 dy + \frac{1}{2} \log_2 (2\pi\sigma_y^2) - \frac{1}{2} \log_2 e - \frac{1}{2} \log_2 (2\pi\sigma_N^2) \\ &= \frac{1}{2} \log_2 e + \frac{1}{2} \log_2 (2\pi\sigma_y^2) - \frac{1}{2} \log_2 e - \frac{1}{2} \log_2 (2\pi\sigma_N^2) \\ &= \frac{1}{2} \log_2 \left( \frac{\sigma_y^2}{\sigma_N^2} \right) \\ &= \frac{1}{2} \log_2 \left( 1 + \frac{\sigma_S^2}{\sigma_N^2} \right) \end{aligned}$$

The next example is related to the LNP models.

**Example 7** (Poisson neuron tuned to a 1D stimulus). Here, the protocol is as described in section 13.9, with  $L$  being the stimulus itself.

The firing rate is given by  $\phi(S)$ , where  $\phi$  is a non-linear tuning curve. Then, it is possible to obtain the probability distribution of the number of spikes,  $n$ , in the time interval  $(0, t)$ , given a stimulus  $S = s$ .

Here  $n$  is distributed as a Poisson random variable with rate  $r \cdot t = \phi(s) \cdot t$ :

$$P(n|s) = \frac{(\psi(s)t)^n}{n!} e^{-\phi(s)t}$$

Then it is possible to compute the mutual information, as before.

## 14.5 Efficient coding hypothesis

It is the hypothesis that parts of the brain have the main goal of maximizing the information transmitted to the other parts of the brain. We can use this principle to explain experimental observations.

## 14.6 What to optimize?

The optimization process we will try to carry out involves maximizing the mutual information between Stimulus and firing rate in the LNP process over the linear filter and over the static nonlinearity in the LNP.

Other types of optimization: what are the possible weights that maximize the information flow, when the input to the neuron come from different sources.

## 14.7 Optimizing the static non-linearity

We assume that the stimulus is given by

$$y = \phi(s) + z$$

where  $s$  is the stimulus,  $\phi$  is a nonlinear function and  $z$  is Gaussian noise  $z \sim \mathcal{N}(0, 1)$ .

Our goal is to maximize the mutual information between the signal and the response of the neuron in the simplified case  $z = 0$ . The mutual information is given by

$$\begin{aligned} I(Y, S) &= H(Y) - H(Y|S) \\ &= H(Y) - H(z) \\ &= H(Y) - \frac{1}{2} \log_2(2\pi\sigma_N^2) \end{aligned}$$

Hence, we need to maximize  $H(Y)$ . The entropy of  $Y$  is maximized when it is uniform.

As a simplification, we consider the case  $z = 0$ .

Now, assuming that  $Y$  takes values in  $(0, r_{max})$ , we seek a function  $\phi : S \rightarrow Y$  that maps the signal/stimulus distribution into a uniform distribution of  $Y$ .

We do so by **Histogram equalization**.

Consider a bin  $[s, s + \Delta s]$  in the probability distribution of  $S$ . This bin has area approximately of  $P(s) \cdot \Delta s$  and is mapped by  $\phi$  into a bin with same area, because probability needs to be conserved (as we are in the case with no noise).

Hence, given that the bin in the  $Y$  space has an area approximately of  $\left(\phi(s + \Delta s)\phi(s)\right) \frac{1}{r_{max}}$  we need to set them equal

$$\begin{aligned} \frac{\phi(s + \Delta s)\phi(s)}{\Delta s} &= r_{max}p(s) \\ \frac{d\phi}{ds} &= r_{max}p(s) \end{aligned}$$

taking the limit as  $\Delta s \rightarrow 0$ . Hence

$$\phi(s) = \int_{-\infty}^s r_{max}p(x)dx$$

Therefore, we choose  $\phi$  to be

$$\phi(s) = r_{max}CDF(s)$$

where  $CDF(s)$  is the Cumulative Distribution function of  $s$ . This is the non-linearity that maximizes the mutual information.

### 14.7.1 Histogram Equalization in the fly visual system

Researchers measured the intensity of light that a fly receives in the experimental setup. Once that is measured, we obtain an estimation of the non-linearity, which can then be compared to actual measurements of the non-linearity on real neurons in the fly visual system, as shown in figure 61. Therefore, we obtain an empirical confirmation of the fact that the retina is indeed maximizing the amount of information (regarding the environment) it transmits to other parts of the brain.

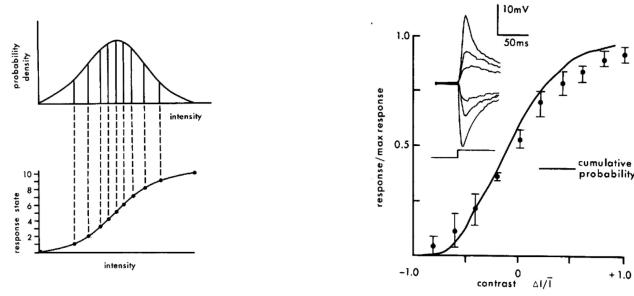


Figure 61:

## 14.8 Optimizing the temporal filter: Whitening

We can also maximize the information transferred with respect to the temporal filter.

We obtain that the output should be as close to white noise as possible, as information transmission is maximized when temporal correlations in the output are minimized.

Therefore, we should choose a temporal filter that maps the stimulus into white noise.

### 14.8.1 Optimal receptive fields with natural images

With this approach, it is possible to reproduce the experimental observations regarding the Off/On-center and Off/on-surrounding receptive fields observed in the retina and V1 (section 13.8.2).

# 15 Population coding and decoding

We will now consider a population of neurons. We are interested in estimating and quantifying the information carried by the population and the information that can be decoded by a neuron from the stimulus, given the output it receives from the rest of the population.

## 15.1 Information measures

- Mutual information

$$I = \int p(\theta) d\theta \int P(r|\theta) \left( \frac{P(r|\theta)}{P(r)} \right) dr$$

- Fisher Information

$$J(\theta) = \int P(r|\theta) \left( \frac{\partial}{\partial \theta} \log P(r|\theta) \right)^2 dr$$

The Fisher Information provides, through the Cramer-Rao lower bound, an intuitive relationship with the error made in estimating the stimulus from the neuronal response. Moreover, it is also often easier to compute than Mutual Information. At the same time, however, FI is defined only for continuous stimuli. For discrete stimuli, MI is the only option.

- Average FI

$$J = \int p(\theta) J(\theta) d\theta$$

## 15.2 Decoding - Estimators

We observe the response  $r$  of  $N$  neurons, driven by a stimulus  $\theta$ .

Our goal is to estimate (decode) the stimulus from the neurons response.

An estimator (decoder) is a function  $\hat{\theta}(r)$ .

### 15.2.1 Examples

**Example 8** (A simple estimator). We have  $N$  neurons. The firing rate of neuron  $i$  in response to a stimulus  $\theta$  is given by

$$r_i = \theta + z_i$$

with  $z_i \sim \mathcal{N}(0, 1)$ .

We define

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N r_i = \theta + \frac{1}{N} \sum_{i=1}^N z_i$$

to be our estimator.

**Example 9** (A more biologically relevant estimator). We have  $N$  neurons. We assume that

$$r_i = r_{max} (1 + \cos(\theta - \theta_i)) + z_i$$

Hence, we are associating to each neuron a preferred stimulus  $\theta_i$ .

Our goal is to estimate  $\theta$ .

Let  $c_i$  be a two-dimensional unit vector

$$c_i = \begin{pmatrix} c_i^x \\ c_i^y \end{pmatrix} = \begin{pmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{pmatrix}$$

and define  $C$  to be

$$C = \frac{2}{r_{max}} \frac{1}{N} \sum_i r_i c_i = \begin{pmatrix} C_x \\ C_y \end{pmatrix} \quad (30)$$

As a function of the angle  $\theta$ , the response of the neuron is given by  $r_i(\theta)c_i$ . We now want to find the expected value of  $C_x$ .

$$\begin{aligned} C_x &= \frac{2}{r_{max}} \frac{1}{N} \sum_i r_i(c_i)_x \\ &= \frac{2}{r_{max}} \frac{1}{N} \sum_i [r_{max}[1 + \cos(\theta - \theta_i)] + z_i] \cos(\theta_i) \\ &= 2 \langle [1 + \cos(\theta - \theta_i) + z_i] \cos(\theta_i) \rangle_{\theta_i, z_i} \\ &= 2 \langle \cos(\theta - \theta_i) \cos(\theta_i) \rangle_{\theta_i} \\ &= \cos(\theta) \end{aligned}$$

where, in passing from the sum to the average, we considered the limit for  $N \rightarrow +\infty$ .

We also need to assume that the  $\theta_i$  are uniformly distributed in  $[0, 2\pi)$ , otherwise we would have  $\langle \cos(\theta_i) \rangle_{\theta_i} \neq 0$ . Similarly we can compute the  $y$ -component of  $C$ ,

$$C_y = \sin(\theta)$$

Hence, defining the decoder as in 30, if we sum over a large number of neurons with randomly distributed preferred orientation  $\theta_i$ , it is possible to obtain  $\theta$  by inverting the components of  $C$ .

### 15.3 Error of the estimator - bias, variance

For a given estimate, we can compute its bias

$$b_{\hat{\theta}} = \langle \hat{\theta}(r) \rangle - \theta = \int P(r|\theta) \theta(r) dr - \theta$$

The Mean Square Error (MSE) is given by

$$\langle (\hat{\theta}(r) - \theta)^2 \rangle$$

#### 15.3.1 Cramer-Rao

For any unbiased estimator

$$MSE \geq \frac{1}{J(\theta)}$$

**Example 10** (Continuation of Example 8). We have  $N$  neurons, with  $r_i = \theta + z_i$  with  $z_i \sim \mathcal{N}(0, 1)$ . An unbiased estimator is given by

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^n r_i$$

We compute the MSE:

$$\begin{aligned} \langle (\hat{\theta}(r) - \theta)^2 \rangle &= \langle \left( \frac{1}{N} \sum_{i=1}^n z_i \right)^2 \rangle \\ &= \frac{1}{N^2} \sum_{i,j=1}^N \langle z_i, z_j \rangle \\ &= \frac{1}{N^2} \sum_{i,j=1}^N \delta_{i,j} \\ &= \frac{1}{N} \end{aligned}$$

To evaluate the Cramer-Rao bound

$$J = \int p(\theta) J(\theta) d\theta$$

we need

$$J(\theta) = \int P(r|\theta) dr \left( \frac{\partial}{\partial \theta} \log P(r|\theta) \right)^2$$

with

$$\begin{aligned} P(r|\theta) &= \prod_{i=1}^N P(r_i|\theta) \\ P(r_i|\theta) &\sim \mathcal{N}(\theta, 1) \end{aligned}$$

To compute  $J(\theta)$  we need

$$\begin{aligned} \log P(r|\theta) &= \log \left( \prod_{i=1}^N P(r_i|\theta) \right) \\ &= \log \left( \prod_{i=1}^N (2\pi)^{-\frac{1}{2}} e^{-\frac{(\theta - r_i)^2}{2}} \right) \\ &= \sum_{i=1}^N -\frac{(\theta - r_i)^2}{2} - \frac{1}{2} \log(2\pi) \end{aligned}$$

which implies

$$\frac{\partial}{\partial \theta} \log P(r|\theta) = -\frac{1}{2} \sum_{i=1}^N 2(\theta - r_i)$$

Hence

$$\begin{aligned} J(\theta) &= \int \prod_{i=1}^N P(r_i) \sum_{i,j} (\theta - r_i)(\theta - r_j) \\ &= \sum_{i=1}^N \int (\theta - r_i)^2 P(r_i) dr_i \\ &= N \end{aligned}$$

because  $\int (\theta - r_i)^2 P(r_i) dr_i$  is the variance of the rates and is equal to 1.

Therefore, we have showed that the MSE is equal to the inverse of the Fisher Information. The estimator we have defined is the best-performing one.