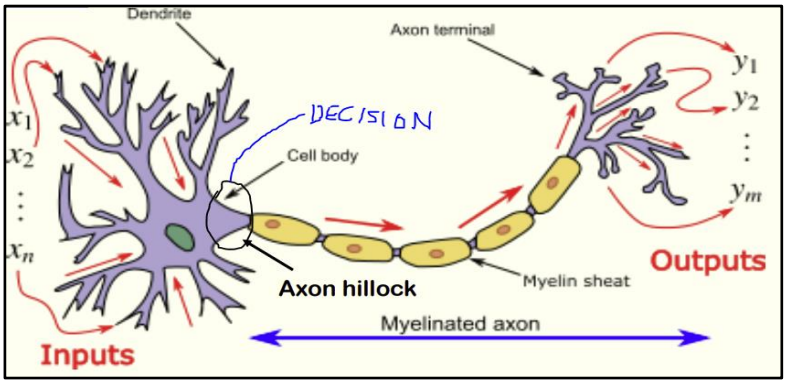
Summary NeuroEngineering [L.Astolfi]

(2020-2021)

2 - The Neural Cell

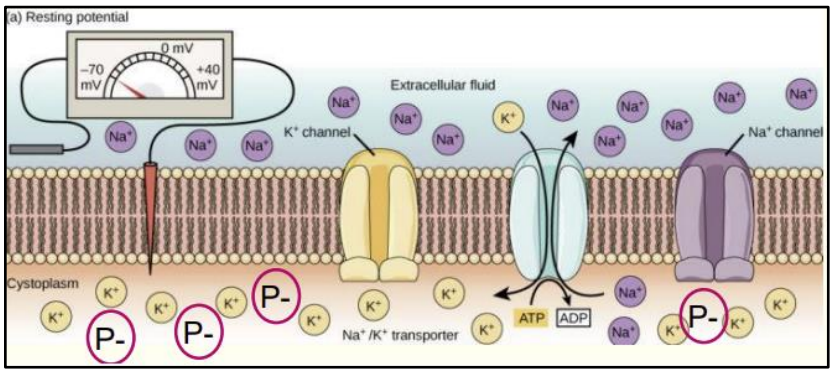
Neuron basic function:

1. Collection of information from multiple sources (each neuron collects multiple information from multiple sources)
2. Integration: the processing of incoming information (in time and space) to provide a binary decision
3. Generation and propagation of the bit of information (binary decision) up to target cells (other neural cells, muscle cells)



Each function is performed by a different part of the neuron.

Multiple input $x1..xn$ but only 1 single binary decision in cell body, then propagated in all the other cells with $y1 = y2 = \cdots = ym$.



The **neural membrane** is most important (interesting) structure in neural cells because first allows all the function of the neuron (everything happen here). The yellow thinks in the membrane are the lipidic (or phospholipid) bilayer, each layer composed by a phosphoric head that is polar so it can connect to water and two lipidic tails that does not allow water to pass. So, this lipidic bilayer block all the exchange between outside the neuron and inside the neuron. **In particular the exchange of the ion Na+ K+ Cl- and Ca++(these are the four main ion families).** We have some doors (Ion pump) where specific ion can pass. If we measure the difference in electrical potential between the interior and external of a neuron at rest it's around -70mV due to the different ion concentrations (internal part more negative) so the membrane is polarized. This value (-70) is given by the

elettrochemical equilibrium, the sum from the diffusional forces (force due to the chemical gradient, the concentration) and the electrical forces (due to the electrical value + - of the ions) of all ion family (tutte insieme) plus the Ion pumps (it is a dynamic equilibrium). Nernst equation:

$\Delta\mu = \mathrm{RT}\ln\frac{[X]_A}{[X]_B} + \mathrm{zF}(\mathrm{E_A} - \mathrm{E_B})$ (first part diffusional forces, second electrical forces)

Ion pumps use energy (ATP) to move ions against their electrochemical gradient (so move from where is less concentrated to where is much concentrated), to maintain this difference in concentration (otherwise the ion pass to membrane passively).

Membrane **depolarization** where or $ion+$ go inside the membrane or $ion-$ go outside the cell -> reduction of the negativity of the internal environment with respect to the external one. The ions that flow where there is a depolarization are called deporalized ions (deporalized current)

Membrane **hyperpolarization** where ion – go inside the membrane or ion + go outside the cell. There is 2 type, or a **Repolarization** if before the membrane was depolarized or a **hyperpolarization** if before the membrane is polarized (when is more negative of -70mV).

Synapses: it's the point where two cells meet and exchange information. The most important (the common one) way to exchange information is with chemical synapses. Not a bidirectional exchange but a unidimensional one from **PRE-synapses** cell to the **POST-synapses** cell. There is no contact between them but there is a space called **Synaptic cleft**. The information is exchange through chemical transmitters called **neurotransmitters**. Each transmitter has the role of open some specific channel (**Ion gated channels**) in the postsynaptic membrane. The ion channel will open when a neurotransmitter (released by presynaptic cell) reach a receptor in the postsynaptic membrane. Neurotransmitter is like a <u>key</u>; the receptor is the <u>lock</u> that open the ion channels.

The synapsis that releases a neurotransmitter that open a port (for example for Na+) and generate a Depolarization are called **Excitatory Synapsis**. On the contrary, if open a port that generate Hyperpolarization (for example Cl-) are called **Inhibitory synapsis**. $\boldsymbol{Excitatory} = sign +$, $\boldsymbol{Inhibitory} = sign -$.

In reality we have different synapsis, so we have different effect that we have to sum together. So, the depolarization has not always the same graphics $mV/Time$ but we can sum different graphics because there is different synapsis (Spatial summation) or because in a single synapsis there are 3 different signal one near the other (Temporal Summation).

When we have a local de/hyperpolarization we have an intra and extra cellular ion currents that compensate the differential potential from this point and the other near it.

Nernst Equation:

$$\Delta\mu = \mathrm{RT}\ln\frac{[X]_A}{[X]_B} + \mathrm{zF}(\mathrm{E_A} - \mathrm{E_B})$$

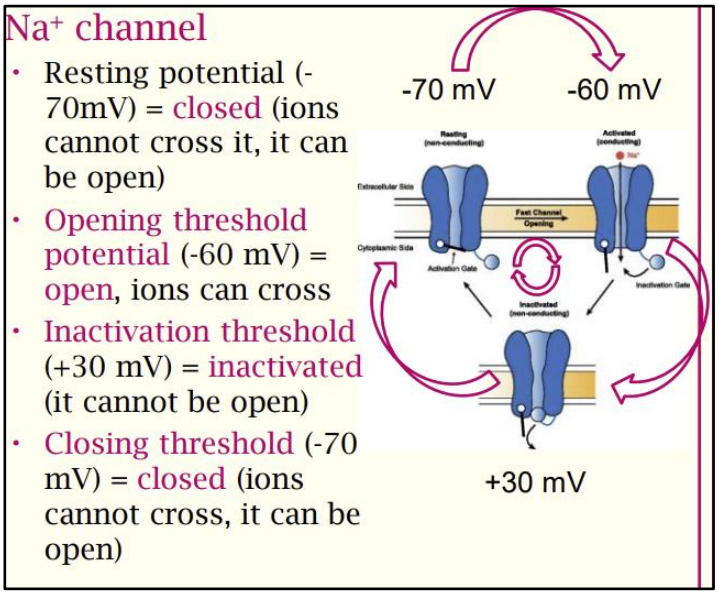
The first part represents the **Thermal energy** of the molecules that moves ions according to the chemical gradient; the second part represent the **Potential energy** that moves ions according to the electrical gradient. When these forces are in equilibrium (**electrochemical equilibrium**) $\Delta\mu = 0$ we obtain that:

$$E_I - E_E = E_j = \frac{RT}{zF} ln\left(\frac{[extra]_j}{[intra]_j}\right)$$

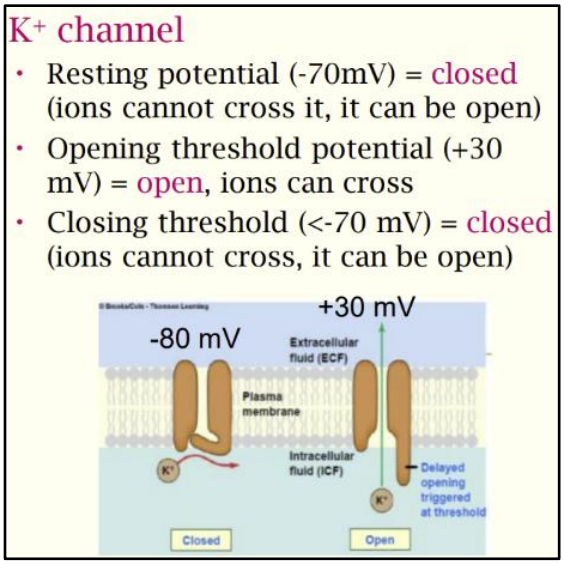
Electrochemical equilibrium potential E_j for each ion family j
Z = valency of ion j
$[extra]_j/[intra]_j$ = extra-/intra-cellular concentration of ion j

For each family E_j is different for example $E_{K+} = -90mV$, $E_{Na+} = 50mV$, $E_{Ca++} = 150mV$.
Thanks to this value we can predict the behaviour if a family of ions is able to move. From rest situation (-70mV) K+ -> hyperpolarization (inhibitory effect), Na+ and Ca++ -> Depolarization (excitatory effect).

The **Action Potential:**
It is a variation of the membrane potential which appears only in neural, muscular, and cardiac cells. The effect of this variation is propagated to other cells and bring information. It is an all-or-none process: if the stimulus does not reach a given threshold, it does not happen. If the threshold is reached, it has always the same shape, duration, and intensity, irrespectively to the stimulus amplitude. So, it is like a binary decision.

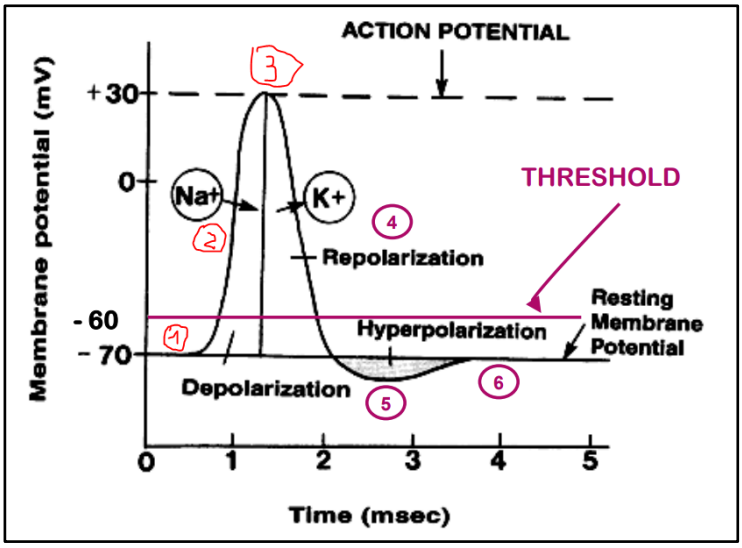


In neural cell for example the sodium channel has 3 stage, the close one (Resting potential) when stay at -70mV, after it there is the open one (Opening threshold potential) that start at -60mV and after the open one there is always the Inactivation threshold (non si può richiudere, deve perforza andare nello stato di inattivazione)



(Inactivation threshold) when stay at +30mV. Next it returns to the closing one. It is a cycle in a one-way direction. So, when we arrive to -60mV always there is a strong Depolarization up to +30mV and then always there is a strong repolarization to -70mV.

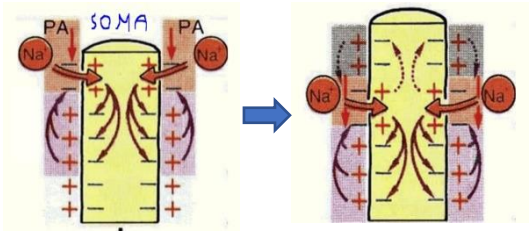
The Potassium channels is simpler. There are only two stages: the resting potential at -70mV (closed) and the opening threshold potential at +30mV. So, we need to have a strong depolarization to open it, then the channel can close again after a strong Repolarization and Hyperpolarization at -80mV.



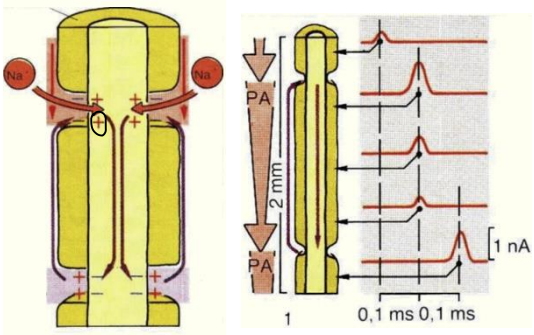
1. **Depolarization** to the Na+ channel opening (at -60mV Na+ channel opens) caused by a perturbation.
2. **Fast depolarization** (the channel Na+ is open, there is a Na+ depolarizing currents from out to inside the cells);
3. Inactivation of the Na+ channels and opening of the K+ Channels and start:
4. **Fast Repolarization** due to the K+ repolarizing currents (from in to out);
5. To close the K+ channel there is also a **Hyperpolarization** to bring -80mV;
6. Return to the resting potential because both the channels are close (due to sodium-potassium pump. Pump work as $Na+$ outside and $K+$ inside)

This graphics is fixed, each step is repeated always in the same ways, with the same shape, with the same values.
From 3 to 6 is called **refractory period**, from 3 to when bring the first time -70mV is the **Absolute refractory period** in which the sodium channel is inactivate and is impossible to re-open it, and the time where it is Hyperpolarized is the **Relative refractory period** where is not impossible to have another action potential (also if it is less likely to occur because there were a stronger perturbation to start another action potential).

The action potential has to reach the other cells, is the output of a cell and the input of other cells in the synapsis. So, it can propagate. There is two kinds of propagation:



Point to point propagation or **unmyelinated axons**: the action potential is continuously generated at each membrane section, and it propagate only unidirectional, from the soma to the synaptic bouton because even if the perturbation goes in both the direction there is the refractory period that does not permit to start another action potential in backward. It is propagated along the fibre up to the synapses.

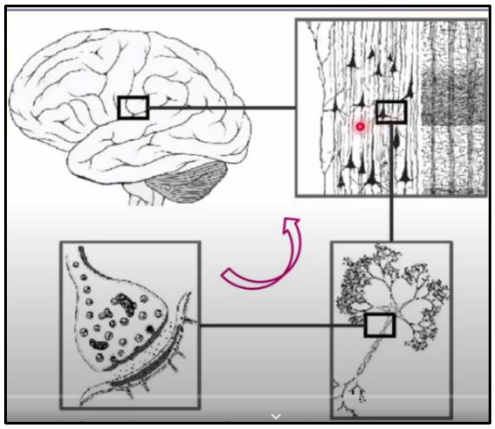


Saltatory propagation or **myelinated axons**: The structure is different, there is a myelinated sheath (1) that provide an electrical isolation with respect to the external environment. Where there is the myelinated sheath we don't have the gated channel. We have the voltage-gated channel only at the Ranvier nodes and only here the action potential is generated. So, the only propagation is given passivly with the flows of ions intra and extra membrane until the next Ranvier nodes (it has not a strong perturbation like in point to point perturbation but it is sufficent to generate another active potential). So, there is an alternate beetween passive conduction (in myelinated sheath) and active re-generation of the potential (in Ranvier nodes) => saltatory conduction. This mechanism is faster and can propagate the action potential for long distance (può raggiungere dal cervello tutto il corpo anche con solo 2/3 step).
The first mechanism is lower and not very common, only when fast communication is not needed (ex. pain).
spike train = sequence of action potential. Spike = action potential.

3 – Principles of Neuroanatomy and Brain organization

Neural organization:

Thousands of information are collected at each time by a single neural cell, processed by the membrane of the cell and if the result of this integration is perturbation of the membrane in the depolarizing sense, which allows the membrane to reach the threshold for the activation, which means the opening of the specific ion channels, which are the sodium voltage-gated ion channels (they are located at the beginning of the axon), an action potential occurs and it's propagated with two different mechanisms (see previously lecture) to other cells. All the cells are interconnected in a huge network. Neurons in the brain are not put in a random way, they are strictly organized in structures, specific fashion (tipo fasci) where we have parallel axons and parallel dendritic trees. The black triangular pyramidal are the somas, or cellular body. (vedere prima slide per capire). The brain cortex is folded like piece of tissues that you can fold together forming diary and sulci. Different part of the cortex or even structures of the brain are organized in brain region which are functionally specialized.



STUDYING THE HUMAN BRAIN (non credo sia importante ai fini dell'esame)

The brain predicts the future on the basis of the past, helping the individual to survive and perpetuate the species. At the beginning the brain was study at first looking by lesion and animal studies. Then by neuroimaging and neuromodulation. Then link to behavioural and clinical data.

THE BRAIN IN TIME AND SPACE

The temporal scale of the brain activity is the temporal scale of the neuronal activity, and the membrane variation of the neuron is the basis of brain activity. This variation is in the range of milliseconds. When we sum up the activity of many neurons, we may lose the need to count each individual millisecond but the temporal scale is still in this range. Most brain changes that we can detect occur in a range of 10 milliseconds. The propagation of electromagnetic fields in the brain is virtually instantaneous, so this also contributes to the preservation of this temporal dynamics.

The spatial scale depends on what we are looking at, because if we focus on the events that occur in the single cell then the spatial scale is very small (micro meters). It is different if we are looking at two events that occur at the opposite side of the membrane (nano meters).

Brain evolution: its tissues are continuously modified by our life, our lifestyle, what we do, if we have pathologies those can also affect our brain structure. There is an individual evolution in time scale that starts before we are born and then it goes on for our entire lifespan. It does not stop when we are grown up and in does not stop even when we are old. The brain is plastic, that means it changes (when we are learning, during memorization, for spontaneous recovery and neurorehabilitation). It changes not in the shape, but in its organization of the cells. There is also a collective evolution in which generation after generation there are small changes in the brain which allow the species to evolve according to the environment. This kind of change is far slower than the individual.

GREY AND WHITE MATTER

Grey matter: it is called in this way since it is grey. If we look at them are mainly made of cell bodies and dendrites of the neurons, so the neural membrane that is around the soma.

White matter: it is usually called fibers and are constituted by myelinated axons.

Their meaning is also different: the grey matter describes the function of the neurons (if they are located in different part of the brain, they have different functionality), while the white describes the communication between the cells. If we are able to trace this fiber, and we are, we are able to trace the network which links different neurons. With the diffusion tensor imaging (DTI) we can made the white matter tractography (we can create an image of it (vedi l'immagine sulla slide che è ciotta). W.M is also called Corpus callosum or the highway of information between hemispheres/brain regions.

BRAIN CORTEX

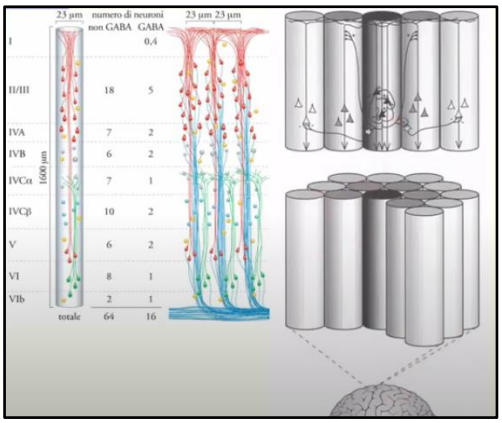
Is the external part of the brain. We don't have a big head to contain the cortical surface that we need because it's folded, that's the reason we have the structures we know of the cortex, that are called gyri and sulci. The last one is 2/3 of the surface and it is the inside part, while the Giri the external (vedi slide per capire).

The brain has 1.5% of the body weight and it uses the 15% of the total blood flow (the whole brain uses 20% of the total energy).

CORTICAL ORGANIZATION

It is organized in six vertical layers, strictly interconnected (they are not independent!). Neurons are organized in columns (0.5 mm of diameter). They are also organized not randomly and they are perpendicular to the cortical surface. Columns are usually composed by neurons with the same behaviour.

The information processing starts from the bottom, so from the internal part of the brain and then spreads to the outside.



We have **DIFFERENT NEURAL CONNECTIONS**: (luchino ha fatto una domanda qua)

-**Feedforward**, or bottom-up: directed from regions at the first processing stages to the following ones.

-**Lateral**: connections linking same stage regions

-**Feedback** or top-down: from advanced stages back to the previous ones.

The majority of connections are lateral ones and all the connection can be excitatory or inhibitory

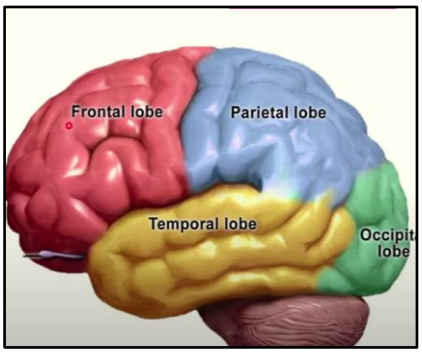
BRAIN LOBES

Frontal (parte davanti fino a metà testa): control center for executive functions, such as reasoning, decision-making, language, cognitive processes as memory, orientation, planning and execution of movement.

Parietal (altra metà della testa): here we have the primary and secondary somatosensory cortex, the part of the brain which allows us to feel for example the object that we touch. It is related to perception and sensory (touch, pressure temperature and pain).

Occipital (parte dietro la nuca): it has visual function (primary visual cortex and it process and interpret the visual information.

Temporal (parte ai lati, ce ne sono due): auditory cortex, center for receptive language and it is a region called hippocampus, a region regarded to memory formation and emotion.



BRODMANN AREAS

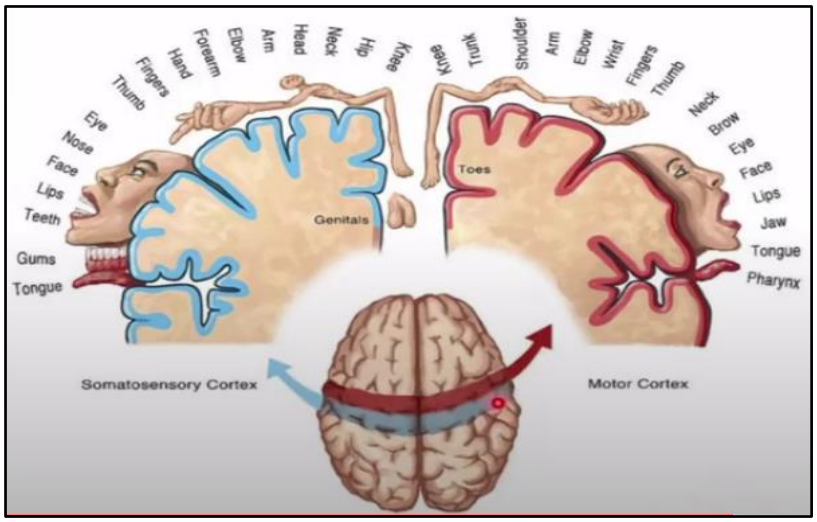
A lobe (le parti viste sopra) hosts a number of functions, but this not means that these functions are mix together. They are organized in specific subregion located in specific brain lobes. This region are called broadmann areas.

VISUAL CORTEX (retinotopic), located in the occipital lobe

It is highly specialized for processing information about static and moving objects and is excellent in pattern recognition. In the retina we have neurons that dives the information in categories, as shape, movement, contrast and so on, and then they integrate all the results together in a mapping called renotopic.

MOTOR/SEMATOSENSORY CORTEX: located in the middle of the head.

These two regions have a somatotopic organization. Somatotopic means for which follows position of the parts of the body. Ogni pezzettino del motor cortex e del somatosensory cortex presenta dei neuroni collegati ad una specifica parte del corpo (vedi figura). Quindi ogni volta che muoviamo una specifica parte del corpo, l'informazione del "muovere" parte da quello specifico pezzettino.



SUBCORTICAL AREAS

They are four: thalamus, brainstem, basal ganglia and cerebellum. Have functions more basic respect the ones provide by the cortex. They are region difficult to reach and to measure, especially noninvasively.

Thalamus: controls traffic from periphery to the cortex. We have two of them, one for hemisphere in the middle of the brain. The cortex is strictly interconnected to the thalamus. It synchronizes the cortex at frequencies belongs to the alpha band, for this reason is also called brain pacemaker.

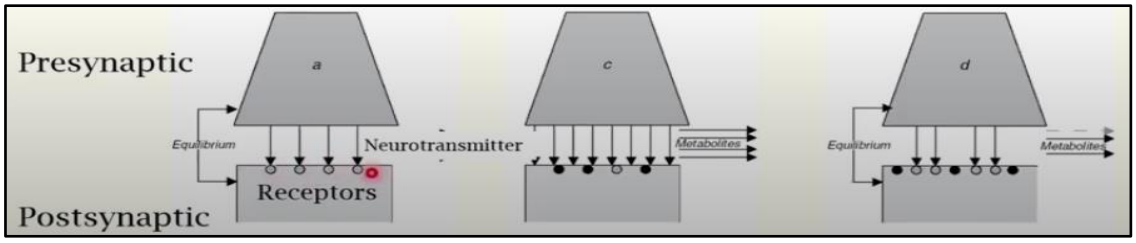
BRAIN PLASTICITY

Critical period (age 0-2): maximum brain reorganization, irreversible changes.

We have evidence of plasticity in adults end even in elderly people: as learning, memory (the brain organization changes when we studied something new or when we "ricordiamo"), consequences of sensory deprivation (amputees, acquired blindness, …) ->there is an adaptation of the brain after these things. Reorganization after brain lesion and effectiveness of neurorehabilitation.

T-Synaptic plasticity

What happens and what can change in the exchange of information as a consequence of brain plasticity: we have to cells, one is a presynaptic that sent an information to the other one, the postsynaptic. The exchange of information is made through the neurotransmitters linked to the receptors of the postsynaptic cell. We repeat this stimulation many times in a short amount of time (faccio lo stesso movimento più e più volte per tanto tempo di fila. Dopo per esempio un'ora di pratica sono più bravo a fare quel movimento. Se non faccio più quel movimento per tanto tempo, tipo anni o mesi, non saprò più farlo bene come prima.)



At synaptic level, when we repeat something several times, we have an increase of neurotransmitter. Il cervello capisce che quello che stiamo facendo è importante perchè stiamo cercando di farlo tante volte, quindi fornisce più collegamenti, rinforzando il circuito alla base del movimento. Non viene fatto un cambio strutturale, nell'organizzazione, ma funzionale-> the metabolic processes that produce and that result in this release of neurotransmitter are increased.

After weeks or months of training something change in the membrane of the cell: the number of receptors increase, so the velocity at which the information will be transferred from the presynaptic cell to the post will increase.

Quindi ho una situazione base. Decido di imparare qualcosa di nuovo e faccio pratica. All'inizio quindi mi aumentano i neurotransmitter (short term plasticity). Dopo tanto tempo avviene il cambiamento, ho più recettori (long term plasticity). Vedi slide per maggiore chiarezza.

Long term plasticity provides structural changes.

We have unmasked connections between neural cells. Più neuroni possono essere collegati tra loro con diverse sinapsi, ma solo una di queste è il collegamento "funzionante". Le altre vengono utilizzate quando la principale si danneggia a causa di una lesione per esempio. Quindi vi è redundacy.

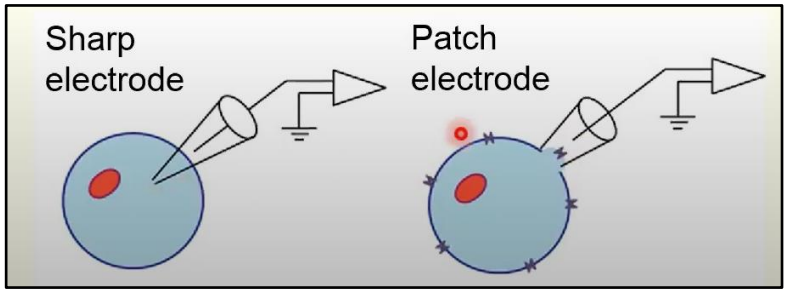
(It is not true that we use 20% of our brain. We use it all, but there is redundancy in the number of connections between cells.)

4 – Electrical Correlates of the Brain Activity

INTRACELLULAR RECORDING

If we want to measure the brain activity, we have to measure the electrical signals generated by the cells: we have a current that cross the membrane of a cell, it moves inside and outside and it also generates a variation of membrane potential.

We can do this recording through a method called intracellular recording: we cross the membrane with an electrode (sharp), or we have some electrodes which are called patch electrodes that do not really cross the membrane. They rather seal with the membrane surface, providing electrical contact with the intracellular fluid.



In both cases there is a reference electrode outside the cell, and we can measure the difference between the inside and the outside, which is the membrane potential.

In these two ways we really capture exactly what's happening in the membrane. How can you insert an electrode inside a neuron?

It is a very complex process. It is usually performed in the soma, so in the membrane surrounding the central body of the cell or in the dendrites. It is more difficult to do that in the axon.

Good points for these methods: the only way to really capture the brain activity, the variation of the membrane potential below the threshold and the action potential.

Limitation: you cannot capture every part of the membrane and it is difficult to perform in vivo, meaning that you are measure the cell when it is connected with the rest of the brain. Otherwise, It is easy to do it in vitro, ergo you remove the cell from the brain, put in vitro and then you stimulate it and see how it reacts. We cannot see what happen when it's interconnected with other cells.

EXTRACELLULAR RECORDING

The electrodes are placed near the membrane, and they do not penetrate it. The difference with the **intracellular** recording is **that extra not measure but detect the action potentials fired by a**

neuron, but not its subthreshold membrane potentials. They measure the exact moment in which an action potential occurs, but not its shape or amplitude.

We can operate this kind of process in vivo. We can see in this case what happen when it's interconnected with other cells. We can also use the axons for the measurement since we do not have to cross the membrane.

RECORDING FROM NEURAL POPULATIONS (not measure one cell but a group of them, their collective activity)

We have to introduce some new terms:

-**Local Field Potentials (LFP)**: it is the extracellular current flow resulting from the linear summation of PSP (Post Synaptic Potential, is the variations due to the post synaptic activity. We are speaking about information received and processed by the cell and not the output of the cell) of neural groups. The spatial resolution is very small: $10^3 \, to \, 1mm^3$. We can measure LFP using thin electrodes put inside the cortex. It is an invasive measure.

-**EEG: originates from synchronous postsynaptic cortical** (in some cases also sub-cortical) currents (sources) of millions/1 billion neurons. If we focus on the scalp EEG there are some properties / events that affect the signal:

Volume conduction: describes what happens to the currents which are ion currents that are propagated across the tissues in the head. The ions can spread since tissues contain salt water, so all tissues of the head allow this propagation.

Electric fields produced by local currents spread instantaneously (at light speed) and sum up linearly.

Among the weaknesses, there is the amplitude of the signal, which is reduced with distance meaning that, when we reach the scalp, it's so attenuated that it moves from the millivolts, that is the scale of the membrane potentials, to the microvolts. Instead, the temporal resolution is very high (milliseconds).

-**Stereo-EEG (S-EEG):** with this family of method, we can measure the activity in subcortical regions, which are much deeper respect of the EEG signals of the cortex. It's still an EEG, but it's called stereo-EEG and it's an invasive measure based on electrodes deeply implanted in the brain. It is used in epileptic patients when the epileptogenic zones are located in depth, and it is not used in general research or in healthy patients.

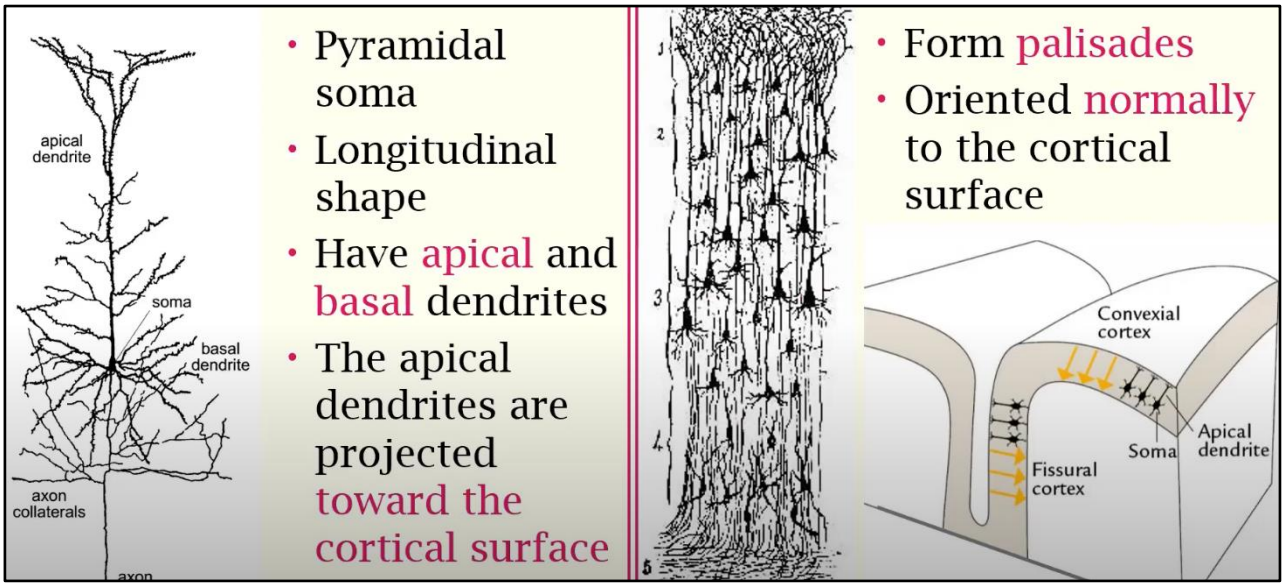
Spatial resolution: $1-4\,mm^3$(very high) and also the amplitude of the signal is stronger that what we get on the scalp because we capture the signal closer to the sources.

(Ti infilano delle sonde nel cervello per farla breve).

-**Electrocorticography (ECoG):** invasive measurement, we can access the cortical activity. It is still an EEG and usually it's performed by means of electrodes arrays put on the cortical, in particular above the pia meter, a very thin membrane covering the cortical surface. We have to open the skull to access the brain, but we don't need to cross the membrane. We can have an effect of volume conduction because what we measure can be the result of the activity of different regions with the propagation across the tissues. Spatial resolution is $1-20m^3$.

GENERATION OF THE EEG SIGNALS

Cortical pyramidal neurons are particular useful to understand the generation of EEG signal and they are mainly responsible for the signal that we gather on the scalp. (figura sotto con caratteristiche scritte)

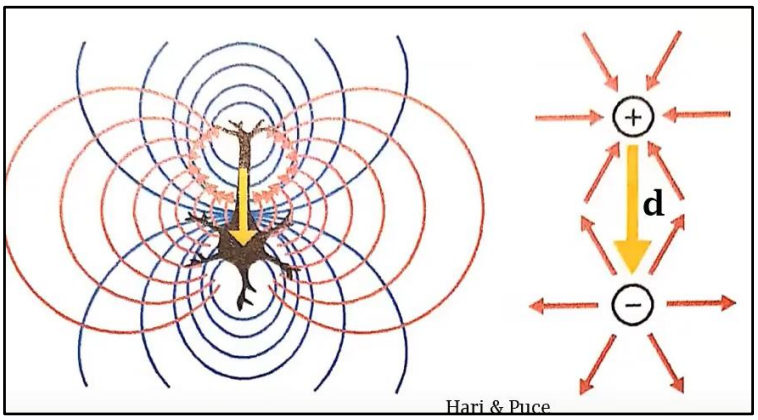


The thing really important of this family of neurons is the shape of the dendritic tree, which is longitudinal.

Form palisades means that they are all parallel and all close to each other as we can see in the picture.

The shape of these neurons allows to model them as a current dipole. So, each neuron behaves like a current dipole. A current dipole is made of two-point sources, positive and negative one, at a distance "d".

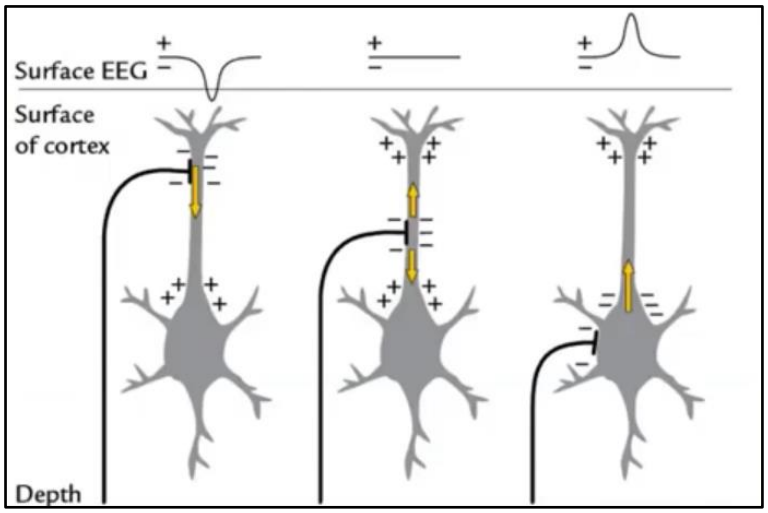
The apical dendrite and the basal dendrite provide the 2 sources and they are divided by a distance "d".



Why can we have this dipolar behaviour? Because we can have synapses in the apical or in the basal dendrite. We can have 2 kinds of postsynaptic potentials: the inhibitory (Potenza con

segno +), in this case located in apical, and the excitatory (Potenza con segno meno), located in the basal. We have also different synapses based on where the presynaptic cell is located: cortico-cortical (CC) synapses usually on the apical dendrites, with presynaptic cell belong to the cortex and the talamo-cortical (TC) synapses usually on the basal dendrites with pre cell belong to the Talamo.

Nota: a seconda di che sinapsi è attiva ho questi casi:



Excitatory synapsis CC has the same sign and direction as inhibitory TC and vice versa.

From now on we will be referring to a group of cells and not to a singular one.

As we said before, pyramidal neurons are perpendicular to the cortical surface, and they have the same orientation of their dendritic tree. Each of this neuron acts like a dipole oriented in the same way respect to the other, which composed the pyramid so even the current they produced has the same direction. They work in synchronous way. This configuration is called palisade.

EEG measures the synchronous electrical activity of neural populations, and the amplitude of the signal is: linearly proportional to the number of synchronous neurons and proportional to the square root of the number of asynchronous neurons, because of intracortical cancellation.

[Se per esempio ho 10^{10}dipoli asincroni allora avrò un segnale pari a 10^5, se invece ho 10^8 dipoli sincroni avrò un segnale di 10^8. Questo implica che se ho meno neuroni ma questi sono sincronizzati, avrò un segnale da loro generato molto più potente di uno generato da molti più neuroni asincroni.]

Gyri produces most of the EEG signal since the favourable orientation (they are parallel to the surface which the electrodes are, and so the currents produced) are perpendicular to the

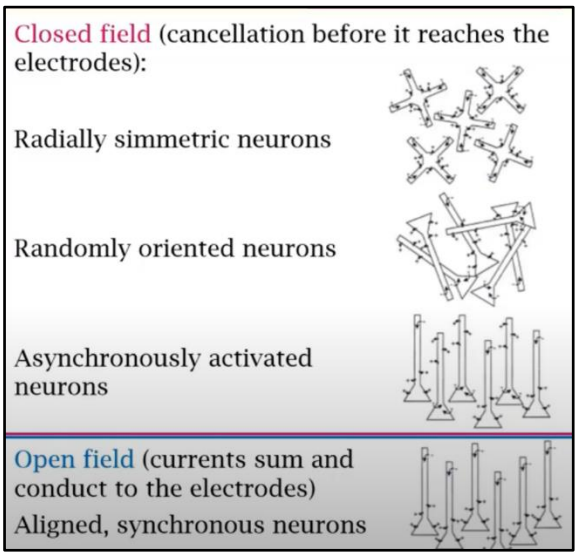
surface. It is easier for the electrodes to catch variations. Summation due to the palisade disposition.

Sulci and fissures produce little or no EEG signal since the orientation is not favourable to the sculp surface.

In Sulci we have always a doble layer of neurons oriented specular with respect to each other; mutual cancellation of opposite cortices.

OPEN AND CLOSE FIELD

Only neurons that produce open field contribute to EEG.



IMPORTANT TO NOTE:

What part of the brain of the neural cell electrical activity is mainly capture by EEG? What kind of membrane potentials are mainly responsible for EEG?

Not action potentials but post-synaptic potentials, because they are slower and can sum up more easily in large groups of neurons, while action potentials are fast and more difficult to add up in time.

LIMITATION OF SCALP EEG

-Spatial blur (attenuation and spread of the potential with distance). The signal must pass through several layers before it

can reach the electrode. Only signals generated by a high number of synchronous neurons can be captured. Furthermore, the deeper the distance from the scalp at which the signal is generated, the lower the intensity with which it is recorded, and vice versa. If the distance is very small, there will be less attenuation.

-Low signal-to-noise ratio

-Multiple sources contribute to the single electrode signal.

-Near electrodes record partially overlapped (correlated) signals.

-Reference choice

If the source that produce the signal is a cortical one, it will produce a strong EEG signals because it is closest to the scalp and since contains a lot of pyramidal neurons (open field produced).

On the contrary, deep sources often produce closed fields, which results in a smaller and weaker EEG signal. They are more distant to the sensors. Resulting signals are more attenuated and there is more spatial blur.

ADVANTAGE

Non-invasive, it is easy to use, is portable, inexpensive, covers the entire cortical surface and it gives excellent temporal resolution.

5.1 - Neural Encoding

First part of the class: We will introduce the concepts of neuronal encoding and decoding. We will see how we will be using not the information about each individual action potential but rather the information coded into the frequency at which such action potentials are produced by the cell. We will introduce two functions that can be associated with the output of a neuronal cell: the first, the **"normal response function"** is a precise description of the output; and the second which is the **"firing rate"** is more an index summarizing the behaviour of the cell. We will introduce the central concept of **"tuning curve"** which is used to describe the behaviour of each individual neuron or of each family of neurons working simultaneously for the same purpose.

We have learned what's inside the cell. We have learned about the neuronal membrane, the structures in the membrane, the dendrites, the ion channels, the gated ion channels, the axon hill log, the propagation of the action potential along the axon, and the synaptic exchange of information between cells. We will look at the neural as a black box. So we will un-zoom we will move a step backward and we will look at the neuron or as you can see in this picture we can also use this model to represent not just the single cell but also groups of cells which act simultaneously processing the same information that is how the brain is organized. So this black box represents either the single cell or a group of cells working simultaneously to process the same information and here for this application we don't focus on the specific mechanisms inside the box we will focus on the relationship between the input and the output of the box which is the brain region (or sub region or the brain cell).

What's the input and what's the output?

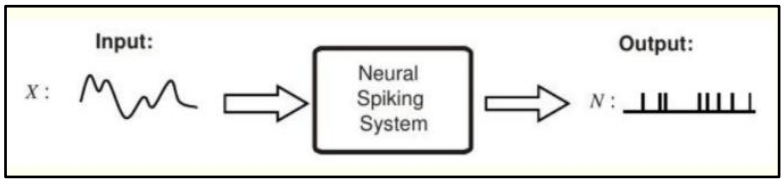


Figure 1

The **output** as we know It is a sequence of action potential produced by the cell as a result of the continuous processing of information received by the cell itself. So, we will have in a time interval in a time window we will have a number of action potentials produced with a temporal distance between them which is affected by the processing which is performed by the cell. So the

output of the cell or the spiking system is a train of action potentials which are also called spikes; so this is the output. The **input** depends on what we want to model so we can decide what is our input here. Usually by this kind of representation we are interested in the specific response of a cell or of a group of cells to a specific aspect of the stimulation that produces that brain activity. So for instance this input can be either the input to the cell so the output of another group of cells because we know that there is no simple relationship between two cells; we usually have thousands of cells affecting as inputs our single cell each time we look at the receiving (the postsynaptic cell). So, this can be a kind of resulting activity of groups of cells or, as we will see today, It can be a property of an external physical stimulation. An example of this can be for instance a visual stimulation in which the input are the variations in time of a specific property of the image that can be the orientation, the contrast, the movement, the distance from the subject… any property of the visual stimulation which we want to study the effect on the specific neuron we are observing. Another example could be a movement made by the subject, can have some properties for instance the direction of the movement can be a specific property of the input to specific neurons in the cortex processing that movement. So the input in our examples will be mainly a physical property of the task or the stimulation that is provided to the subject. So between the stimulation and the neuron we know that there will be other cells but we will focus on the effects of that property of the stimulus on the specific cell that we are studying. So we will focus on single cell recordings: we know that they can be done either in vivo or in vitro, they can be intracellular or extracellular. So extracellular recordings on single cells in vivo will be the way in which the data were recorded.

What are we looking for with the neural encoding and encoding procedure?

We are going to summarize, to describe the behaviour of a single cell or the spiking system with a mathematical function, linking the input with the output. So we are looking for a function of the stimulus or better a function of a property of the stimulus so not a function of the image but rather a function of the orientation of the stimulus, of the visual stimulus or the function of the movement of the angle at which a stimulus is produced and so on. So a function of this property of the stimulus which returns some properties of the neural response which is that spike train. So we aim with this encoding procedure to summarize the behaviour of the neuron by means of a simple mathematical formula linking the input with the output. It's possible and It's interesting to do to look at the neuron in the other way around. So starting from the

response produced by a cell we might be interested in guessing which was the stimulus that produce that response. This is the reverse procedure with respect to the encoding (decoding). We will begin by simplifying things. . Why this? [Empirical experiment: Correlation between the stimulus (pressure in the footpad) and the number of action potentials]. The number of action potentials in the time interval is the definition of frequency. Another reason in support of this hypothesis is that we know that effects of an action potential can be isolated or can be summed up with others and this summation for the same pre-synaptic cell is basically a temporal submission. Meaning that if two spikes are close enough to each other in time their effects will be summed up, and the final effect on the postsynaptic cell will be increased. On the contrary the farther they are the less uh the effect will be on the postsynaptic cell. So, there are both theoretical and empirical proofs of this rate coding hypothesis. Why is this hypothesis so convenient for us? Because It allows to move from the exact description of the output of the cell to a much more synthetic and simple way to describe the cellular output. Let's go through the extended description and the synthetic description.

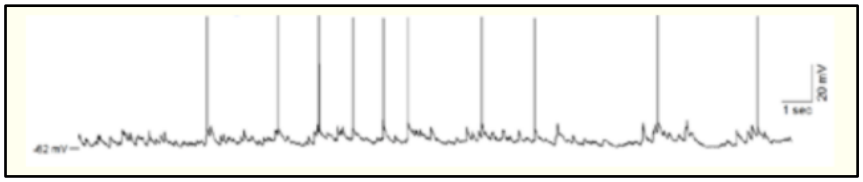


Figure 2

The **extended** description uh needs to report the exact sequence of action potentials that are produced by the cell. So in this example this is a you see that this is an intracellular measurement because we have also the variations of the membrane potential which are below the threshold. So with this accurate recording of the membrane potential of a neuronal cell we see a sequence of action potentials which as we know are very short in duration less than one millisecond, very high in their amplitude (more or less uh 100 millivolts), and they can occur at any time during the processing performed by the membrane of the cell. We know that these action potentials have all the same duration hold all the same shape all the same amplitude. What's interesting is their temporal occurrence. So if we want to completely describe this output we can use a mathematical function which is the Dirac delta, representing each spike each action potential. Since we have more than one we will use this function here

$$\rho(t) = \sum_{i=1}^{n} \delta(t - t_i)$$

Figure 3

which is a train of Dirac deltas. Each of each of them centred on a specific time instance.

The trial is a repetition of the experiment under the same conditions. Usually in any experiment based on neuronal electrical recordings either single cells or groups of cells, intracranial or non-invasive, in all these range of methods we usually perform multi-trial recordings. Meaning that we don't just do the experiment once (we repeat the stimulation, or we repeat the task many times as many as possible) and what you need to take care of is that each trial is repeated under the same circumstances as much as possible. Which means we don't need to change the stimulus, to change the environment, to make different requirements to the subject and so on… We need to keep the conditions controlled so that the task is repeated exactly under the same circumstances and with within temporal windows that must be of the same length. Regardless of the number of repetitions, all repetitions will have the same duration. So T is the trial duration. For each repetition of the experiment the cell will produce a number of action potentials. The number of action potentials produced in the trial is n. So, we have n spikes in a temporal window of T duration, and we will have n Dirac deltas representing this sequence of spikes that It's the output of the cell. From a mathematical perspective we can represent all of this by *Figure 3*, which is called "the neural response function". It is a train of delta of Dirac deltas each centred in a specific temporal instant which is related to the specific spike. This is an exact representation of the output of the cell. From this function we know exactly how many spikes we have and when each of them occurs. This is the most complete description that we can have for the cell output, but this is not what we will be using why because we use the previous definition the firing rate coding hypothesis according to which the information is not coded by the exact temporal instance but rather by the number of spikes which are produced along a task and the temporal distance between them (the rate at which they are produced by the cell).

A simpler definition of firing rate is the "spike count firing rate definition". The spy count is very simple, It's based on counting how many spikes I have in my temporal window and dividing by the temporal window itself. It's important that the temporal direction is expressed in seconds. What happens if I have more than one trial? I will average this result along trials.

Let's move on to what these measure means for our neuronal cells and how we deal with multi-trial recordings. Why do we use the um the firing rate? Because It's the measure of the effect that the neuronal output will have on the following cells; so the higher

the firing rate the higher the temporal summation and higher the postsynaptic potentials that are obtained by mean as a result of this cell activity. However, if the stimulus has some temporal characteristics this is reflected by variations of the processing which will be performed by the cell and so variations and we expect variations in the firing rate of the cell. So, if we want to use the this simple spike count firing rate definition we need to keep the stimulus stationary as I said so we can't just move the properties of the stimulus in time as we want we have to keep them fixed for a temporal interval in which we record the activity of the cell and we have to keep that stationary. The second thing that we need to keep in mind is that this response is not just due to the response of the single cell, is not just due to the stimulus and to the single cell. There are many cells processing the same stimulus and interacting together in networks of neurons. So, the behaviour of the single neuron is mediated by the network, is explained fully explainable only if we look at the entire network. Let's keep in mind that these cells cannot act independently from each other, but they need to be connected in neural networks. We try to keep the conditions stable across trials. During the experiments we need to keep everything as It is the same stimulation the same conditions the subject is kept in the same environment with the same instructions we try to keep everything fixed.

$$\langle r \rangle = \frac{\langle n \rangle}{T} = \frac{1}{T}\int_0^T r(t)dt$$

Figure 4

Since It's all linear [*Figure 4*] what we do is to compute the average number of spikes in each trial, and we will divide It by the trial duration.

What's a tuning curve?

A tuning curve is a graph of neuronal response (usually measured in action potentials or spikes per unit time) as a function of a continuous stimulus attribute, such as orientation, wavelength, or frequency. A neuron is said to be “tuned” for the stimulus that evokes the greatest response, and the width of the curve from the half-maximum response on either side of the peak indicates how broadly or narrowly tuned a neuron is for a particular stimulus attribute. So in our course, It's something that is built in 2 steps: the first step is based on experiments on real recordings and the second step is mathematical, so we look for a function approximating our data. So, there's an empirical step and a theoretical step to build a tuning curve.

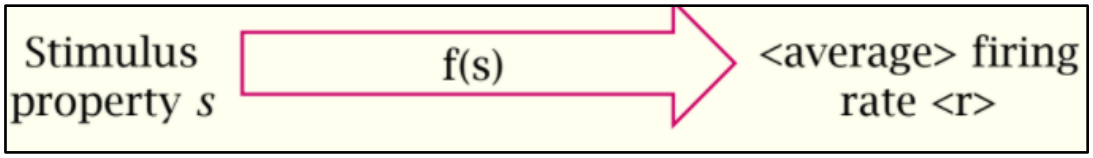


Figure 5

So, let's summarize what we are looking for. We are looking for a function of the stimulus property that can produce or that can predict the firing rate that we will get on average from our cell or from our spiking system. To do that we need to summarize the neural behaviour by this function f. f will explain the behaviour of our neuron. To do that we will start from an experiment from empirical data and then we will look for a mathematical function. We will see how a tuning curve is built and even more importantly we'll see how a tuning curve is used and interpreted.

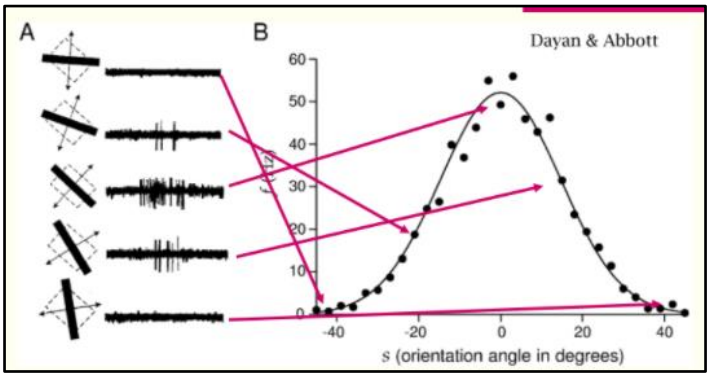


Figure 6

We choose the first orientation, we repeat the experiment for different trials, we compute the average firing rate by means of the spike count definition (so we count how many spikes we have, we divide by the duration of the trial, we average this long trials, and we finally get the number which is the firing rate expressed in hertz, we report this number on the Gaussian as a ●). -45 +45 It's the entire range of positions that the bar can have. By changing the orientation of the bar, we will travel through the entire Gaussian. We finally have enough data to infer the shape of the curve and to look for a function representing approximating these points that we have obtained empirically. Why do we need the function? We need a function to summarize the behaviour of the cell and also to make predictions about points that we have not tested with real data. How can we describe the behaviour of this neuron by looking at this curve? [*Figure 6*] As we can see the neuron responds in a different way to different angles so It's not the same if I use one orientation or another. The maximum response to a specific orientation is 0, that's why we can called It the preferred angle. It's the angle producing the maximum response by the cell, which is between 50 and 55 Hz, and then the response will decrease with the distance from the angle 0. So, when we move

from the preferred angle to other angles the response of the cell will decrease according to this function ending up in being zero or close to zero.

So, we can summarize the behaviour of this neuron by saying that a specific neuron of the visual cortex is devoted to recognizing a specific orientation which is the zero angle. So, the closer the stimulus is to the preferred orientation the stronger will be the neural response. The further the stimulus is from the preferred orientation the smaller will the neural response be.

How do we use this curve? To understand the meaning and behaviour of the neuron and to make predictions. if I have an effect, I am always able to predict which will be the result.

5.2 - The Poisson Spike Generator

Second part of the class: We will introduce a simulator of the neuronal behaviour. We will learn how to describe the neuron not in terms of the specific action potential but in terms of the statistical properties of the neuronal output. What we aim to obtain is not an object behaving exactly like the neuron at any time instant, because even the neuron itself is not able to replicate exactly its behaviour across trials. So we don't want to emulate the identical behaviour of the actual neuron but rather to produce a kind of output which has the same properties of the output of the neuron. We will summarize the statistical properties of the neural output and we will see that we are able to understand what these statistical properties are once we know the firing rates. So, we start from the firing rate (which is obtained from the tuning curve) and then we will produce a generator of output which behaves like the cell, on the basis of the Poisson distribution and on the basis of a simple threshold mechanisms.

Let's start from the meaning of building a stochastic model of the neural response. Do we need to build the stochastic model because we are not aware of what happens in the cell or we are not aware of the mechanisms that produce the cellular output? The answer is no. We are able to build a deterministic model of the neuronal response there are models available which represent exactly the behaviour of the neuronal membrane at any specific point of the cell, everything can be modelled. We need to build a very complex model for any specific neuron to obtain the exact sequence of action potentials produced by that cell this is very computationally demanding, very complex in terms of modelling properties, and finally we have the model of a single cell. We have learned that cells are not acting individually they are acting in a network. If you really want to understand the brain functioning, we need to put these single models of each cell into a network. So, if this neuron the model of the single neuron is very complex to put very complex models in a network would get to a modelling complexity which is impossible to manage. It is not also so useful because finally what we are interested in, is not the exact sequence of action potentials but rather the firing rate of the cellular output.

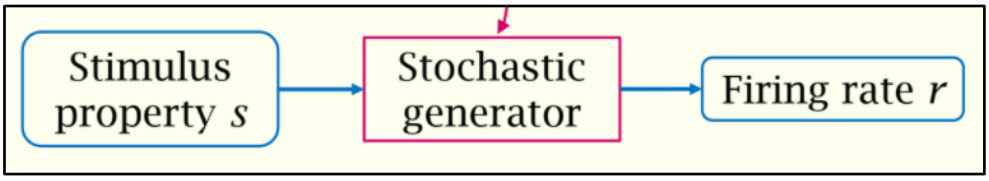


Figure 7

What we are looking for is a stochastic generator, the simulator is acting like a neuron in the stochastic sense that: given a stimulus property s (like the orientation of the stimulus the direction of the movement the retinal disparity whatever property we want to use as an input to our model) It returns a spike trains with the same firing rate that we get on real cells. This is much simpler than having to model all the processes that lead to a specific response by the cell.

What do we need to build this generator? We need two things: the first one **(1)** is a description of the stochastic behaviour of the cell, and this will be provided by a distribution which is the Poisson distribution; and **(2)** we need information about the behaviour of the actual cell in terms of link between input and output. Summarizing, the general behaviour (general rule) is given by the Poisson distribution but the numbers inside the Poisson distribution are provided by experimental data.

So, to build this generator we need again two steps: the first one is the is experimental based on the building of the tuning curve so we need to measure the behaviour of the actual neuron that we want to simulate, and as a result of that we have the tuning curve expressing the firing rate as a function of the stimulus $f(s)$ estimated on our real data. So given the stimulus we get the resulting firing rate. This firing rate we use It in the stochastic generator, in the Poisson distribution (that has as output a spike train) has the same firing rate of our initial data.

Let's go into details and let's start from the **Poisson distribution**. The first important hypothesis that we have to make if we want to use this distribution is that any action potential, we assume It to be independent of the presence or timing of other action potentials (in all the spike train). (However, we will see that It's not completely true that any action potentially is independent from the other, It is true if they are distant in time at least a few milliseconds, but not closer than that.)

Another thing that we need to decide in advance is if we want to consider the firing rate stationary over time or variable over time the same. If we consider the spike count firing rate which is stationary over time during the stimulation and during the response that we want to reproduce then we can use a homogeneous Poisson distribution (which is what we are going to do). If we use a time dependent firing rate, we need to switch to inhomogeneous Poisson processes.

How do we use the Poisson process? Why we use this? What's the meaning of what appears in the distribution when referred to the neuronal response? Poisson stochastic process is used to model or

to reproduce a number of statistically independent discrete events. So, we have discrete events the action potentials statistically independent (integer and non-negative). So, these are the hypotheses under which we are able to use the Poisson distribution to model the behaviour of a stochastic process.

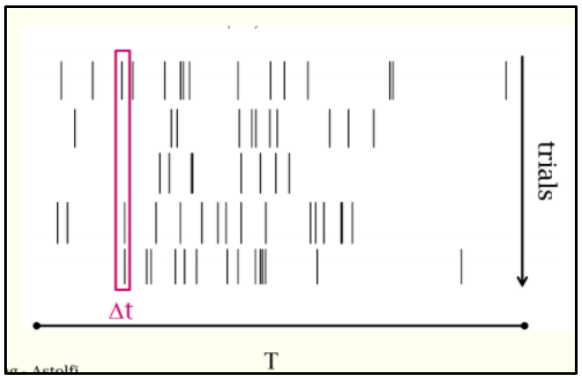


Figure 8

In this example [*Figure 8*] we have 5 trials (the repetitions of the process), the duration T of this trial is fixed. We can consider having an integer number of events, we have to divide the trial into small windows Δt close to 0 or tending to 0. So we divide T into small windows (as small as possible so we can have no more than 1 action potential), and in each of them we can either have an action potential or not okay so we can have a 0 or a 1 in each Δt (this is call event and so can have 0 or 1 action potential

We will focus on two distributions derived from the Poisson distribution process description: the first one is the probability of having n spikes in a trial of duration T (we talk about probability because we cannot be certain of how many spikes we will have); the second distribution is the distribution of the intro spike interval. Why are we interested in the distribution of the intro spike interval? What's the intro spike interval? The intro spike interval It's the temporal distance between two spikes. Of course, It changes along the trial. What's the probability of having an inter-spike interval of tot milliseconds? Let's start from the first point which is the probability of having exactly n spikes in a trial of duration T .

How do we describe the probability of having n spike in a trial of duration T?

$$P_T[n] = \frac{(rT)^n}{n!} e^{(-rT)}$$

Figure 9

This is [*Figure 9*] the probability obtained on the basis of the Poisson process. $P_T(n)$ represents the probability (0 <= $P_T(n)$ <=1) of having an specific number n of spikes during a trial of duration T . The second part of the formula is expressed as a function of n, T and of r which is the firing rate. What's the value of r that we put here into the formula? We put here the value of r which we obtained experimentally from our data, so this is the estimated r we obtained on average across trials.

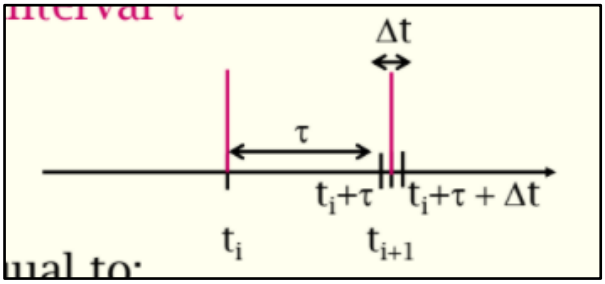


Figure 10

Let's imagine that to compute the probability of a given interval τ we need two spikes (the previous spike t_i and the following spike t_{i+1}, the distance between them is such that $t_i + \tau, \leq t_{i+1} \leq t_i + \tau + \Delta t$) .

The two conditions (that happens simultaneously) for having the probability of having this distance $t_i + \tau, \leq t_{i+1} \leq t_i + \tau + \Delta t$ between 2 spikes are: the probability of having zero spikes during τ; the second condition is that we have then a spike in the following window Δt. So, to have exactly the distance equal to τ between two spikes we need both things to occur.

To have the distribution of the inter spike interval we need to multiply:

probability of no spike for a time τ * probability of a spike in Δt

$e^{(-r\tau)}$ $r\Delta t$

Figure 10, 11, 12

which leads to this formula: $r\Delta t e^{(-r\tau)}$.

The most likely inter spike intervals are short ones (for homogeneous Poisson processes). At the same time, it is dependent on the firing rate, because **smaller is the firing rate the sparser are the spikes in the train.** When I have a stronger r the inter spike interval will be shorter. Since we are in hypothesis of homogeneous Poisson process, it can be demonstrated that the mean of the inter spike interval is inversely proportional to the r and the variance is inversely proportional to the square of the firing rate.

The Poisson spike generator consist in 2 steps:

1. Experimental estimation of the firing rate r_{est} (tuning curve)
2. Then we are able to build a spike generator with $r = r_{est}$, and based on a vity mechanism.

At each fixed size time (Δt) steps (time size of a single action potential) the system produces a random number chosen uniformly between 0 and 1. The random number is independent from the threshold which changes randomly at each step. Then for each Δt the system multiplies by the estimate firing rate (r_{est}) by Δt and compare the result with the threshold. If it is greater the threshold, we will have an action potential; if it's below the threshold we won't have an action potential.

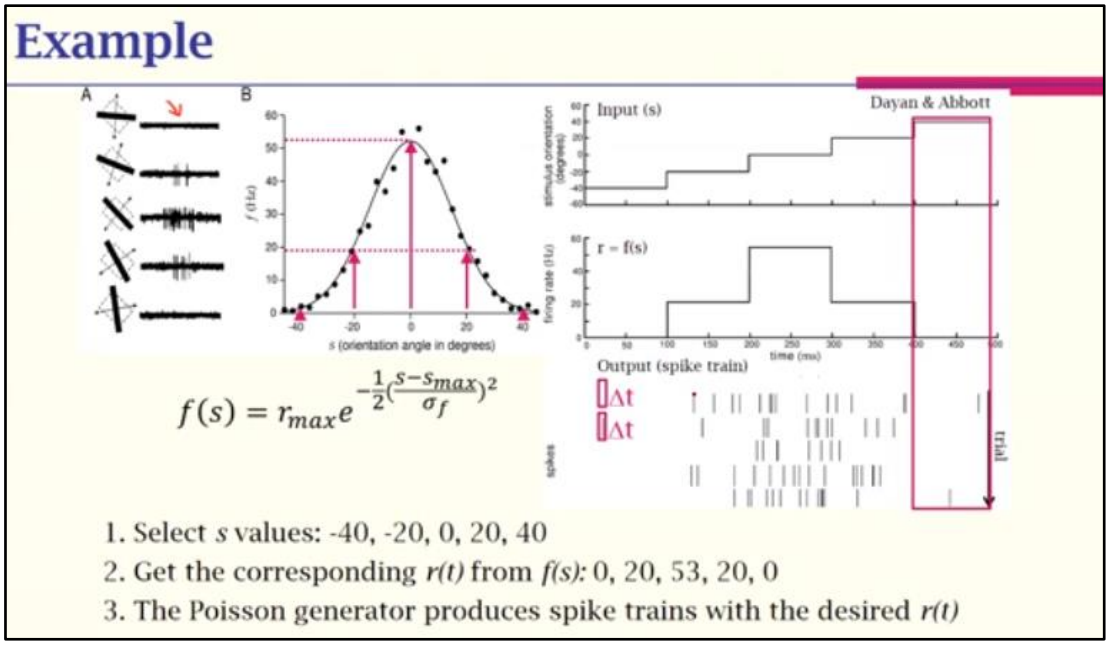


Figure 13

Our aim is to reproduce a neural behaviour the most similar possible to the real one. To do that the process should be:

1. Choose a value of the stimulus $\boldsymbol{s}$
2. We derive the relating firing rate $\boldsymbol{r}$
3. For any $\boldsymbol{\Delta t}$ we multiply the $\boldsymbol{r}$ by $\boldsymbol{\Delta t}$ and we compare the result to the random threshold generated at any temporal interval (0:1).
4. The Poisson generator produces spike trains with the desired $\boldsymbol{r}$

https://drive.google.com/drive/folders/1iZ9aGFHhOQRK4k0AzoC77fwp4NYuHBz (16:00-31:00).

LIMITATION OF POISSON GENERATOR

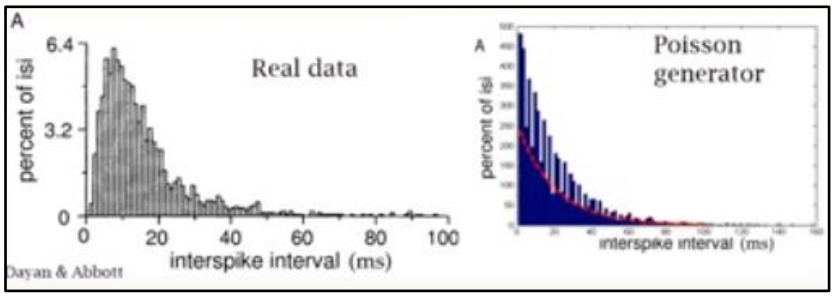


Figure 14

In comparison with real data, **it works quite well** (looking the trend), the main differences are the lower values of $\boldsymbol{\tau}$ (short

distances) due to the hypothesis of the independent spikes. **In the real word we have two refractory period**, in particular in the relative one (couple of millisecond after the absolute refractory period) it's very unlikely to have another action potential. **This fact goes in contrast with our previous hypothesis.**

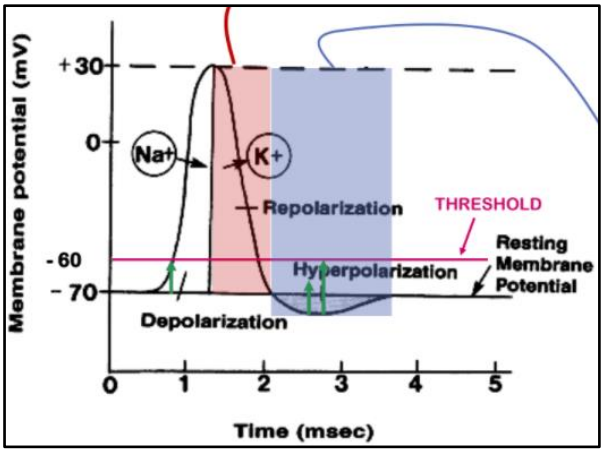


Figure 15

We know that there's a refractory period and more specifically we have two refractory periods: the **first one** which includes the shorter distance between two spikes is an **absolute refractory period** which is due to the inactivation of the sodium voltage-gated channels. During this temporal interval no action potential can be generated. The **second one** is the **relative refractory period** which is due to the potassium voltage-gated channels which produce and hyperpolarization of the membrane and during this second temporal interval (which is of about a couple of milliseconds after the after the peak of the action potential) it is not impossible to have another action protection potential, it's just much less likely. This is the reason why we have that difference in the distribution of the inter spike intervals.

So, what can we do either accept the limitation of the generator or we can try to correct the generator by including the refractory periods. This means that we have to constrain the minimum inter spike distance. So, we start from the Poisson generator which is again a very simple algorithm, and we add up some constraints to the generator.

We need to include the **strong constraint** for the shortest temporal distance between two spikes to simulate the absolute refractory period, and the **weaker constraint** for the following temporal interval to simulate the relative refractory period.

CONSTRAINED POISSON GENERATOR

We have to include the strong constraint of the absolute refractory period and another weaker constraint for the relative one.

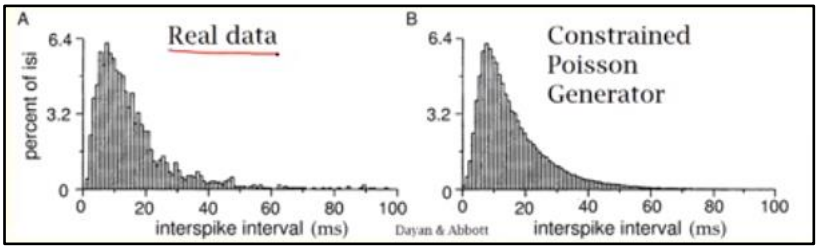
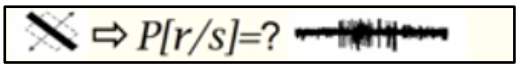


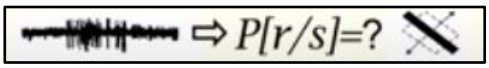
Figure 16

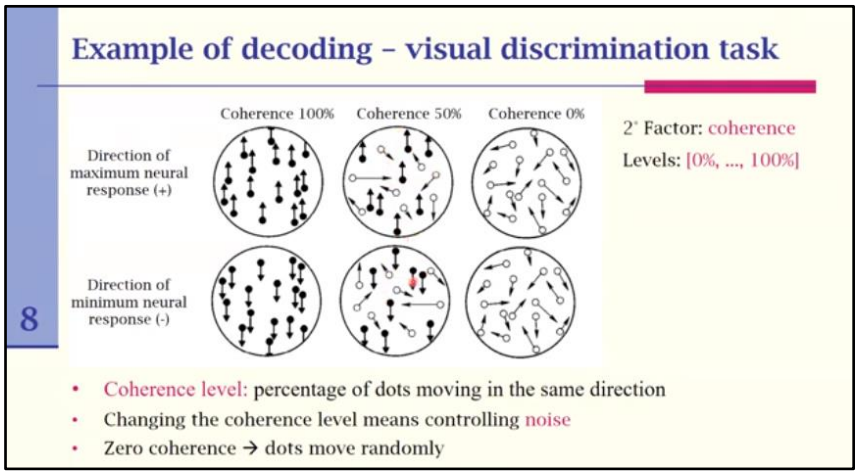
So, to sum up: we can even use the unconstrained generator, but we need to expect a different behaviour with respect to real data, if we are just interested in the firing rate it can be okay. On the other hand, if we are interested in the more realistic and the more likely behaviour, we need to improve the model by adding these constraints.

6 - Neural decoding

Consists in guessing. **Find the stimulus that more likely produces the spike train that it has induced.** It is important in order to understand the behaviour of our brain and to project devices that have to interface with brain. I need the encoding to know how to perform the decoding so they are strictly related each other.

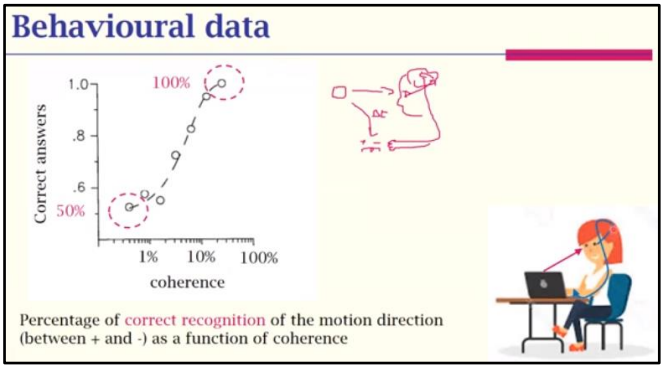
Encoding: 

Decoding: 



It is a visual discrimination task: in this case consists in the recognition of motion direction.

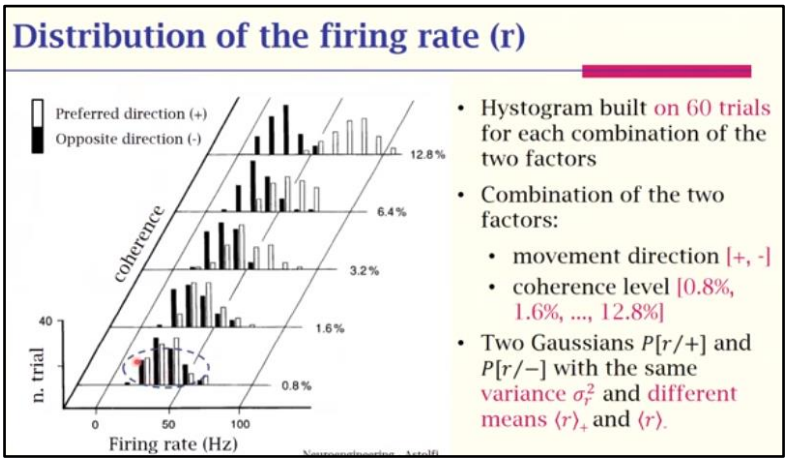
Two types of motion: from the bottom to the top (**+**) and vice-versa (**-**). Noise is increased by adding dots moving in random directions, the amount of noise will influence the amount of coherence. Zero coherence means all random movements.



(see the logarithmic scale). We can note how after the 10% of coherence our brain reaches almost the 100% of correct answerers.

https://drive.google.com/drive/folders/1iZ9aGFHhOQRK4k0AzoC77fwp4NYuHBz (1:06:00-end).

DISTRIBUTION OF FIRING RATE (w.r.t previous example)



Our firing rate is between 0 and 100 Hz. The Graph combinates different combinations of variables, like directions and coherence, and reports a different distribution for both directions. At the beginning (low coherence) the two distributions are mixed together, and the result is a well-shaped gaussian. **This result comes from a complete scrambled image and the brain is not able to recognize a main direction due to the noise.**

We can see after how the white distribution, with the increasing of coherence, slides to the right. This means that the cell which causes this white firing rate distribution is sensitive to the increasing of coherence and vice versa for the black distribution. This is the reason why the + direction is called preferred (the neuron changes the response) and – direction is called opposite.

THRESHOLD CLASSIFICATION

I aim to recognize the stimulus by looking the output (like the graph before) without do mistakes.

Discriminability: the capability to distinguish two conditions/cases.

The Discriminability when two different distributions are gaussian can be defined as the difference between their means divided by the standard deviation.

$$d' = \frac{< r^+ > - < r^- >}{\sigma_r}$$

This will help to predict which condition will be easier for us to discriminate between the two directions.

So given a threshold, between the two distributions, ***z***:

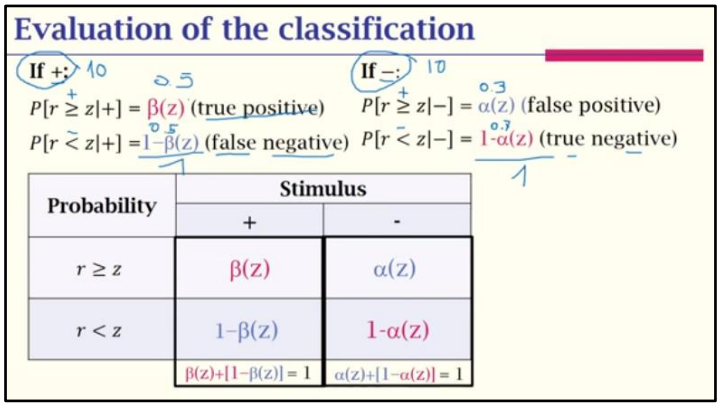
- If ***r*** >= z ➔ the inferred direction is **+**
- If ***r*** <= z ➔ the inferred direction is **–**

Ideally the threshold should be the intermediate points between the means of the distributions. **Bigger is the discriminability the better is the placing of the threshold** because it ensures areas of indecision.

EVALUATION OF CLASSIFICATION

- **TP**
- **TF**
- **FP**
- **FN**

They 4 types of results depend on *z*.

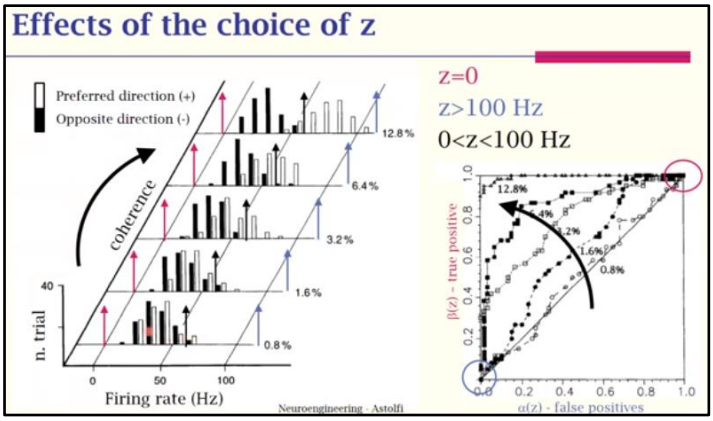


As we can see the **two rows of the table are not independent**, but the two columns instead are independent. I need only, for example, the first row to describe entirely my model. Ideally a want that **TP** = 1 and **FP** = 0.

ROC

First, I need a couple from the four possible outcomes, for example: **TP** and **FP**.

The ideal classifier in on the top left of the graph. Each dot represents a value of ***z*** and there is a curve for each level of coherence. If the value of **z** is 0 all the points of the **–** direction will be detected above the **z** so they will be classified **FP**, but I will have all **TP** for the **+** direction, same but inverted with choosing the **z** equal to 1.



With the increasing of the coherence the discriminability will improve and also choosing the ideal z I will achieve the best solution. I can use the curves to select the better value for z. To quantify the result of the model we the **AUC** (area under the curve). The minimum **AUC** is 0.5 that is the minimum amount of good response (0.8 of coherence), the maximum **AUC** is almost 1. **In this case the classifier performance, using the threshold mechanism, are comparable to a real subject.** A single neuron shows the same behaviour of the human so we can say that there is a family of neural cell that together describes the behaviour of the human by comparing the behaviour with the classification of the single cell.

7 - Brain Networks

So far, we have analysed each signal independently from other (**Univariate Analysis**). But in complex systems it is necessary to analyse the signals **and** their **interdependency** (**Multivariate Analysis**).

The brain is a complex system, and it is possible to build, at different scales (molecules, neurons, brain areas and social systems), dynamic **brain networks**. The analysis of the brain in the perspective of brain networks is called **network neuroscience** and aims to **map**, **record**, **analyse and model** the elements and interactions of neurobiological systems.

Brain connectivity is a **description of brain networks**. These networks can have different natures: **anatomical**, **functional**, **effective**.

- **Anatomical**, physical or structural connections linking sets of neurons or neuronal elements. Relatively **stable** and **shorter time scales**, thus, during a recording session of a couple of hours we do not expect changes. Anatomical connectivity is at the basis of functional/effective connectivity, but **cannot explain it**.
- **Functional**, it is essentially a description of what is happening in the brain in terms of similarity between the activity in different regions. Usually it is based on **correlation**, **coherence**, and **transfer entropy**.
- **Effective**, it refers explicitly to the **influence** that one neural system exerts over another, either at a synaptic or population level. It corresponds to **parameters of a model** that tries to **explain observed dependencies**. Based on concept of **causality**.

Synchronization of brain regions is one of the most important aspects to understand brain functioning.

Hebbian theory: cells that **fire** together **wire** together.

Frequency analysis of time series

A time series is a not continuous signal *x[n]* with a sampling interval *T*.

Autocorrelation function

$$r_{xx}[k] = \sum_{n=-\infty}^{\infty} x^*[n]x[n+k]$$

K is the lag between two samples.

It is aimed to describing the temporal evolution of the time series, focusing on similarities internal to the time series in time.

Cross-correlation function (given two time series *x[n]* and *y[n]*)

$$r_{xx}[k] = \sum_{n=-\infty}^{\infty} x^*[n]y[n+k]$$

Focuses on similarities between two time series.

Fourier transform:

$$X(f) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j2\pi fnT}$$

It transforms the time signal to the frequency domain.

Power Spectral Density (PSD):

$$S_{xx}(f) = |X(f)|^2$$

How the power of the signal is **distributed** at different frequencies. It is convenient to focus on PSD with respect to Fourier transform, in fact, while it is not always possible to compute the Fourier transform, the **PSD always exists.**

Wiener-Khinchin Theorem: PSD can be computed by Fourier-transforming the autocorrelation function. Useful if the Fourier-transform of the signal does not exist.

$$S_{xx}(f) = \sum_{n=-\infty}^{\infty} r_{xx}[k]\, e^{-j2\pi fnT}$$

Note that $r_{xx}[k]$ can always be Fourier-transformable.

Given two time series *x[n]* and *y[n]*

Mutual Power Spectral Density: it represents how the two signals **share** power at different frequencies.

$$S_{xy}(f) = X(f)Y^*(f)$$
$$S_{yx}(f) = Y(f)X^*(f)$$

It is **symmetrical**, so $\boldsymbol{S_{xy}(f) = S_{yx}(f)}$

Wiener-Khinchin Theorem: PSD can be computed by Fourier-transforming the cross-correlation function of the two signals

$$S_{xy}(f) = \sum_{n=-\infty}^{\infty} r_{xy}[k]\, e^{-j2\pi f n T}$$

It is possible to store all the spectral information in a mathematical object, called **spectral matrix *S(f)***.

$$S(f) = \begin{matrix} S_{xx}(f) & S_{xy}(f) \\ S_{yx}(f) & S_{yy}(f) \end{matrix}$$

It is **symmetrical**. Indeed, we know the two signals share power, but we do not know the direction of this exchange.

Ordinary coherence

Coherence means a frequency version of correlation in time, how signals are synchronized in the frequency domain.

Ordinary coherence is the **linear correlation** between **two signals** at a given frequency:

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{|S_{xx}(f)||S_{yy}(f)|}$$

The numerator is proportional to how much the two series are synchronized at a specific frequency. This would be sufficient to determine how much the two signals are synchronized and at which frequency, **but** we need to **normalize** it to take into account **the ratio of exchanged power on the total power of the signals**. Indeed otherwise, we would not be able to understand if a specific value of the numerator were high or low.

The **maximum** of the power that the two signals can share is equal to the product of the two individual energies.

The result is **bounded in the interval [0, 1]** and thus it represents a percentage.

Ordinary coherence is symmetric $C_{xy}(f) = C_{yx}(f)$, so it does not give any information about the **direction** of the interaction. It measures **synchronicity** but not **causality**. It is bivariate, what if there are more than two signals?

Correlation is **NOT causation**.

It is not possible to infer causation just by looking at correlation, we first have to consider all the possible, even external, sources.

8 - Brain Networks

Causality:

- **Physical influence:** is when changing causes changes their consequences. To test this kind of causality it is necessary to stimulate the brain during its activity. **Interventional** procedure.
- **Temporal precedence:** causes precede their consequences. Based on observation of the system, not requiring perturbing it, testing for improvement in **predictive capacity** between temporally distinct neural events. **Observational** procedure. Less invasive and accurate. Is more practical than the physical one.
- **Statistical definition of causality (given by Wiener):** given two **simultaneously measured** signals, if one can **predict** the first signal better by incorporating the **past information** from the second signal than using only information from the first one, then the second signal can be called **causal** to the first one.

Strong points of the statistical definition are:

- It does not require any **intervention.**
- It gives **directionality.**
- It is relatively easy to compute.

What we are missing now is a way to perform the predictions and compare them, we need a **model.**

For this reason, has been introduced the **Wiener-Granger causality**:

An observed time series $a(n)$ is said to **Granger-cause** another series $b(n)$ if knowledge of *a(n)'s* past significantly improves prediction of *b(n)* by an **autoregressive (linear) modelling.**

- *b* is modelled on the base of its own past, trying to predict next sample of *b*:

$$b(n) = B_1 b[n-1] + \cdots + B_N b[n-N] + e_B(n)$$

- *b* is modelled on the base of its and *a* past, trying to predict next sample of *b*:

$$b(n) = B_1 b[n-1] + \cdots + B_N b[n-N] + A_1 a[n-1] + \cdots + A_M a[n-M] + e_{B,A}(n)$$

We have that *a* **Granger-cause** *b* if $e_{B,A}(n) < e_B(n)$, where $e_{B,A}(n)$ and $e_B(n)$ are the **residuals or prediction errors.**

These gives us the directionality, but just in one direction.

Linear Autoregressive (AR) Model:

$$\boxed{x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + e[n]}$$

x[n] is the **time series**

p is the **model order**, in EEG is usually between 10 and 20.

a[k] is the **autoregressive parameter**, lag *k*

e[n] is the **model residual**

With this model it is possible to represent signals that respect following **hypothesis**:

- *x[n]* wide-sense stationary (do not change the stimulus during experiment)
- *e[n]* zero mean, uncorrelated white noise. Constant value in the frequency domain.

An AR model can be used as a **linear predictor**, meaning that the linear combination of the past samples is an estimation of the next one.

$$x'[n] = -\sum_{k=1}^{p} a[k]x[n-k]$$

$$e[n] = x[n] - x'[n]$$

x[n] is the recorded value, while *x'[n]* is the predicted one. *e[n]* is the prediction error.

We want to determine which coefficients a are the ones that minimize the squared prediction error (not part of the course).

Bivariate Autoregressive Modelling:

The autoregressive prediction of *y* is made by including information about **the past samples of another signal** *x*.

This has *x* as target

$$x[n] = \sum_{k=1}^{p} a_{xy}[k]x[n-k] + \sum_{k=1}^{p} b_{xy}[k]y[n-k] + e_{xy}[n]$$

This has *y* as target

$$y[n] = \sum_{k=1}^{p} a_{yx}[k]x[n-k] + \sum_{k=1}^{p} b_{yx}[k]y[n-k] + e_{yx}[n]$$

The two residuals will be in principle different from the ones obtained by the univariate model; indeed, it depends on the past values of **both** signals.

We are mainly interested in comparing these residuals, but doing it sample by sample would be noisy. So, we will use their **variance**.

Granger causality test:

The prediction performances for both models can be assessed by the **variances of the prediction errors**:

- For **univariate** models:

$$V_{x|x} = var(e_x) \text{ and } V_{y|y} = var(e_y)$$

- For **bivariate** models:
  - $V_{x|x,y} = var(e_{xy})$ and $V_{y|y,x} = var(e_{yx})$

Where X|X and X|X,Y indicate predicting X by its past values only and by past values of both X and Y, respectively.

If $V_{x|x,y} < V_{x|x}$ then Y **causes** X in the sense of **Granger causality** and a measure of that is given by:

$$G_{\{y\to x\}} = \ln(\frac{V_{x|x}}{V_{x|x,y}}) \quad \neq \quad G_{\{y\to x\}} = \ln(\frac{V_{y|y}}{V_{y|y,x}})$$

Let's analyze for instance $G_{\{y\to x\}} = \ln(\frac{V_{x|x}}{V_{x|x,y}})$

This is testing the existence of causality directed from *y* to *x*.

We have to use the equations defining the value of *x[n]* and in particular we have to focus on their **residuals** using them in the above formula.

We can end up with three possible outcomes:

- The two residuals $e_x[n]$ and $e_{xy}[n]$ are **equals** and in particular:

$$V_{x|x} = V_{x|xy}$$

  So, the past of the second signal added in the bivariate model is **useless**, we can conclude that there is no cause effect directed from *y* to *x*.

$$G_{\{y\to x\}} = \ln(1) = 0$$

- There is a reduction of the residual due to the past of the second signal, thus:

$$V_{x|x} > V_{x|xy}$$

$$G_{\{y\to x\}} = \ln(> 1) = \uparrow$$

This yields to an improvement of the index, meaning that there is a **causality**.

- The third case is when there is an increase of the residual of the bivariate model with respect to the univariate one. This cannot happen, adding new information to the same used previously will never produce a worsening of the performances.
  The **Granger** index should never be lower than 0.

$$G_{\{y \to x\}} \in [0, \infty]$$

The higher is the improvement of the index value, the higher is the strength of the causality.

Advantages:

- **Directionality** $G_{\{y \to x\}} \neq G_{\{x \to y\}}$
- **Statistical** meaning

Limitations:

- Defined in **time domain**, we have lost the frequency information.
- **True causality** can **only** be assessed if the set of two time series contains **all possible relevant information and sources** of activities for the problem.

Yes, but…What if we have more than two time series?

Let's analyse the **pairwise** and **multivariate** approaches.

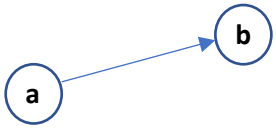
If we have only **two-time** series, we simply look at the temporal precedence, we apply the definition, we use the model and end up with a value for the **Granger causality** index, quantifying the causality.

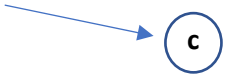
Pairwise approach

A first simple idea is to consider the signals taken pairwise analyzing them couple by couple using the methods seen before (Granger definition, ordinary coherence…). This not means we have just two time series, but that we analyze the time series taking each possible combination of two time series.

But, the set of two time series does not contain all possible relevant information. This recalls the second **limitation** of the **Granger** definition.

To explain this, let's consider a set of three signals ***a***, ***b***, ***c***. Their causal relation is depicted below:





So, there is no direct causal effect between *b* and *c*. But they are not independent, they both suffer from the common effect of sample *a*. Thus, using the Granger causality index we will end up with **spurious link** between *b* and *c* which is not due to any true causality, neither direct nor indirect, but just to the common effect of sample *a*.

Any pairwise model analysing the relation between signals *b* and *c* **cannot recognize** that the statistical link between them is due to the **common effect of signal *a***, which is not included in the model.

This is the problem of the **confounding source**, and it gets worse and worse while increasing the number of signals.

To solve this problem, we can use a **multivariate approach**.

The connectivity pattern is obtained by a unique generation model estimated on the entire set of data and takes into account all their interactions.

Multivariate methods, by building a **unique model** base on **all the (measurable) signals**, use **all the information** at disposal and thus allow a **better comprehension** of the relationship between the signals. No **spurious link** due to not considered common external sources.

A problem that even this kind of models cannot solve is the presence of signals that we have not been able to record. Problem of the **hidden source**.

Using multivariate approach will not solve totally the problem of spurious links, but it reduces it a lot.

Multivariate Autoregressive Models (MVAR):

It is basically very similar to the bivariate one, but combines the past of the whole set of samples in the dataset.

If we have N time series, we will have N equations. For instance, let's analyse one

$$x_1[n] = -\sum_{k=1}^{p} a_{11}[k]x_1[n-k] - \sum_{k=1}^{p} a_{12}[k]x_2[n-k] - \cdots - \sum_{k=1}^{p} a_{1N}[k]x_N[n-k] + e_1[n]$$

That is the linear combination of its own past, plus the pasts of all the other samples. Again, we will have a **residual**, defined again as the difference between the estimated sample and the

measured one. It is essentially the same model, bigger and with a new **problem**:

In the bivariate approach we focused just on the residuals, comparing the one obtained with the bivariate model with the univariate one. But if we do the same here, given, for instance, a reduction of the residual, we would know that there is a signal in the dataset linked to the target one, but we would not know **which**.

Here we have a lot of possible causes to which it is possible to attribute this improvement in the prediction.

A possibility is the **partial analysis**, according to which you build many models in which each time you remove one signal and compare the results trying to understand the source. We will not consider this.

Another possibility shifts our attention from the residuals, that are not specific for a couple of signals, to the **parameters** (e.g. a_{12}). Each set of parameters is specifically referred to a couple of signals (note that $a_{12} \neq a_{21}$). Another advantage of this method is that, since the parameters are a sequence of numbers, we can also derive some information about the **frequency content of causality** linking those regions. (with residuals this was not possible since they are numbers and constant in frequency by hypothesis).

The model parameters are N*N*p, one set for each couple of signals and in both directions.

An example: $a_{ij} = [a_{ij}[1], \dots, a_{ij}[p]]$

The goal of the next method is to merge the advantages of the methods seen so far: the directionality of Granger index, and the spectrality (meaning that it is in frequency domain) of ordinary coherence.

First thing to do is to transform the **MVAR** model in the frequency domain.

Starting from this equation we can bring all the summations to the first member, including the term $x_1[n]$ in the first summation.

$$x_1[n] = -\sum_{k=1}^{p} a_{11}[k]x_1[n-k] - \sum_{k=1}^{p} a_{12}[k]x_2[n-k] - \cdots - \sum_{k=1}^{p} a_{1N}[k]x_N[n-k] + e_1[n]$$

So,

$$\sum_{k=0}^{p} a_{11}[k]x_1[n-k] + \sum_{k=1}^{p} a_{12}[k]x_2[n-k] + \cdots + \sum_{k=1}^{p} a_{1N}[k]x_N[n-k] = e_1[n]$$

With $a_{11}[0] = 1$

Let's now focus **on a single summation per time,** it can be re-written as:

$$\sum_{k=0 \vee k=1}^{p} A_{ij}[k]X_j[n-k] \; = E_j[n]$$

Considering A[k] as a time series $a_{ij}[k] = a_{ij}[0], a_{ij}[1], \dots, a_{ij}[p]$ it is possible to see the above formula as the **convolution** of the two signals. Then, there is a theorem according to which the convolution of two signals in time corresponds to the product of their Fourier transformed version, thus we can re-write the above equation in frequency as:

$$A_{ij}(f)X_{ij}(f) = E_j(f)$$

Where $A_{ij}(f) = \sum_{k=0}^{p} a_{ij}[k]e^{-j2\pi fTk}$

So, any of the summations of the time equation is reduced to this equation.

$$\sum_{k=1}^{p} a_{ij}[k]x_j[n-k] \; => A_{ij}(f)X_j(f)$$

Even the residuals are transformed, obtaining a vector *E(f)*.

For the whole model we will end up with

$$A(f)X(f) = E(f)$$

Where:

$$A(f) = \begin{matrix} A_{11}(f) & \dots & A_{1N}(f) \\ \vdots & \ddots & \vdots \\ A_{N1}(f) & \dots & A_{NN}(f) \end{matrix} \; ; \; X(f) = \begin{matrix} X_1(f) \\ \vdots \\ X_N(f) \end{matrix} \; ; \; E(f) = \begin{matrix} E_1(f) \\ \vdots \\ E_N(f) \end{matrix}$$

Note that the **directionality** is preserved, indeed *A(f)* is not symmetric (differently from the spectral metric).

Partial directed coherence (PDC):

- **Partial** because (because it is **multivariate**) we look for the interaction between two time series in a larger domain.
- **Directed,** because it measures **causal** effects with its **directionality.**

- **Coherence** because it is spectral, it is a measure of a link in the **frequency domain** (like ordinary coherence).

So, we can say it is **multivariate**, **causal**, **spectral** index of connectivity.

From *j* to *i* is defined on the basis of the matrix *A*:

$$\pi_{ij}(f) = |A_{ij}(f)|^2$$

Different **normalizations** of PDC are provided, for instance:

$$\pi_{ij}(f) = \frac{|A_{ij}(f)|^2}{\sum_{m=1}^{N}|A_{im}(f)|^2} \text{ , where: } \sum_{n=1}^{N}\pi_{in}(f) = 1$$

The denominator represents the sum of all values arriving to *i* including *j*. This index is bounded in the interval [0, 1].

The PDC assumes maximum value of 1 when the considered source (*j*) is the only one influencing the target *i*.

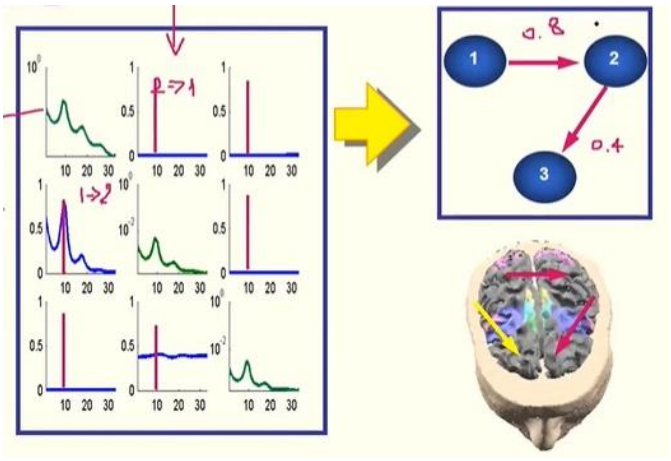
PDC value represents a percentage of the target's power spectral density due to a certain source.

The value of PDC_{ij} at a certain frequency f_0 represents the existence of a causality link **directed from j to i**.

$$\pi_{ij}(f) \neq \pi_{ji}(f)$$

PDC_{ij} estimates the influence of the region j toward the region i at a given frequency f.

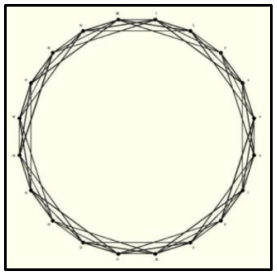
So, we end up with a matrix of PDC functions then, to build a **network** starting from it we need to consider a chosen frequency (e.g. 10Hz) and link with an arrow the nodes representing the signals, when the PDC has a value > 0 (or > of a fixed threshold), and weight this arrow with the value of the PDC function.



Multivariate approach:

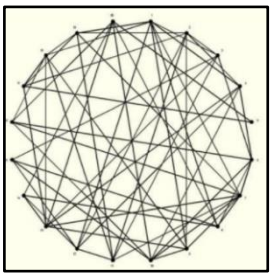
- **Advantages:**
  - **Better estimation** performances.
  - Allows for inserting **all data sources** in the model.
- **Limitations:**
  - Limitation in the number of channels/signals that can be modelled -> **more data required.**

Regular Networks



- Each node is linked to a small number of other nodes.
- The degree is the same for all nodes.
- Efficient communications between small groups, un-efficient communications at the entire network level

Random Networks



- Each node is linked to the others randomly.
- There are no small groups with a strong internal communication

9 - Graph theory

A **graph** is a way to represent a network. We consider two objects:

- **Nodes**, that represent an entity. For us nodes are neurons, electrodes, brain regions…
- **Edges** or links, that make connections between nodes assuming different semantic meanings.

Main properties of a graph are **directionality** and **weight.**

Four graph categories: **undirected**, **directed**, **binary**, **weighted.**

The weight can depend either on the way in which the network is built and on the way we look at it. It is possible to make a weighted graph binary just by imposing a threshold and transforming to 0 all the links below the threshold and to 1 the ones above. We use binary graph because we want to focus and the topology of the network and because it is easier to compute the indices. The directionality and weight of graphs are independent, so it is possible to have any combination of the properties, in general we will focus on binary graphs.

Birth of graph theory attributed to **Euler** in 1786. Social networks are represented as a binary, undirected graph.

Graph representation is not only a graphical support, it is useful because we associate to it a **matrix** that allow us to compute indices able to describe properties of the network that sometimes can be difficult to extract just looking at it.

It allows to extract quantifiable, objective, measurable and concise indices to be used:

- For a statistical comparison between conditions, time points, groups of subjects…
- As markers of specific pathological conditions
- As features for a classification aimed to decode the brain activity

Adjacency matrix is a way to represent graphs. Each element of the matrix represents a link between each couple of nodes. If the element is equal to zero, there is no link between those two nodes, if the element is different from zero, there is a link between them (in weighted graphs the element assumes as value the weight of the link).

If the graph is undirected the matrix will be symmetric.

Basic measures

Density: *density **k** of a binary graph is the number of connections which are present in the graph normalized by the maximum number of edges that we could have.*

$$k = \frac{L}{L_{tot}} \text{ with } L_{tot} = N(N-1)$$

It is **necessary** to normalize it to give a significant meaning to the result.

Since for undirected graphs the adjacency matrix is symmetric, the number of existing arcs can be determined from **half** of the matrix. Moreover, the total number of nodes will be

$$L_{tot} = \frac{N(N-1)}{2}$$

In undirected graph the total number of nodes is halved with respect to directed ones.

For directed graph to compute the actual number of nodes present in the graph we need the **full** matrix.

Application: check effects of therapy that in some case can cause a rearranging of our brain networks increasing the density.

Degree: *the degree of a node i is the number of edges connected to that node.*

$$g_i = \sum_{j=1, i \neq j}^{N} a_{ij} \text{ with } g_i \in [0, N-1]$$

Density is a network global measure, instead degree is a **local** measure computed for each node. To compute it we need the **full** matrix even for undirected graphs. For undirected graphs the degree of a node is the sum of the ones on the corresponding row or column.

The node with highest degree is the one with major role in the graph.

For **directed graphs** we can distinguish between:

- **In-degree** of a node, as the number of edges **directed to** the node

  $$g_i^{in} = \sum_{j=1, i \neq j}^{N} a_{ji} \text{ with } g_i^{in} \in [0, N-1]$$

  Sum of the ones on the adjacency matrix **row** corresponding to the node.
- **Out-degree** of a node, as the number of edges **originated from** the node

  $$g_i^{out} = \sum_{j=1, i \neq j}^{N} a_{ij} \text{ with } g_i^{out} \in [0, N-1]$$

  Sum of the ones on the adjacency matrix **column** corresponding to the node.

- **Degree** of a node, as the total number of edges originated from and directed to the node

$$g_i = g_i^{in} + g_i^{out}$$

In neuroscience is useful to understand if a region is important in collecting or in generating information for other regions.

Distance: *logical distance measured in terms of number of steps needed to move from a node to another.*

A **path** is **any possible sequence** of edges linking two nodes. There can be multiple paths for the same couple of nodes. The **path length** is given by the number of edges (**not** by the physical distance).

Distance *d(i, j)* between nodes *i* and *j* is the **shortest path** between them.

For directed graphs we have to take care of the direction of the links. When a node it is not reachable from another one, their distance is **infinite**.

It is possible to define a **distance matrix**, it has 1 in the positions in which there is a 1 in the adjacency matrix; an infinite where there is no linking between two nodes. On the main diagonal there are all zeros. The other positions have to be computed.

The lower is the average distance between nodes the more is the network **efficient**.

Global efficiency: *the global efficiency E_g of a graph is the **average** of the **reciprocal** of the **distances** between any pair of nodes.*

$$E_g = \frac{\sum_{i,j=1,i\neq j}^{N} \frac{1}{d_{ij}}}{N(N-1)} \quad \text{with } E_g \in [0,1]$$

d_{ij} is one element of the distance matrix.

The denominator represents the total number of possible edges, we divide by it because the highest efficiency we can get is when the matrix is fully connected, and so is full of ones. To take this into account we divide for the maximum number of possible links.

If the graph is **fully disconnected** the global efficiency is **zero**. When all the nodes are connected with the minimum distance (1) the global efficiency is **one**.

Global efficiency measures how efficiently the information is exchanged in the network, meaning how fast in terms of number of steps to reach the destination.

For undirected graphs it is needed to take into account just **half** of the matrix and normalizing for **half** of the maximum number of links.

$$E_g = \frac{2\sum_{i,j=1,i\neq j}^{N}\frac{1}{d_{ij}}}{N(N-1)} \text{ with } E_g \in [0,1]$$

In general, it is not true that having all direct link is the best option because they have not so high **local efficiency** and it is not convenient in terms of energy, cost, resources spent by the network.

10 - Graph theory

(Average) Local Efficiency: *formal definition?*

It gives an idea of how well the network is organized. Is a measure of how much the network is robust to the removal of a specific node.

For each node *i* we can extract a **subnetwork** S_i made of **all the nodes directly connected** (both incoming and outcoming) to *i* (**but not including *i* itself and all its edges in and out**) and compute its **global efficiency** $E_g(S_i)$. The average of $E_g(S_i)$across all the nodes is called **local efficiency**.

$$E_l = \frac{1}{N}\sum_{i=1}^{N} E_g(S_i)$$

A node *i* is more important for the graph if, once built its subgraph and removed *i* from it, there are some nodes that become disconnected from the rest of the graph.

The lower is the local efficiency of a node the higher is its importance.

Modularity: *Modularity Q measures the tendency of the subnetworks to form communities*

A community is a group of nodes that has more connections between them with respect to the average number of connections they have with any nodes in the network.

Integration in graph sense: a network is more integrated when the parts that make it are more strictly interconnected. Higher is the density, the global efficiency higher is the integration, meaning that the network works more as a unique system.

Segregation means that the network can be divided in subnetworks, the higher is the local efficiency the more the network is segregated. A network with high local efficiency is a network that tends to form sub-groups

Integration and segregation are one the opposite of the others.

Higher the modularity higher the segregation of the network.

To compute modularity, we need to define **two or more classes** (group) of nodes. For simplicity we will consider just two groups. On the base of them it is possible to compute the **modularity**.

For **undirected graphs**:

$$Q = \frac{1}{L}\sum_{i,j=1}^{N}(a_{ij} - \frac{k_i k_j}{L})\delta(C_i, C_j)$$

For **directed graphs**:

$$Q = \frac{1}{L}\sum_{i,j=1}^{N}(a_{ij} - \frac{k_i^{out}k_j^{in}}{L})\delta(C_i, C_j)$$

Where:

$\boldsymbol{L}$ is the **number of edges** in the network.

$\boldsymbol{a_{ij}}$ is an **element of the adjacency matrix** (often it is binary).

$\boldsymbol{K_i}$ is the **degree** of node i .

$\boldsymbol{C_i}$ is the **community** to which node i belongs.

$\boldsymbol{\delta(C_i, C_j)}$ is a function which returns **1 if $Ci = Cj$, 0 otherwise.** It selects just the cases in which the nodes belong to the same class.

Thus, modularity focuses on **intra-class** connections.

The term $\frac{k_i k_j}{L}$ is the **probability of having a connection** between node i and node j, by chance, in the network. This means that if it is high is because they have many connections in general, not just because they belong to the same community.

Let's now analyze the term $(a_{ij} - \frac{k_i k_j}{L})$, we can end up with four cases:

- a_{ij} is **1**, so there is a connection between the two nodes
  - $\frac{k_i k_j}{L}$ is a **high value**. So, the probability of having a random connection is very high and for this reason the fact that they are linked is not so meaningful.
  - $\frac{k_i k_j}{L}$ is a **low value**. So, the probability of having a random connection is very low and for this reason the fact that they are linked is meaningful.

- a_{ij} is **0**, so there is not a connection between the two nodes
  - $\frac{k_i k_j}{L}$ is a **high value**. So, the term assumes a negative result, Meaning that those two nodes could be very easily connected by chance, but they are not. So, they really do not belong to the same community.
  - $\frac{k_i k_j}{L}$ is a **low value**. So, they are not connected and it is not so meaningful since the probability was low.

Now, let's analyze the possible results for Q:

- **Positive**, the two communities act as a community
- **Negative**, the communities are not communities
- **Very close to zero**, random division

Usually, we look for a positive modularity to say that the network can be divided into communities, meaning that these communities will behave as a group, not being just a group of randomly chosen nodes.

High modularity means that the chosen communities have a meaning has a group. The more the modularity, the more the network can be divided in **meaningful** subnetworks.

Divisibility: Divisibility *D* is a measure of the **segregation** between two communities

$$D = \frac{L}{\sum_{i,j=1}^{N} a_{ij}\left[1-\delta\left(C_i, C_j\right)\right] + k} = \frac{L}{L + \sum_{i,j=1}^{N} a_{ij}\left[1-\delta\left(C_i, C_j\right)\right]}$$

K is a constant (to avoid divergence), usually it is equal to *L* so that $D \in [{}^{1}/_{2}, 1]$

It is again about on testing the division of the networks in the two subnetworks but focusing in particular on the **inter-class** connections (connections linking nodes belonging to different classes).

Since $\delta\left(C_i, C_j\right)$ is equal to 1 when the two nodes belong to the same class and 0 otherwise, multiplying by $(1-\delta(C_i, C_j))$ has the effect of selecting only the connections between nodes belonging to different classes. So, in the denominator we are just **counting** how many connections between different-class nodes there are. Then it is normalized (and reversed) by *L* to express the concept that the number of connections between nodes of different classes is **inversely proportional** to the divisibility.

The *L* at denominator is useful first to avoid divergence, in case there are no connections between the communities, then it is done to set the divisibility in the range [1/2, 1] making much easier its interpretation.

When the divisibility is equal to **1**, the two classes are completely non connected.