Neural Information Processing Systems

Membranes and Synapses

Lecture 2

Drift and Diffusion

Diffusion current: the movement of ions caused by variation in the carrier concentration. Direction of the diffusion current depends on the slope of the carrier concentration.

Drift current: the movement of ions caused by electric fields. Direction of the drift current is always in the direction of the electric field.

The electrical mobility μ is the ability of charged particles to move through a medium in response to an electric field that is pulling them. The drift velocity is given by $v_d = \mu E$, where E is the applied field.

Diffusion is described by Fick's Law:

$$J_{\text{diff}} = -D\frac{\partial[C]}{\partial x}$$

J_{diff} is the diffusion flux $(\text{molecule}/s \cdot cm^2)$
D is the diffusion coefficient (cm^2/s)
$[C]$ is the ion concentration $(\text{molecule}/cm^3)$

Drift is described by Ohm's Law of Drift:

$$J_{\text{drift}} = -\mu z[C]\frac{\partial V}{\partial x}$$

J_{drift} is the drift flux $(\text{molecule}/s \cdot cm^2)$
V is the electric potential resulting from charge diffusion (V)
μ is the mobility (cm^2/Vs)
z is the ion valency
$[C]$ is the ion concentration $(\text{molecule}/cm^3)$

The Nernst Equation

Einstein's relation between diffusion and mobility describes the connection between J_{diff} and J_{drift} in equilibrium when the net flux is equal to zero. In this case we have:

$$J_{\text{drift}} + J_{\text{diff}} = 0$$
$$-\mu z[C]\frac{\partial V}{\partial x} - D\frac{\partial[C]}{\partial x} = 0$$

We can rewrite this using the chain rule:

$$\frac{\partial[C]}{\partial x} = \frac{\partial[C]}{\partial V}\frac{\partial V}{\partial x}$$

Hence we have:

$$-\mu z[C]\frac{\partial V}{\partial x} - D\frac{\partial [C]}{\partial V}\frac{\partial V}{\partial x} = 0$$
$$\left(-\mu z[C] - D\frac{\partial [C]}{\partial V}\right)\frac{\partial V}{\partial x} = 0$$

This holds at every location x, so the term in the parentheses must be zero:

$$D\frac{\partial [C]}{\partial V(x)} = -\mu z[C]$$
$$D = -\frac{\mu[C]}{q\frac{\partial [C]}{\partial V(x)}}$$

Using the Maxwell-Boltzmann distribution, the concentration is proportional to the exponential:

$$[C] \propto \exp\left[-\frac{zqV(x)}{k_B T}\right]$$

This allows us to compute the derivative:

$$\frac{\partial [C]}{\partial V(x)} = -\frac{zq}{k_B T}[C]$$

Now the diffusion coefficient is given by:

$$D = -\frac{\mu[C]}{-\frac{zq}{k_B T}[C]}$$
$$D = \frac{z\mu k_B T}{q}$$

Substituting this back into the expression for the total ion flux yields:

$$J_{net} = J_{\text{drift}} + J_{\text{diff}}$$
$$= -\mu z[C]\frac{\partial V}{\partial x} - D\frac{\partial [C]}{\partial x}$$
$$J_{net} = -\left(\mu z[C]\frac{\partial V}{\partial x} + \frac{\mu k_B T}{q}\frac{\partial [C]}{\partial x}\right)$$

This is called the Nernst-Planck Equation. In molar form it becomes:

$$\frac{J}{N_A} = -\left(\frac{\mu z[C]}{N_A}\frac{\partial V}{\partial x} + \frac{\mu k_B T}{qN_A}\frac{\partial [C]}{\partial x}\right)$$
$$\frac{J}{N_A} = -\left(uz[C]\frac{\partial V}{\partial x} + u\frac{RT}{F}\frac{\partial [C]}{\partial x}\right)$$

u is the molar mobility $(cm^2/V \cdot sec \cdot mol)$
R is the ideal gas constant $(1.98\, cal/K \cdot mol)$
F is Faraday's constant $(96{,}480\ C/mol)$

Multiplying this equation by zF yields the current density:

$$I_{net} = -\left(uz^2[C]F\frac{\partial V}{\partial x} + uzRT\frac{\partial [C]}{\partial x}\right)$$

The Nernst equation can be derived from the current density form of the Nernst-Plank Equation when the net current density over the membrane is equal to zero.

$$\begin{aligned}
I_{net} &= 0 \\
&= \left(uz^2[C]F\frac{\partial V}{\partial x} + uzRT\frac{\partial [C]}{\partial x}\right) \\
&= zF[C]\frac{\partial V}{\partial x} + RT\frac{\partial [C]}{\partial x} \\
\frac{\partial V}{\partial x} &= -\frac{RT}{zF}\frac{1}{[C]}\frac{\partial [C]}{\partial x} \\
V(x) &= -\frac{RT}{zF}\int_{x_1}^{x_2}\frac{1}{[C]}\frac{\partial [C]}{\partial x}dx \\
V(x) &= -\frac{RT}{zF}\ln\left(\frac{[C_{x_2}]}{[C_{x_1}]}\right) \\
E_i &= \frac{RT}{z_iF}\ln\frac{[C]_{out}}{[C]_{in}}
\end{aligned}$$

This equation gives the potential energy for each charge across the membrane due to forces of drift and diffusion. There is a unique equilibrium potential (also called reversal potential) for each ion species i. If the membrane potential is equal to this equilibrium potential, there will be no *net* flow of that ion across the membrane.

If a specific ion is at its equilibrium potential, it means the that inside and outside concentrations are such that:

$$E_i = E_m$$

Under typical conditions in a neuron, Na^+ and K^+ ions are <u>not</u> at their equilibrium potentials.

Reversal Potentials

Various active pumps and exchangers plus leakage channels -> ionic concentrations inside and outside the cell -> reversal potential for each ion -> equilibrium membrane potential.

The resting potential for the cell membrane as a whole is given as a weighted average of the concentrations and conductances of the individual ions, as expressed by the Goldman equation:

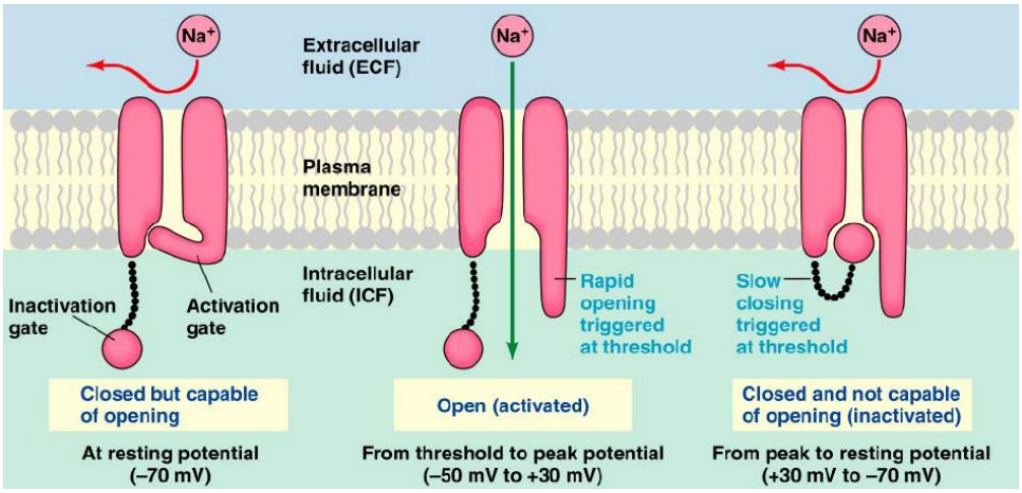
$$E_m = \frac{RT}{F}\ln\left(\frac{p_K[K^+]_{out} + p_{Na}[Na^+]_{out} + p_{Cl}[Cl^-]_{out}}{p_K[K^+]_{in} + p_{Na}[Na^+]_{in} + p_{Cl}[Cl^-]_{out}}\right)$$

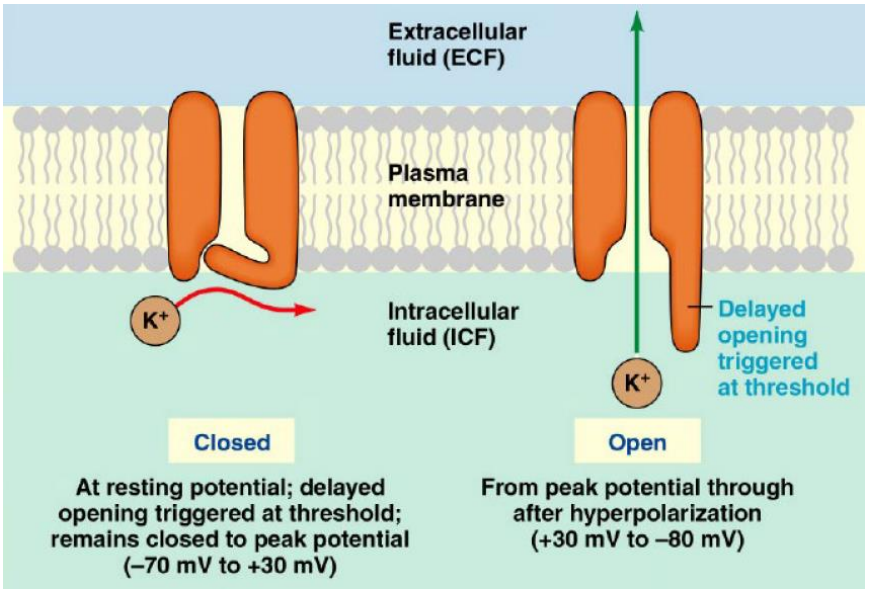
Where p_i is the permeability of ion i. Note that during membrane depolarisation or hyperpolarisation, the ionic concentration doesn't change very much. Mostly the change in membrane potential is brought about by changes in the membrane permeability of a specific ion. For example, when the sodium channels open, p_{Na} increases dramatically, thereby pushing the membrane potential closer to the equilibrium potential for sodium.

Lecture 3

Ion Channels

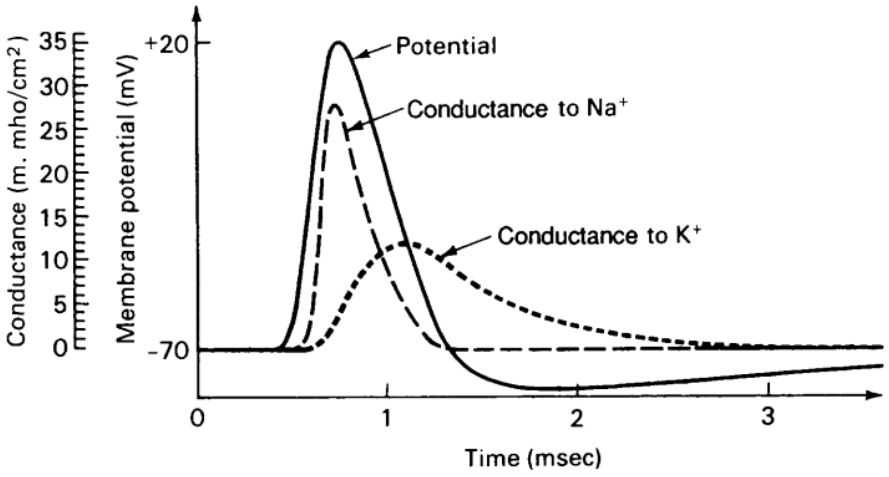
Membrane ion channels are like gates, which selectively permit or block the passage of particular ions across the plasma membrane. Changes in conformation of the proteins open and close the channel. The three major types of channels are: voltage gated, chemically gated, and mechanically gated. Channels can be either activated, deactivated, or inactivated (unable to open).

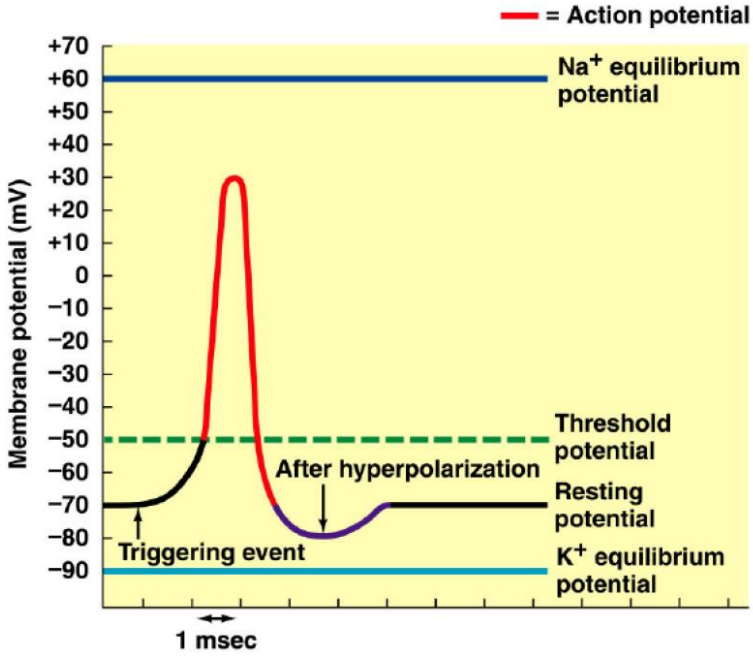




Action Potential Initiation

Action potentials are brief, all-or-nothing reversals of the membrane potential, brought about by rapid and transient changes in membrane ion permeability. Action potentials are generated by depolarisation on the surface of the membrane causing voltage-gated ion channels to open. This leads to g_{Na} increasing, pushing the cell membrane closer to the equilibrium for sodium. However, as the cell continues to depolarise, voltage-gated potassium channels open, thereby increasing g_K and hence pushing the cell membrane back towards the equilibrium potential for potassium.





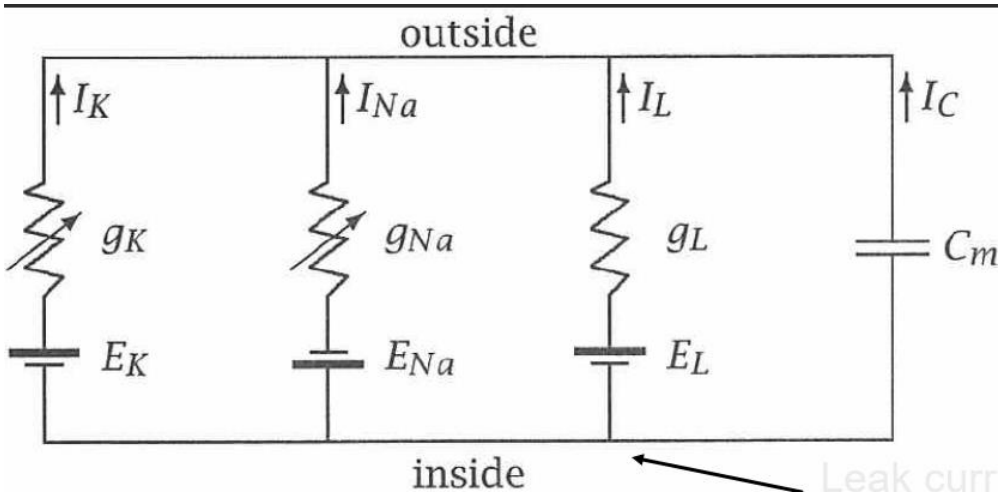
Action Potential Propagation

The initial action potential is usually confined to a localised area on the cell membrane surface. This depolarised area of the cell creates a local current sink, which tends to depolarise the regions around it. This sets up a wave of depolarisation that spreads throughout the entire surface of the membrane. After the depolarisation wave passes by, the sodium channels remain inactivated for a period of time, resulting in an absolute refractory period. During this period, the cell cannot depolarise again, hence preventing the action potential from travelling backwards.

Cell type	Range of T_d (ms)	Speed (m/s)
Nerve	0.5 – 1.0	20 – 140
Skeletal muscle	2.0 – 5.0	3.0 – 5.0
Cardiac muscle	150 – 300	0.03 – 0.4

Hodgkin and Huxley Model

Hodgkin and Huxley modelled the neuron current based on the following circuit:



This yields the equation:

$$I_m = C_m \frac{dV}{dt} + I_K + I_{Na} + I_L$$

Written in terms of conductances and voltages this becomes:

$$I_m(t) = C_m \frac{dV(t)}{dt} + g_K(V,t)(V - E_K) + g_{Na}(V,t)(V - I_{Na}) + g_L(V - E_L)$$

Note that here it has been assumed that the leakage conductance is constant with respect to time and voltage, and the capacitance is also assumed to be constant. The capacitance is determined by the phospholipid bilayer, and is not affected by the ion channels.

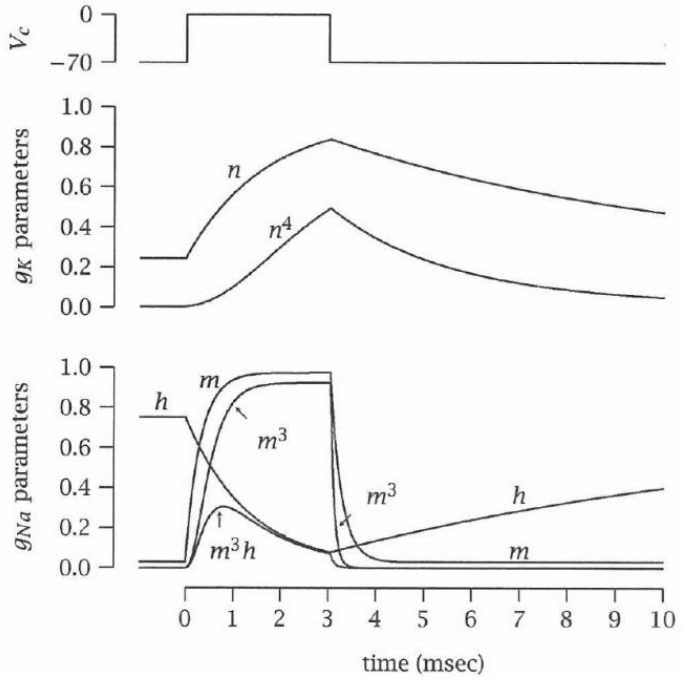
Based on a series of voltage clamping experiments made with giant squid axons, Hodgkin and Huxley parameterised the conductances as follows:

$$g_K(V,t) = G_K n^4$$
$$g_{Na}(V,t) = G_{Na} m^3 h$$

Where G_k and G_{Na} are experimentally determined maximum conductances, and n, m, and h are functions that describe the activation of K channels, activation of Na channels, and inactivation of Na channels respectively. They have the forms:

$$\frac{dm}{dt} = \alpha_n(1-n) - \beta_n n$$
$$\frac{dm}{dt} = \alpha_m(1-m) - \beta_m m$$
$$\frac{dh}{dt} = \alpha_h(1-h) - \beta_h h$$

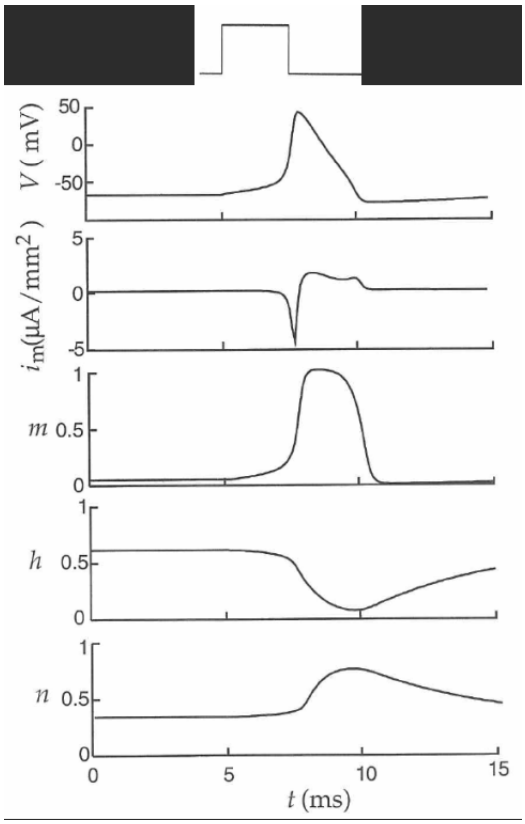
Solving these differential equations yields the following plots:



Substituting in these forms we have the Hodgkin-Huxley equations:

$$I_m(t) = C_m\frac{dV(t)}{dt} + G_K n^4(V - E_K) + G_{Na}m^3h(V - I_{Na}) + g_L(V - E_L)$$

We can observe the solutions on the following plots:



Lecture 4

Synaptic transmission

Neurons transfer signals across the synaptic cleft by releasing chemical messengers called neurotransmitters. These bind to postsynaptic receptors, causing ion channels to open and hence generating a postsynaptic potential, which can be either excitatory or inhibitory.

Note that the words presynaptic and postsynaptic only refer to a single synapse. Most neurons are presynaptic to one group of neurons and postsynaptic to another group. Neurons in the CNS typically receive 10,000-100,000 synaptic inputs.

Neurotransmitters

Neurotransmitters in the presynaptic neuron are stored in the synaptic knob in vesicles. A change in potential caused by an action potential triggers the opening of voltage gated Ca2+ channels in the synaptic knob, causing it to flow into the cell. Different neurotransmitters cause different effects on the postsynaptic neuron:

- Glutamate is always excitatory.
- Glycine is always inhibitory.
- Norepinephrine can be either excitatory or inhibitory.

The depolarisation or hyperpolarisation produced by release of neurotransmitters by a single presynaptic neuron is called a postsynaptic potential (PSP). These can either by excitatory depolarisation (EPSPs) or inhibitory hyperpolarisation (IPSPs).

Neuromodulators are chemical messengers that bind to neuronal receptors but do not generate PSPs. Neuromodulators bring about long-term changes that modulate (depress or enhance) synaptic effectiveness. They may act either pre-synaptically or post-synaptically.

Summation of PSPs

A single PSP is usually not enough to generate an action potential. Instead, usually multiple PSPs must be combined. This can occur due to either spatial summation, where PSPs over different presynaptic inputs that are spatially separated are combined, or temporal summation, where the PSPs due to one or more presynaptic inputs arriving in a short period of time are added together.

Networks of neurons

Neurons are linked to each other through enormous networks involving convergence and divergence of the neural connectivity. Neural connectivity changes mostly during early development, while synaptic efficiency changes regularly during life. Synaptic efficiency refers to how strong a given synaptic connection is, and is determined by the amount of neurotransmitter released by the presynaptic neuron, and the magnitude of the resulting response in the postsynaptic neuron.

Neural Coding

Lecture 5

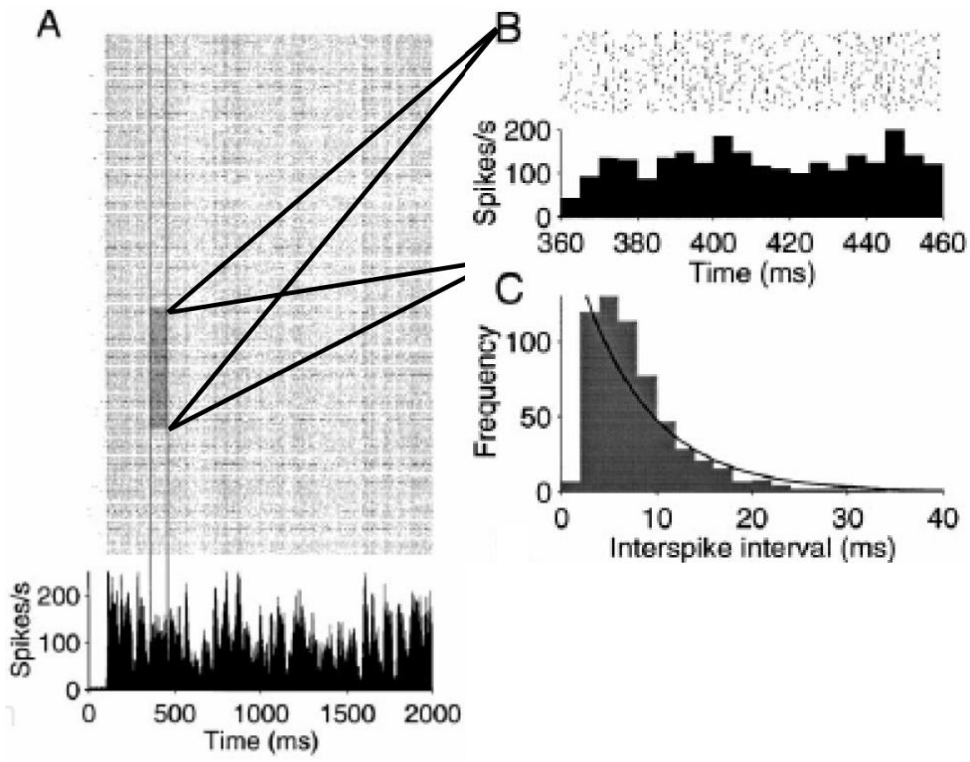
Types of Neural Coding

Neurons carry information in the timing of their action potentials. Because all action potentials are essentially identical, we can describe the full spike train as a sequence of firing times:

$$F = \{t_1, t_2, \dots, t_n\}$$

We can write this in terms of the sum of idealised infinitesimally narrow spikes represented by delta functions:

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



There are many different forms of neural coding found in the brain:

- Rate code: The average rate of firing is important (the timing of individual APs is stochastic).
- Population code: Information is carried by the instantaneous pattern of activity of a population of neurons.
- Place code: Information contained in the set of neurons that are active, where different neurons respond to different subsets of the receptive field.
- Temporal code: The timing of the individual APs carries information (e.g. in phase locking).
- Spatiotemporal code: Includes both place and temporal aspects.

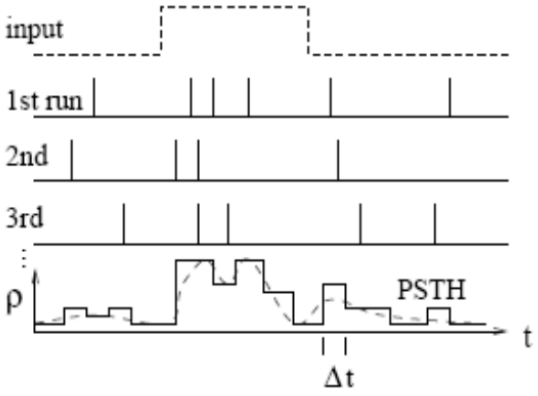
Neuron Firing Rates

There is no single method of defining the firing rate of a neuron. The simplest method is to simply count the number of action potentials in time T and divide by that interval:

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

Alternatively, one can average over several repetitions of the same stimuli for the same neuron:

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



Spike density in PSTH

$$\rho = \frac{1}{\Delta t}\frac{1}{K} n_k\left(t; t+\Delta t\right)$$

Peri-stimulus time histogram

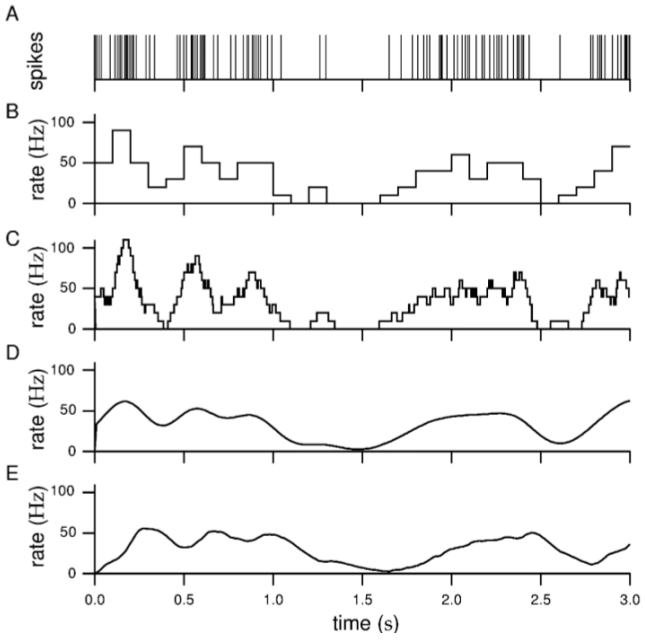
Count spikes in time bins over all repetitions of the experiment.

Measuring the firing rate over time is difficult, since there are only a finite number of action potentials, so there is insufficient information to precisely define the rate as a continuous function over time. The simplest approach is to simply count the number of action potentials in a set of pre-defined and positioned bins. This, however, leads to a lumpy estimate which is dependent on the positioning of the bins. To avoid this, a window function can be integrated with the spike train, effectively allowing bins to be applied to each time interval:

$$r_{est}(t) = \int_0^\infty w(\tau) S(\tau) d\tau$$

Potential window functions include rectangular, Gaussian, or exponential:

$$w(t) = \begin{cases} \frac{1}{\Delta t} \text{ for } -\Delta t/2 \le t \le \Delta t/2 \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right) \\ \alpha^2 t \exp(-\alpha t) \end{cases}$$



A. Spike train from a neuron in the inferotemporal cortex of a monkey recorded while the animal watched a video.

B. Discrete-time spiking rate obtained by binning time and counting spike with Δt = 100 ms.

C. Approximate spiking rate obtained by sliding a rectangular window function along spike train, Δt = 100 ms.

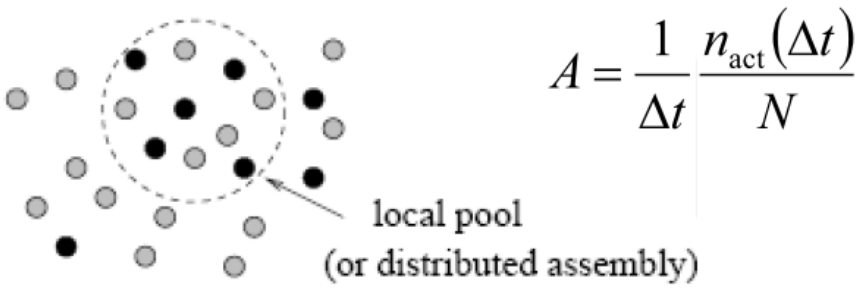
D. Approximate spiking rate obtained by sliding a Gaussian window function along spike train, σ = 100 ms.

E. Approximate spiking rate obtained by sliding a window function along spike train of the form $w(\tau) = [\alpha^2 \tau \exp(-\alpha\tau)]_+$ with $1/\alpha$ = 100 ms.

Population Coding

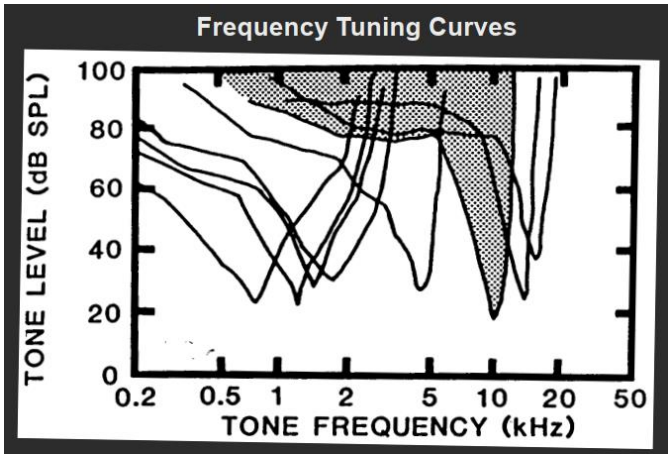
Population codes are defined in terms of activity, which refers to the firing rate of a pool of neurons over a single trial.

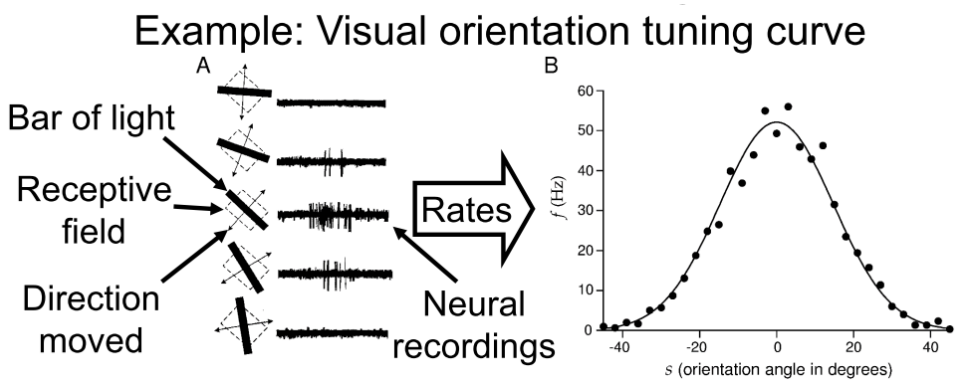
$$A = \frac{1}{\Delta t}\frac{n(\Delta t)}{N}$$



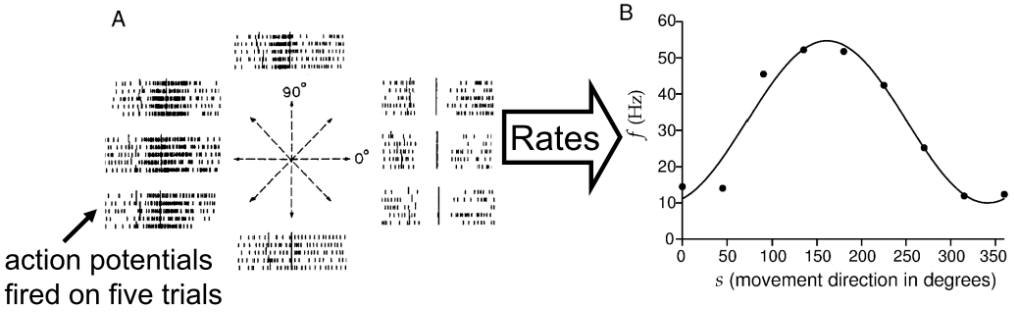
Place Coding

Place coding involves information being encoded by which specific neurons in a particular population are firing, given that different neurons respond to different stimuli in accordance with their tuning curves. For example, typical frequency tuning curves in the normal mammalian cochlea.





Example: Motor orientation tuning curve



Temporal Coding

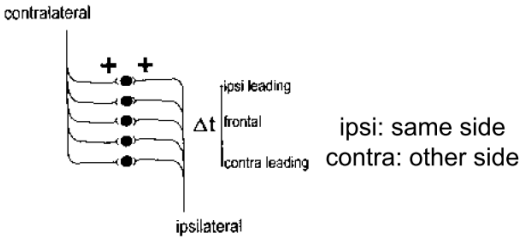
Temporal coding involves information being encoded in the specific timing of the arrival of spikes, or at least high-frequency variation in the rate code. This provides much more precise information about the temporal variation of the stimulus. This can also be called correlation coding, since it entails that information is carried not just by the rate of firing, but by correlations between firing times. For example, information could be carried by spike-time intervals.

An example of where temporal coding is used is in sound localisation, where cells only fire if they detect coincident activity from each ear.

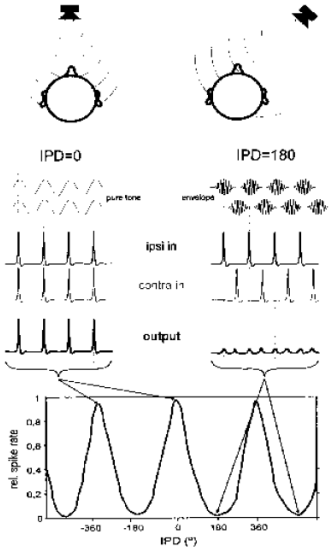
Example: ***Sound localisation*** is a cognitive function in which temporal information is used.

Involves comparison of the time of arrival of spikes from the two ears.

Neural circuitry in auditory brainstem:



ipsi: same side
contra: other side



The measure of synchronization r between two spike trains can be defined in terms of how much the N spikes arranged into M bins, with h_m spikes per bin, resemble sine or cosine functions:

$$S_S = \frac{1}{N}\sum_{m=1}^{M} h_m \sin\left(\frac{2\pi m}{M}\right)$$

$$S_C = \frac{1}{N}\sum_{m=1}^{M} h_m \cos\left(\frac{2\pi m}{M}\right)$$

$$r = \sqrt{S_s^2 + S_C^2}$$

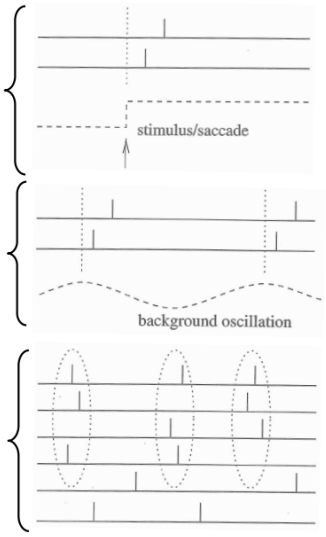
Other mechanisms of temporal including are summarised below.

Other temporal codes

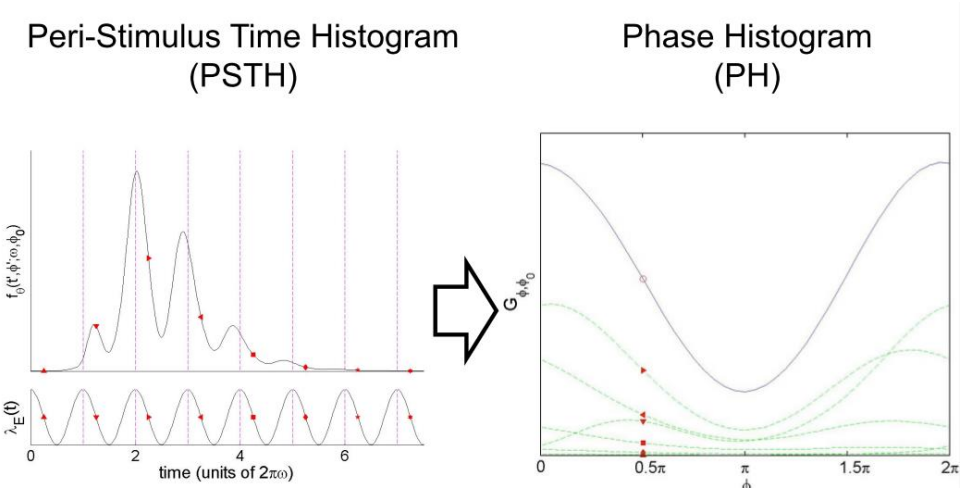
Time-to-first spike: one neuron responds faster to a change in a stimulus than another.

Phase: two neurons fire at different phases with respect to a periodic stimulus.

Synchrony: groups of neurons are nearly synchronous while other are not.



The peristimulus time histogram or poststimulus time histogram, both abbreviated PSTH, are histograms of the times at which neurons fire relative to the time of the stimulus.

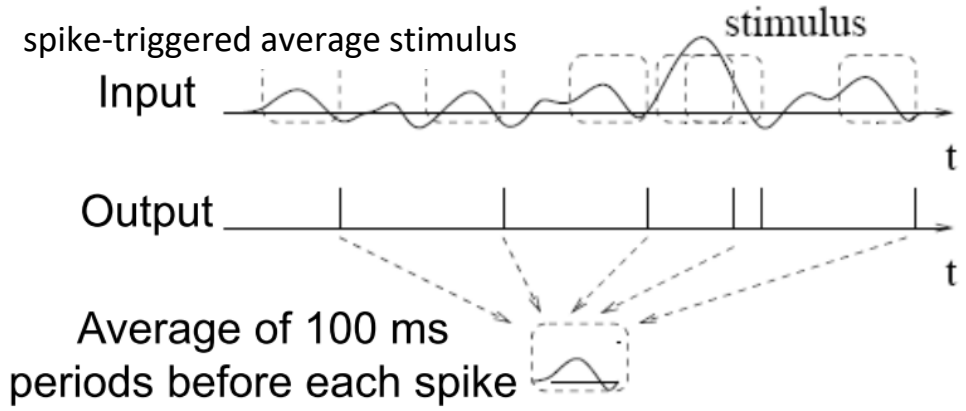


Many neurons respond to place (spatial) and temporal input. This can be quantified using a technique called reverse correlation, in which we determine what stimulus is responsible for neural firing. This is quantified using the spike-triggered average, which is the average input of the stimulus $s(t)$ immediately preceding each action potential at time t_i.

$$C(t) = \frac{1}{n}\sum_{(i=1)}^{n} s(t_i - t)$$

This can be rewritten as an integral over the neural response function:

$$C(t) = \frac{1}{n}\int_0^T S(t)s(t-\tau)\,d\tau$$



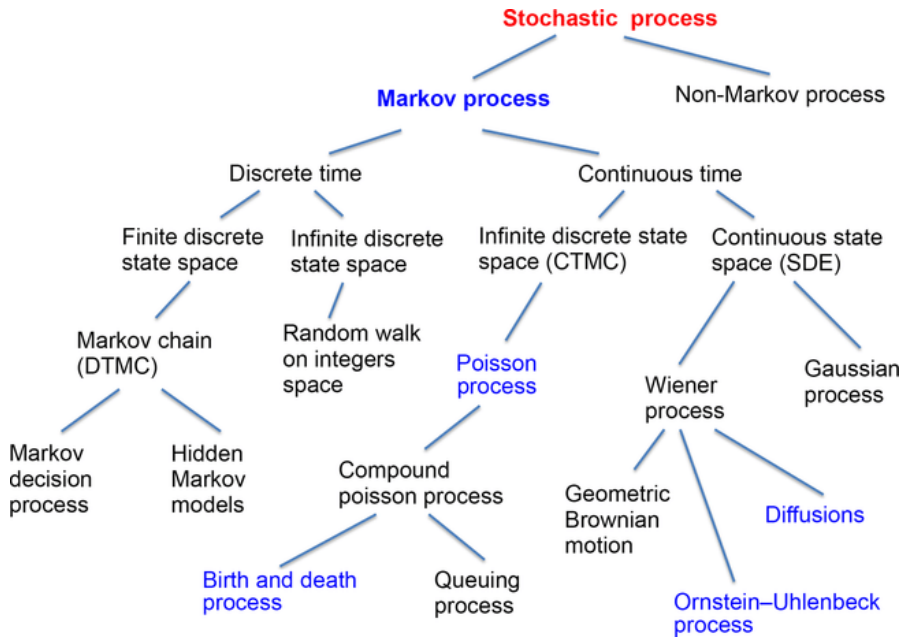
Lecture 6

Stochastic Processes

We can model the spike times as a stochastic process, with the assumptions we make affecting the properties of the resulting spike train.

- A **point process** is a stochastic process that generates a sequence of events, such as action potentials. The probability of a spike at any given time could depend on the entire history of preceding spikes.
- A **renewal process** is a point process where the probability of an event occurring at a specific time depends only on the immediately preceding spike (intervals between spikes are independent).
- If there is no dependence at all on preceding spikes (the spikes are statistically independent), the point process is called a **Poisson process**.

The simplest type of point process is a **homogenous Poisson process**, where the firing rate is constant over time.



Homogenous Poisson Process

The probability of a specific sequence of spikes occurring at arbitrary times is denoted $P[t_1, t_2, \dots, t_n]$. Under a homogenous Poisson process, the probability of such a sequence is given by:

$$P[t_1, t_2, \dots, t_n] = n!\, P_T[n] \left(\frac{\Delta t}{T}\right)^n$$

Where $P_T[n]$ is the probability of a sequence of n spikes occurring with time T. The factor of $n!$ Adjusts for the number of combinations of n spikes that is possible, and the final term is for normalisation.

To compute $P_T[n]$, we divide the time T into M bins of size $\Delta t = T/M$ and take the limit $\Delta t \to 0$, at which limit there is no possibility of two spikes occurring in the same bin. The probability of n spikes appearing one each in n bins is $(r\Delta t)^n$, while that of the remaining $M - n$ bins each having no spikes is $(1 - r\Delta t)^{M-n}$. The number of ways of putting n spikes in M bins is given by the binominal coefficient, yielding the expression:

$$P_T[n] = \lim_{\Delta t \to 0} \frac{M!}{(M-n)!\, n!} (r\Delta t)^n (1 - r\Delta t)^{M-n}$$

In taking the limit we note that since $M \to \infty$ whle n is fixed (by assumption), we have $M - n \approx M = T/\Delta t$, and $\frac{M!}{(M-n)!} \approx M^n = \left(\frac{T}{\Delta t}\right)^n$. Thus we find:

$$\begin{aligned}
P_T[n] &= \lim_{\Delta t \to 0} \frac{M^n}{n!} (r\Delta t)^n \left((1 - r\Delta t)^{\frac{-r\Delta t}{-r\Delta t}}\right)^{\frac{T}{\Delta t}} \\
P_T[n] &= \frac{1}{n!} \lim_{\Delta t \to 0} \left(\frac{T}{\Delta t}\right)^n (r\Delta t)^n \left((1 - r\Delta t)^{\frac{1}{-r\Delta t}}\right)^{-rT} \\
P_T[n] &= \frac{1}{n!} (rT)^n \lim_{\Delta t \to 0} \left((1 - r\Delta t)^{\frac{1}{-r\Delta t}}\right)^{-rT} \\
P_T[n] &= \frac{1}{n!} (rT)^n \exp(-rT)
\end{aligned}$$

Substituting this into the formula for the probability of a generic sequence we have:

$$P[t_1, t_2, \dots, t_n] = r^n (\Delta t)^n \exp(-rT)$$

This is a Poisson distribution.

Interspike Interval Distribution

For a Poisson distribution, the probability of an interspike interval falling between τ and $\tau + \Delta\tau$ is found by setting $n = 1$ and $T = \tau$ in the above formula:

$$P[\tau \le t_{i+1} - t_i < \tau + \Delta t] = r\Delta t \exp(-r\tau)$$

The density function for infinitesimal Δt becomes:

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

The interspike interval therefore has the properties:

$$\mu_\tau = \frac{1}{r}$$

$$\sigma_\tau^2 = \frac{1}{r^2}$$
$$C_V = \frac{\sigma_\tau}{\mu_\tau} = 1$$

If the coefficient of variation of the spike count is greater than 1, this means it is not a Poisson process. In such cases the variance grows faster than the mean as the trial duration increases, indicating the presence of long-term correlations in the data.

We can summarise this as follows. If for every t > 0 the number of arrivals in the time interval [0,t] follows the Poisson distribution with mean $r\Delta t$, then the sequence of inter-arrival times are independent and identically distributed exponential random variables having mean $1/r$.

Note that a Renewal Process is a generalisation of a Poisson process, in which the holding times do not need to be exponentially distributed.

The Poisson Spike Generator

A series of spikes can be generated from a known firing rate $r(t)$ using the Poisson Spike Generator method. Each period of time Δt, we simply generate a random number x uniformly distributed between 0 and 1. If $x < r(t)\Delta t$, then a spike is fired, otherwise no spike is fired. This will deliver a series of Poisson distributed spikes (with exponential interspike intervals) so long as Δt is small compared to $r(t)$.

An alternative, and generally faster way to do this which works for constant firing rates, is to randomly draw the interspike interval instead of the number of spikes per interval, as we can skip all the intervals without any spikes. This works by randomyl generating a random number x and setting $\Delta t = t_{i+1} - t_i$ using the equation to compute the next spike time t_{i+1}:

$$\begin{aligned} x &= \exp\big(-r(t_{i+1} - t_i)\big) \\ -r(t_{i+1} - t_i) &= \ln(x) \\ t_{i+1} - t_i &= -\frac{1}{r}\ln(x) \\ t_{i+1} &= t_i - \frac{1}{r}\ln(x) \end{aligned}$$

Lecture 7

Definition of Entropy

Entropy is a measure of variability, or the amount of information, in a set of responses. The unit of information is the bit, where 1 bit is the amount of information sufficient to choose between two equally likely alternatives.

The information content of a single response is equal to:

$$s(x_i) = -\log_2(P[x_i])$$

The entropy of an ensemble of outcomes, namely a random variable, is:

$$S(X) = -\sum_i P(X)\log_2\big(P(X)\big)$$

For K equally likely outcomes, the entropy becomes:

$$S(X) = -\sum_{k} \frac{1}{K} \log_2\left(\frac{1}{K}\right)$$
$$S(X) = \log_2(K)$$

The entropy thus turns out to be the number of digits required to write K as a binary number less one.

Mutual Information

Applied to the case of a stimulus s generating firing rate outputs r, we can define the response entropy of the firing rate output as:

$$S_r = -\sum_{r} P[r] \log_2(P[r])$$

The amount of entropy that is left over after conditioning on the stimulus is called the conditional entropy, and is given by:

$$S_{r|s} = -\sum_{r} P[r|s] \log_2(P[r|s])$$

The average conditional entropy is known as the noise entropy. It gives the amount of entropy in r that is not attributable to the stimuli. It is written as:

$$S_{noise} = \sum_{s} P[s]\, S_{r|s}$$
$$S_{noise} = -\sum_{s} P[s] \sum_{r} P[r|s] \log_2 P[r|s]$$

The mutual information entropy is the reduction in entropy of r that occurs when we learn the value of stimulus s. It thus represents how much the firing rate tells us about the stimulus. It defined as the full response entropy minus the noise entropy:

$$\begin{aligned} I_m &= S_r - S_{noise} \\ &= -\sum_{r} P[r] \log_2(P[r]) + \sum_{s} P[s] \sum_{r} P[r|s] \log_2 P[r|s] \\ &= -\sum_{r}\sum_{s} P[s]P[r|s] \log_2(P[r]) + \sum_{s}\sum_{r} P[s]P[r|s] \log_2 P[r|s] \\ &= \sum_{r}\sum_{s} P[s]P[r|s] \left(-\log_2(P[r]) + \log_2 P[r|s]\right) \\ I_m &= \sum_{r}\sum_{s} P[s]P[r|s] \log_2 \frac{P[r|s]}{P[r]} \end{aligned}$$

Entropy of Spike Trains

The information associated with a specific interspike interval of τ using time bins of size $\Delta\tau$ is:

$$-\log_2(p[\tau]\Delta\tau)$$

To calculate the entropy associated with the source neuron, we need to take the expectation over all possible lengths of interspike intervals, and then multiply by the number of intervals $N = \frac{T}{\Delta\tau} = rT$:

$$S = N \times E[-\log_2(p[\tau]\Delta\tau)]$$
$$S = -rT\int_0^\infty p[\tau]\log_2(p[\tau]\Delta\tau)\,d\tau$$

For a homogenous Poisson process, we have $p[\tau] = r\exp(-r\tau)$, making the integral:

$$S = -rT\int_0^\infty r\exp(-r\tau)\log_2(r\Delta\tau\exp(-r\tau))\,d\tau$$
$$S = -rT\int_0^\infty r\exp(-r\tau)\frac{\ln(r\Delta\tau\exp(-r\tau))}{\ln(2)}d\tau$$
$$S = -\frac{r^2T}{\ln(2)}\int_0^\infty \exp(-r\tau)\ln(r\Delta\tau\exp(-r\tau))\,d\tau$$

Let $u = r\tau, du = rd\tau$

$$\begin{aligned}
S &= -\frac{rT}{\ln(2)}\int_0^\infty \exp(-u)\ln(r\Delta\tau\exp(-u))\,du \\
&= -\frac{rT}{\ln(2)}\int_0^\infty \exp(-u)\,[\ln(r\Delta\tau) - u]\,du \\
&= -\frac{rT}{\ln(2)}\left(\ln(r\Delta\tau)\int_0^\infty \exp(-u)\,du - \int_0^\infty u\exp(-u)\,du\right) \\
&= -\frac{rT}{\ln(2)}\left(\ln(r\Delta\tau)\,[-\exp(-u)]_0^\infty - [-u\exp(-u) - \exp(-u)]_0^\infty\right) \\
&= -\frac{rT}{\ln(2)}\left(\ln(r\Delta\tau)\,[1] - [--1]\right) \\
&= -rT\left(\log_2(r\Delta\tau) - \frac{1}{\ln(2)}\right) \\
&= -rT(\log_2(r\Delta\tau) - \log_2(e)) \\
&= -rT\log_2\left(\frac{r\Delta\tau}{e}\right) \\
&= rT\log_2\left(\frac{e}{r\Delta\tau}\right) \\
&= rT(\log_2(e) - \log_2(r\Delta\tau)) \\
&= rT\log_2(e)\left(1 - \frac{\log_2(r\Delta\tau)}{\log_2(e)}\right) \\
S &= \frac{rT}{\ln(2)}(1 - \ln(r\Delta\tau))
\end{aligned}$$

Entropy is maximised at the rate:

$$\begin{aligned}
\frac{dS}{dr} &= \frac{T}{\ln(2)}(1 - \ln(r\Delta\tau)) + \frac{rT}{\ln(2)}\left(-\frac{\Delta\tau}{r\Delta\tau}\right) \\
\frac{T}{\ln(2)}(1 - \ln(r\Delta\tau)) &= \frac{rT}{\ln(2)}\left(\frac{\Delta\tau}{r\Delta\tau}\right) \\
(1 - \ln(r\Delta\tau)) &= 1 \\
\ln(r\Delta\tau) &= 0 \\
r\Delta\tau &= 1 \\
r &= \frac{1}{\Delta\tau}
\end{aligned}$$

Neuron Models

Summary of Neural Models

Model	Equations	Notes
McCulloh-Pitts	$S_i(t+\Delta t) = \begin{cases} 1 \text{ if } \sum_{j=1}^{N} w_{ij}x_j(t) \geq \theta_i \\ 0 \text{ if } \sum_{j=1}^{N} w_{ij}x_j(t) < \theta_i \end{cases}$	Binary output, deterministic
Hopfield Neurons	$S_i(t+\Delta t) = \begin{cases} +1 \text{ with prob } g\left(\sum_{j=1}^{N} w_{ij}x_j(t)\right) \\ -1 \text{ with prob } 1-g\left(\sum_{j=1}^{N} w_{ij}x_j(t)\right) \end{cases}$	Binary output, stochastic.
Firing-Rate Neurons	$S_i(t+\Delta t) = g\left(\sum_{j=1}^{N} w_{ij}x_j(t) - \theta\right)$	Rate-based, deterministic.
Leaky Integrator	$C\frac{dV_i}{dt} + \frac{V_i}{R} = I_i(t)$ $S_i(t) = g(V_i(t) - \theta_i)$	Rate-based, deterministic, subthreshold dynamics.
Leaky Integrate and Fire	$C\frac{dV_i}{dt} + \frac{V_i}{R} = I_i(t)$ $S_i(t) = \begin{cases} \delta(t) \text{ if } V_i(t) \geq V_{\text{th}} \\ 0 \quad \text{if } V_i(t) < V_{\text{th}} \end{cases}$	Spiking, deterministic, subthreshold dynamics.
Stochastic Leaky Integrate and Fire	$C\frac{dV}{dt} + \frac{V}{R} = a_e\frac{dN_e(t)}{dt} - a_i\frac{dN_i(t)}{dt}$ $S(t) = \begin{cases} \delta(t) \text{ if } V(t) \geq V_{\text{th}} \\ 0 \quad \text{if } V(t) < V_{\text{th}} \end{cases}$	Spiking, stochastic, subthreshold dynamics.
Conductance-based Neurons	$I(t) = g(t-t_{in})\big(V_E - V(t)\big)$ $C\frac{dV(t)}{dt} + \frac{V(t)}{R} = g\left(V_{syn} - V(t)\right)$ $S(t) = \begin{cases} \delta(t) \text{ if } V(t) \geq V_{\text{th}} \\ 0 \quad \text{if } V(t) < V_{\text{th}} \end{cases}$	Spiking, deterministic, subthreshold dynamics, conductance-based.
Hodgkin-Huxley neuron	$I_m = C_m\frac{dV}{dt} + g_K(V - V_K) + g_{Na}(V - V_{Na}) + g_L(V - V_L)$ $g_K(V,t) = G_K n^4$ $g_{Na}(V,t) = G_{Na}m^3h$ $\frac{dn}{dt} = \alpha_n(1-n) - \beta_n n$ $\frac{dm}{dt} = \alpha_m(1-m) - \beta_m m$ $\frac{dh}{dt} = \alpha_h(1-h) - \beta_h h$	Spiking, deterministic, subthreshold dynamics, conductance-based.
FitzHugh-Nagumo	$\frac{dV}{dt} = V - \frac{V^3}{3} - W + I$ $\frac{dW}{dt} = \phi(V + a - bW)$	Spiking, deterministic, subthreshold dynamics, conductance-based.

Lecture 8

McCulloch-Pitts Neurons

The simplest neuron model, in which the output of the point neuron is either 0 or 1, and the activity is computed incrementally at discrete moments of time in terms of the $j = 1, \dots, N$ inputs neurons, each with activity x_j. The output S_i at the next time step is given in terms of the output threshold θ_i:

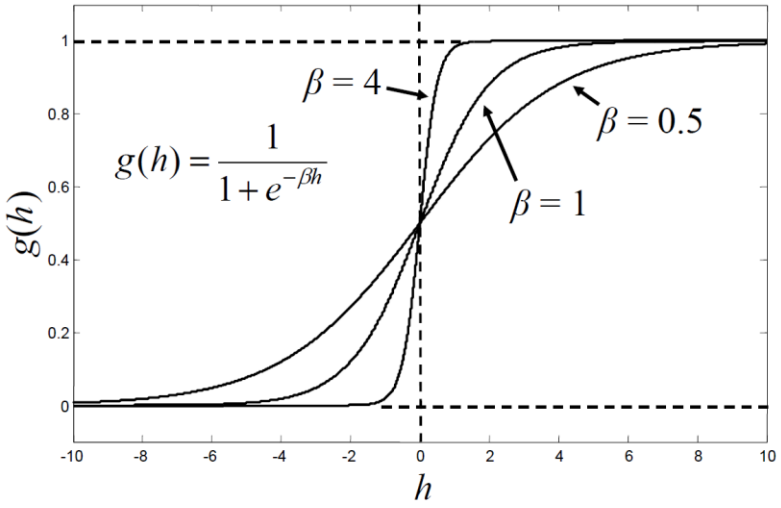
$$S_i(t + \Delta t) = \begin{cases} 1 \text{ if } \sum_{j=1}^{N} w_{ij} x_j(t) \geq \theta_i \\ 0 \text{ if } \sum_{j=1}^{N} w_{ij} x_j(t) < \theta_i \end{cases}$$

Hopfield Neurons

This adds stochastic behaviour to the McCulloch-Pitts neuron.

$$S_i(t + \Delta t) = \begin{cases} +1 \text{ with prob } g\left(\sum_{j=1}^{N} w_{ij} x_j(t)\right) \\ -1 \text{ with prob } 1 - g\left(\sum_{j=1}^{N} w_{ij} x_j(t)\right) \end{cases}$$

Where $g(h) = \frac{1}{1+\exp(-\beta h)}$, with β being the inverse temperature, a measure of noise. In the limit where $\beta \to \infty$, the rule becomes deterministic and the Hopfield neuron reduces to the McCulloch-Pitts neuron.



Firing-Rate Neurons

In these models the neural output is not binary, but is described as a firing rate:

$$S_i(t + \Delta t) = g\left(\sum_{j=1}^{N} w_{ij} x_j(t) - \theta\right)$$

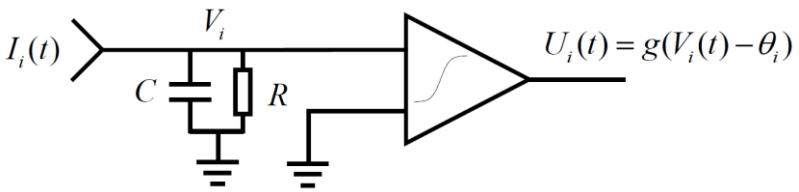
The activation function g can take various forms, such as a sigmoid or threshold-linear function.

Leaky Integrator Neurons

Continuous time rate-based models are sometimes called 'leaky integrators', as they simulate a membrane potential which gradually adds up input signals over time. The evolution of synaptic voltage in such models is given by the equation:

$$C\frac{dV_i}{dt} + \frac{V_i}{R} = I_i(t)$$

This is essentially an RC-circuit, with time constant $\tau = RC$, which acts like a low-pass filter.



The output in terms of firing rate is written as a function of the voltage and the threshold:

$$S_i(t) = g(V_i(t) - \theta_i)$$

In response to a step current I switched on at $t = 0$, the voltage is given by:

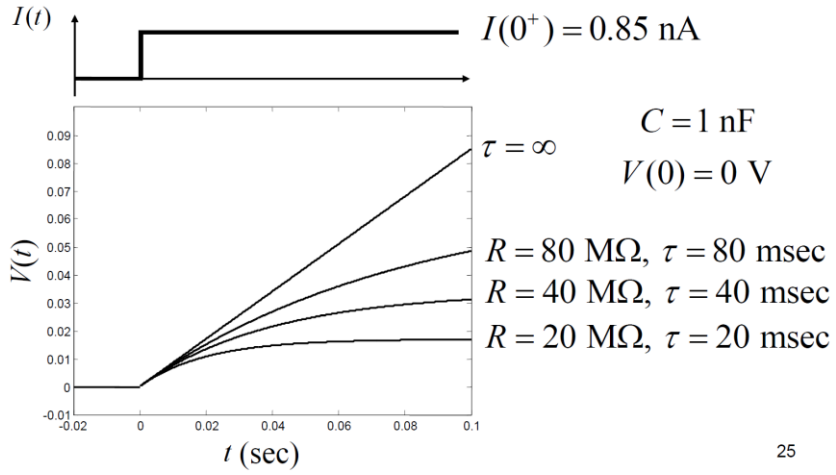
$$C\frac{dV_i}{dt} + \frac{V_i}{R} = I$$
$$V_i' + \frac{1}{RC}V_i = \frac{I}{C}$$

Solve using the integrating factor $e^{\int \frac{1}{RC}dt}$:

$$V_i' e^{\int \frac{1}{RC}dt} + \frac{1}{RC}V_i e^{\int \frac{1}{RC}dt} = \frac{I}{C}e^{\int \frac{1}{RC}dt}$$
$$\frac{d}{dt}\left[V_i e^{\left(\frac{t}{RC}\right)}\right] = \frac{I}{C}e^{\left(\frac{t}{RC}\right)}$$
$$\int_0^{t'} \frac{d}{dt}\left[V_i e^{\left(\frac{t}{RC}\right)}\right] dt = \frac{I}{C}\int_0^{t'} e^{\left(\frac{t}{RC}\right)} dt$$
$$\left[V_i e^{\left(\frac{t}{RC}\right)}\right]_0^{t'} = IR\left[e^{\left(\frac{t}{RC}\right)}\right]_0^{t'}$$
$$V_i e^{\left(\frac{t}{RC}\right)} - V_i(t=0) = IR\left[e^{\left(\frac{t}{RC}\right)} - 1\right]$$
$$V_i e^{\left(\frac{t}{RC}\right)} = IR\left[e^{\left(\frac{t}{RC}\right)} - 1\right] + V_i^0$$
$$V_i(t) = IR\left[1 - e^{(-t/RC)}\right] + V_i^0 e^{(-t/RC)}$$

For very large values of RC and $V_i(t=0) = 0$:

$$V_i(t) \approx IR\left[1 - \left(1 + \left(\frac{-t}{RC}\right)\right)\right]$$
$$V_i(t) = \frac{I}{C}t$$



25

Given the firing rate, a series of output spikes can be generated by modelling the neuron as a Poisson neuron, in which output spikes are generated randomly as an inhomogeneous Poisson process governed by the time-varying output rate $S_i(t) = g(V_i(t) - \theta_i)$, which depends on the voltage.

Continuous output rate-based neurons can be combined to form neural networks. In this case the synaptic input is added to the input current:

$$\boxed{C\frac{dV_i}{dt} + \frac{V_i}{R} = I_i(t) + \sum_{j=1}^{N} w_{ij}\, S_j(t)}$$

$$\boxed{S_i(t) = g(V_i(t) - \theta_i)}$$

Delta function current

The simplest model of synaptic inputs is instantaneous current injection:

$$I_{syn}(t) = \sum_{j=1}^{N} w_{ij} \sum_{k} \delta\left(t - t_{jk}\right)$$

Where t_{jk} is the time when the jth input neuron fires its kth spike.

For simplicity, we can consider a single spike at one synapse of the form:

$$I_{syn}(t) = Ca\delta(t - t_{in})$$

In this case the output becomes:

$$C\frac{dV_i}{dt} + \frac{V_i}{R} = Ca\delta(t - t_{in})$$
$$V_i' + \frac{1}{RC}V_i = a\delta(t - t_{in})$$

Solve using the integrating factor $e^{\int \frac{1}{RC}dt}$:

$$V_i' e^{\int \frac{1}{RC}dt} + \frac{1}{RC}V_i e^{\int \frac{1}{RC}dt} = a\delta(t - t_{in})e^{\int \frac{1}{RC}dt}$$

$$\frac{d}{dt}\left[V_i e^{\left(\frac{t}{RC}\right)}\right] = a\delta(t - t_{in})e^{\left(\frac{t}{RC}\right)}$$

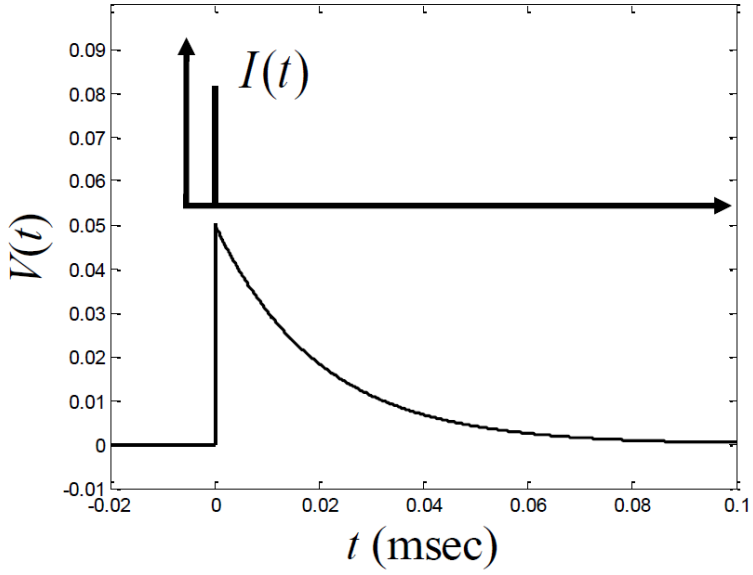
$$\int_0^{t'} \frac{d}{dt}\left[V_i e^{\left(\frac{t}{RC}\right)}\right] dt = a\int_0^{t'} \delta(t - t_{in})e^{\left(\frac{t}{RC}\right)} dt$$

For the RHS we need $t - t_{in} > 0$, so use Heaviside function:

$$\left[V_i e^{\left(\frac{t}{RC}\right)}\right]_0^{t'} = aH(t - t_{in})e^{\left(\frac{t_{in}}{RC}\right)}$$

$$V_i e^{\left(\frac{t}{RC}\right)} - V_i(t = 0) = aH(t - t_{in})e^{\left(\frac{t_{in}}{RC}\right)}$$

$$V_i(t) = aH(t - t_{in})e^{-(t-t_{in})/RC}$$



Exponential current

Alternatively, the synaptic input can be modelled as an exponential current:

$$I_{syn}(t) = H(t - t_{in})\frac{Ca}{\tau_s}\exp(-(t - t_{in})/\tau_s)$$

In this case the output becomes:

$$C\frac{dV_i}{dt} + \frac{V_i}{R} = H(t - t_{in})\frac{Ca}{\tau_s}\exp(-(t - t_{in})/\tau_s)$$

$$V_i' + \frac{1}{RC}V_i = H(t - t_{in})\frac{a}{\tau_s}\exp(-(t - t_{in})/\tau_s)$$

Solve using the integrating factor $e^{\int \frac{1}{RC}dt}$:

$$V_i' e^{\int \frac{1}{RC}dt} + \frac{1}{RC}V_i e^{\int \frac{1}{RC}dt} = H(t - t_{in})\frac{a}{\tau_s}\exp\left(-\frac{(t - t_{in})}{\tau_s}\right)e^{\int \frac{1}{RC}dt}$$

$$\frac{d}{dt}\left[V_i e^{\left(\frac{t}{RC}\right)}\right] = H(t - t_{in})\frac{a}{\tau_s}\exp\left(-\frac{(t - t_{in})}{\tau_s}\right)e^{\left(\frac{t}{RC}\right)}$$

$$\int_0^{t'} \frac{d}{dt}\left[V_i e^{\left(\frac{t}{RC}\right)}\right] dt = \frac{a}{\tau_s}\int_{t_{in}}^{t'} e^{\left(-\frac{t-t_{in}}{\tau_s}+\frac{t}{RC}\right)} dt$$

$$\left[V_i e^{\left(\frac{t}{RC}\right)}\right]_0^{t'} = \frac{a}{\tau_s}\int_{t_{in}}^{t'} e^{-\frac{RC}{\tau_2 RC}t+\frac{t_{in}RC}{\tau_s RC}+\frac{\tau_s}{RC\tau_s}t} dt$$

$$V_i e^{\left(\frac{t}{RC}\right)} - V_i(t=0) = \frac{a}{\tau_s} e^{\left(\frac{t_{in}}{\tau_s}\right)} \int_{t_{in}}^{t'} e^{\left(\frac{\tau_s - RC}{RC\tau_s}\right)t} dt$$

$$V_i e^{\left(\frac{t}{RC}\right)} = \frac{a}{\tau_s} e^{\left(\frac{t_{in}}{\tau_s}\right)} \frac{RC\tau_s}{\tau_s - RC} H(t-t_{in}) \left[e^{\left(\frac{\tau_s - RC}{RC\tau_s}\right)t}\right]_{t_{in}}^{t'}$$

$$V_i e^{\left(\frac{t}{RC}\right)} = a e^{\left(\frac{t_{in}}{\tau_s}\right)} \frac{RC}{\tau_s - RC} H(t-t_{in}) \left(e^{\left(\frac{\tau_s - RC}{RC\tau_s}\right)t} - e^{\left(\frac{\tau_s - RC}{RC\tau_s}\right)t_{in}}\right)$$

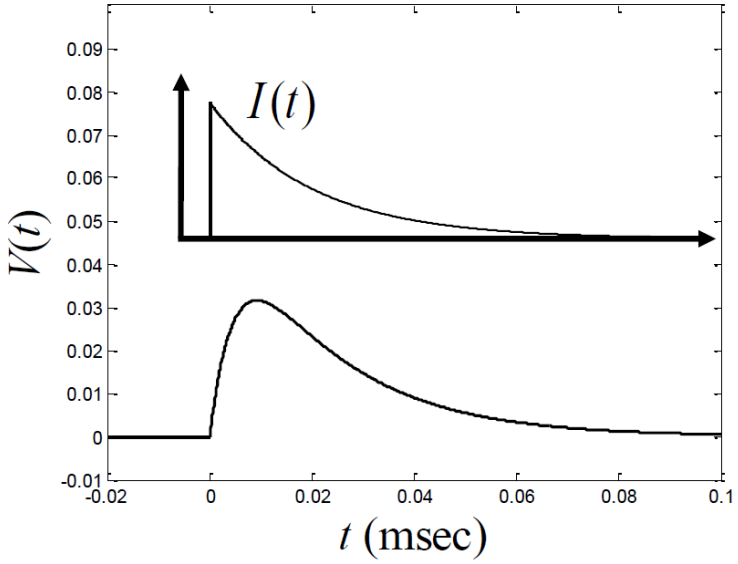
Setting $RC = \tau$ we can simplify this to:

$$V_i(t) = \frac{a\tau}{\tau_s - \tau} e^{\left(\frac{t_{in}}{\tau_s}-\frac{t}{\tau}\right)} H(t-t_{in}) \left(e^{\left(\frac{\tau_s - \tau}{\tau\tau_s}\right)t} - e^{\left(\frac{\tau_s - \tau}{\tau\tau_s}\right)t_{in}}\right)$$

$$= \frac{a\tau}{\tau_s - \tau} H(t-t_{in}) \left(e^{\left(\frac{t}{\tau}-\frac{t}{\tau_s}+\frac{t_{in}}{\tau_s}-\frac{t}{\tau}\right)} - e^{\left(\frac{t_{in}}{\tau}-\frac{t_{in}}{\tau_s}+\frac{t_{in}}{\tau_s}-\frac{t}{\tau}\right)}\right)$$

$$= \frac{a\tau}{\tau_s - \tau} H(t-t_{in}) \left(e^{\left(-\frac{t}{\tau_s}+\frac{t_{in}}{\tau_s}\right)} - e^{\left(\frac{t_{in}}{\tau}-\frac{t}{\tau}\right)}\right)$$

$$V_i(t) = \frac{a\tau}{\tau_s - \tau} H(t-t_{in}) \left(e^{-(t-t_{in})/\tau_s} - e^{-(t-t_{in})/\tau}\right)$$



Regardless of the spike function used, the total synaptic current due to several input synapses is:

$$I_{syn}(t) = C\sum_{k=1}^{n_E} a_{Ek} S_{Ek}(t) + C\sum_{k=1}^{n_I} a_{Ik} S_{Ik}(t)$$

Where n_E and n_I are the numbers of excitatory and inhibitory neurons respectively, each with its corresponding constant a and firing rate $S_k(t)$.

To implement a rate-based model with delta-function inputs, the following steps can be used:

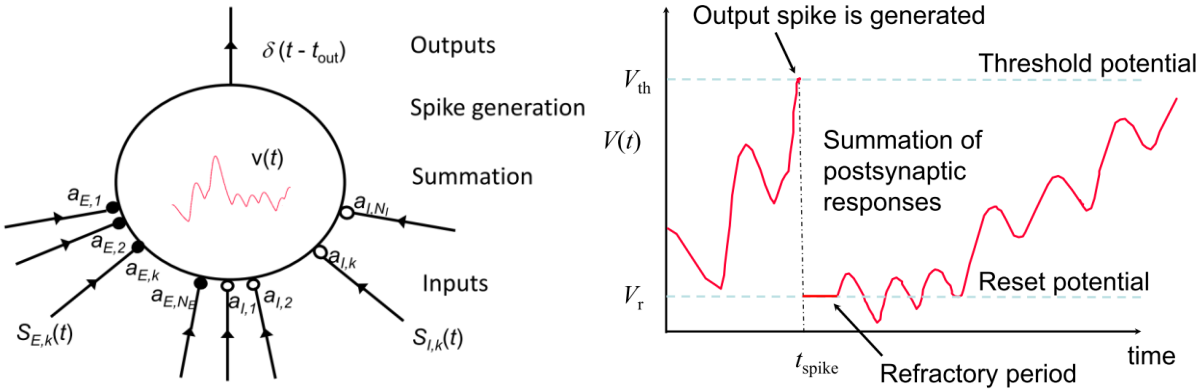
1. Decay the membrane potential using $V(t+\Delta t) = e^{-\Delta t/\tau}V(t)$.
2. Update the time increment $t = t + \Delta t$.
3. Add synaptic inputs to $V(t)$ using $V(t+\Delta t) = V(t+\Delta t) + I_{syn}(t)$.
4. Calculate the output using $S_i(t+\Delta t) = g(V(t+\Delta t) - \theta)$.
5. Determine if the neuron generates a spike at this time, by testing if $S_i(t+\Delta t) > \text{rand}(1)$.

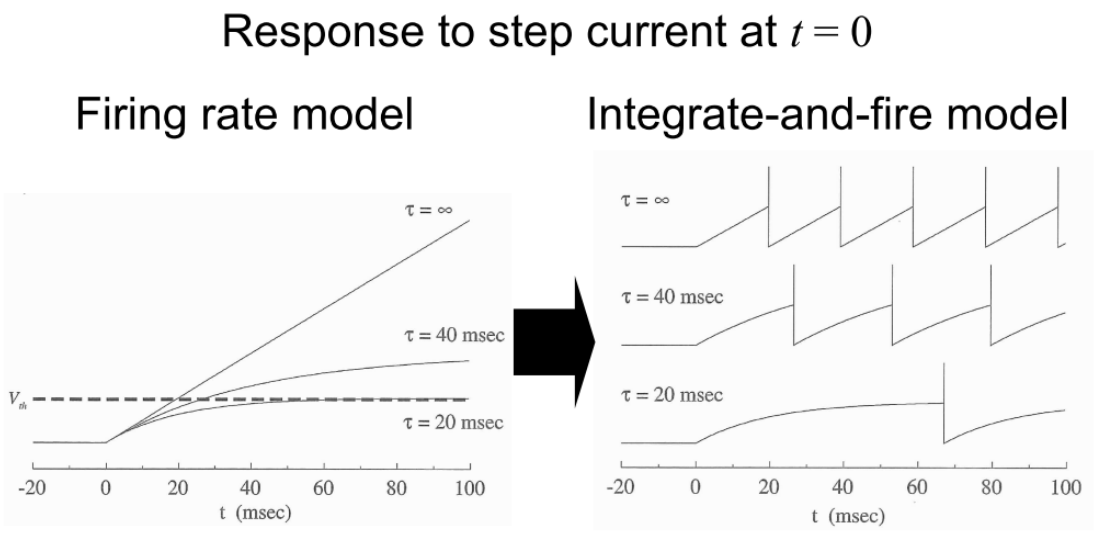
Note that the procedure is slightly different for exponential inputs.

Lecture 9

Leaky Integrate-and-Fire Neurons

Integrate and fire neuron models have two separate stages: a sub-threshold summation of inputs which occurs in the same way as in the rate-based model, and then the generation of a stereotyped action potential when threshold is reached. When this occurs, the membrane potential is reset.



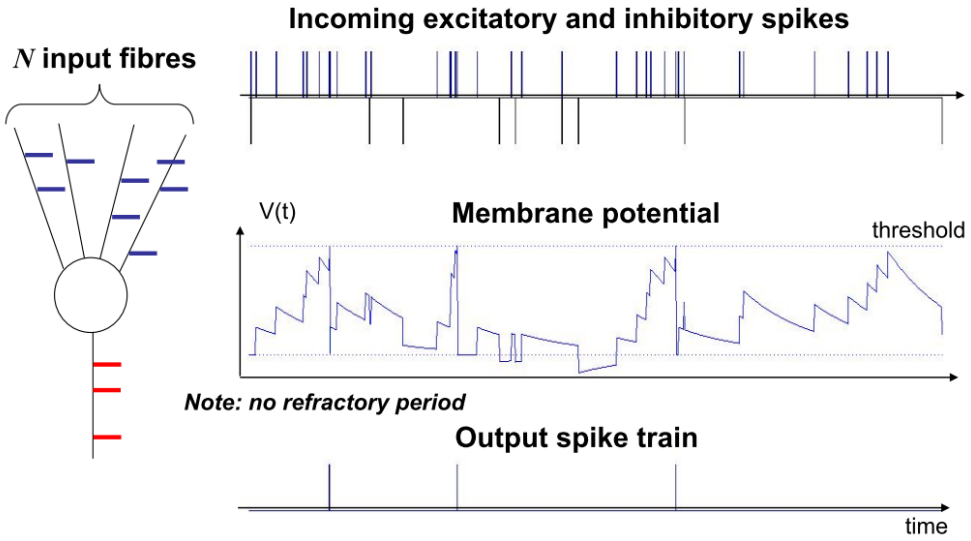


The equation of an integrate-and-fire neuron is given by:

$$\boxed{C\frac{dV_i}{dt} + \frac{V_i}{R} = I_i(t)}$$

$$\boxed{S_i(t) = \begin{cases}\delta(t) & \text{if } V_i(t) \geq V_{\text{th}} \\ 0 & \text{if } V_i(t) < V_{\text{th}}\end{cases}}$$

Note that in this model spikes are generated directly from the membrane voltage, rather than indirectly through the firing rate function $S\big(V(t)\big)$.



The equation for the membrane potential with constant input is the same as for the leaky integrator model:

$$V_i(t) = IR\big[1 - e^{(-t/RC)}\big] + V_i^0 e^{(-t/RC)}$$

The minimum current needed to trigger an action potential is called the threshold current:

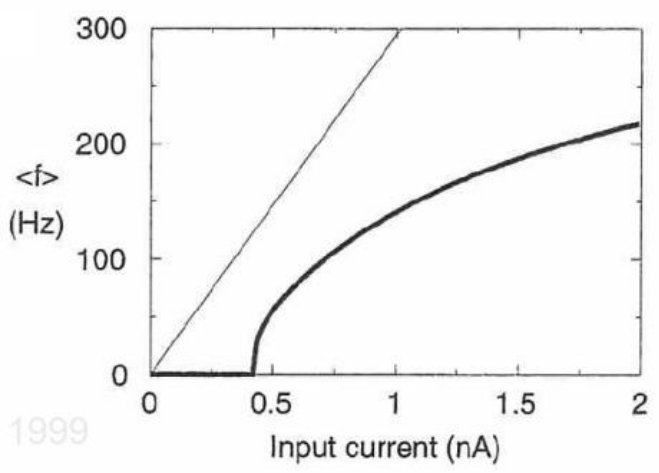
$$I_{\text{th}} = \frac{V_{\text{th}}}{R}$$

If the initial voltage is zero, then an output spike will be generated at time:

$$\begin{aligned} V_{\text{th}} &= IR\big[1 - e^{(-t/RC)}\big] \\ \frac{V_{\text{th}}}{IR} &= 1 - e^{(-t/RC)} \\ 1 - \frac{V_{\text{th}}}{IR} &= e^{(-t/RC)} \\ -\frac{t}{RC} &= \log\left(1 - \frac{V_{\text{th}}}{IR}\right) \\ t_{th} &= -RC\log\left(1 - \frac{V_{\text{th}}}{IR}\right) \end{aligned}$$

If the refractory period is t_{ref}, then the continuous firing rate f is given by:

$$f = \frac{1}{t_{\text{th}} + r_{\text{ref}}} = \frac{1}{r_{\text{ref}} - RC\log(1 - V_{\text{th}}/IR)}$$



Using this equation, we can plot the firing rate as a function of the input I. As shown in the figure below, if I is too low, the threshold is not reached an no spikes occur. In the limit of $I \to \infty$, the firing rate saturates at $f = 1/r_{\text{ref}}$.

Stochastic non-Leaky IF Neurons

This is a simplified integrate-and-fire model that allows us to introduce how stochastic inputs work. In the absence of a leak current the equation is:

$$C\frac{dV}{dt} = I(t)$$
$$S(t) = \begin{cases} \delta(t) & \text{if } V(t) \geq V_{\text{th}} \\ 0 & \text{if } V(t) < V_{\text{th}} \end{cases}$$

If the neuron receives excitatory input from a Poisson distributed process $N_e(t)$ with mean rate μ_e and synaptic weight w, then the membrane potential is a random process that shows random jumps:

$$I(t) = a_e \frac{dN_e}{dt}$$

Hence, we can solve the membrane potential equation as:

$$C\frac{dV}{dt} = a_e \frac{dN_e}{dt}$$
$$V(t) = \frac{a}{C} N_e(t)$$

If $a/C > V_{\text{th}}$, then each synaptic input is sufficient by itself to cause a spike, and so the waiting time between spikes is just the same as the waiting time between synaptic inputs. For a Poisson process the distribution for the next spike is:

$$P(t_{\text{th}} \leq t) = 1 - e^{-\mu_e t}$$

If instead n inputs are needed to trigger a spike, then we have:

$$n = \frac{V_{\text{th}}}{a_e}$$

Now the probability density for getting n inputs by time t is:

$$p_n(t) = \frac{\mu_e(\mu_e t)^{n-1} e^{-\mu_e t}}{(n-1)!}$$

This is called the nth-order gamma density. This gives rise to the following properties of the interspike interval t_{th}:

- $\text{mean}(t_{th}) = n/\mu_e$
- $\text{var}(t_{th}) = n/\mu_e^2$
- $C_V(t_{th}) = 1/\sqrt{n}$

This means that the distribution becomes more concentrated about the mean as n increases.

Incorporating an absolute refractory period results in a higher mean interspike interval without changing the variance, hence the C_V decreases. Thus, refractory periods reduce the relative variability of spikes.

We can modify this approach to incorporate both excitatory and inhibitory input:

$$V(t) = \frac{a_e}{C} N_e(t) - \frac{a_i}{C} N_i(t)$$

The expected value of the voltage of such a process is $a_e\mu_e t - a_i\mu_i t = \mu t$. This is called the drift. The variance of the membrane voltage is $a_e^2\mu_e t - a_i^2\mu_i t = \sigma^2 t$. This is the variance parameter. Note that because there is no current leakage, prior to the production of a spike, voltage mean and variance increase linearly with time.

Now we have the following interspike interval properties:

- $\text{mean}(t_{th}) = \frac{V_{\text{th}}}{\mu} = \frac{na_e}{\mu}$
- $\text{var}(t_{th}) = \frac{V_{\text{th}}\sigma^2}{\mu^3} = \frac{na_e\sigma^2}{\mu^3}$
- $C_V(t_{th}) = \left(\frac{1}{V_{\text{th}}}\frac{\sigma^2}{\mu}\right)^{1/2}$

Note that if $\mu_i = 0$ so there is no inhibition:

$$\begin{aligned}\text{mean}(t_{th}) &= \frac{na_e}{\mu} \\ &= \frac{na_e}{a_e\mu_e} \\ \text{mean}(t_{th}) &= \frac{n}{\mu_e}\end{aligned}$$

$$\begin{aligned}\text{var}(t_{th}) &= \frac{na_e\sigma^2}{\mu^3} \\ &= \frac{na_e a_e^2\mu_e}{a_e^3\mu_e^3} \\ \text{var}(t_{th}) &= \frac{n}{\mu_e^2}\end{aligned}$$

Which are the same as the values given before for the excitatory only case. It is evident that increased inhibitory input increases the mean interspike interval, and also increases the jitter (variance).

Lecture 10

Stochastic Leaky IF Neurons

Now we add back the leak current to the stochastic integrate-and-fire model:

$$C\frac{dV}{dt}+\frac{V}{R}=a_e\frac{dN_e(t)}{dt}-a_i\frac{dN_i(t)}{dt}$$
$$S(t)=\begin{cases}\delta(t) \text{ if } V(t)\geq V_{\text{th}}\\ 0 \quad \text{if } V(t)<V_{\text{th}}\end{cases}$$

This has the solution for the voltage of:

$$\frac{dV}{dt}+\frac{V}{RC}=\frac{a_e}{C}\frac{dN_e(t)}{dt}-\frac{a_i}{C}\frac{dN_i(t)}{dt}$$

Solve using the integrating factor $e^{\int\frac{1}{RC}dt}$ and assuming $V(0)=0$:

$$V'e^{\int\frac{1}{RC}dt}+\frac{1}{RC}Ve^{\int\frac{1}{RC}dt}=\frac{a_e}{C}\frac{dN_e(t)}{dt}e^{\int\frac{1}{RC}dt}-\frac{a_i}{C}\frac{dN_i(t)}{dt}e^{\int\frac{1}{RC}dt}$$
$$\frac{d}{dt}\left[Ve^{\left(\frac{t}{RC}\right)}\right]=\frac{a_e}{C}\frac{dN_e(t)}{dt}e^{\left(\frac{t}{RC}\right)}-\frac{a_i}{C}\frac{dN_i(t)}{dt}e^{\left(\frac{t}{RC}\right)}$$
$$\int_0^{t'}\frac{d}{dt}\left[Ve^{\left(\frac{t}{RC}\right)}\right]dt=\frac{a_e}{C}\int_0^{t'}\frac{dN_e(t)}{dt}e^{\left(\frac{t}{RC}\right)}dt-\frac{a_i}{C}\int_0^{t'}\frac{dN_i(t)}{dt}e^{\left(\frac{t}{RC}\right)}dt$$
$$\left[Ve^{\left(\frac{t}{RC}\right)}\right]_0^{t'}=\frac{a_e}{C}\int_0^{t'}\frac{dN_e(t)}{dt}e^{\left(\frac{t}{RC}\right)}dt-\frac{a_i}{C}\int_0^{t'}\frac{dN_i(t)}{dt}e^{\left(\frac{t}{RC}\right)}dt$$

Taking expectations of both sides yields:

$$E\left[V(t)e^{\left(\frac{t}{RC}\right)}\right]=\frac{a_e}{C}\int_0^{t'}E\left[\frac{dN_e(t)}{dt}\right]e^{\left(\frac{t}{RC}\right)}dt-\frac{a_i}{C}\int_0^{t'}E\left[\frac{dN_i(t)}{dt}\right]e^{\left(\frac{t}{RC}\right)}dt$$
$$E[V(t)]e^{\left(\frac{t}{RC}\right)}=\frac{a_e}{C}\int_0^{t'}\mu_e e^{\left(\frac{t}{RC}\right)}dt-\frac{a_i}{C}\int_0^{t'}\mu_i e^{\left(\frac{t}{RC}\right)}dt$$
$$E[V(t)]e^{\left(\frac{t}{RC}\right)}=R(a_e\mu_e-a_i\mu_i)\left[e^{\left(\frac{t}{RC}\right)}-1\right]$$
$$E[V(t)]=R(a_e\mu_e-a_i\mu_i)\left[1-e^{(-t/RC)}\right]$$

In the limit where $t\to\infty$ notice that the expected value tends to $R\mu$, rather than diverging as in the non-leaky case.

Conductance-based Neurons

In conductance-based models, the input current is now dependent upon the membrane voltage itself, thus introducing an additional complexity into the model.

The membrane voltage at which the postsynaptic current changes between excitatory and inhibitory is called the synaptic reversal potential, V_E for excitatory input and V_I for inhibitory input. Note that $I_{syn}=gV_{syn}$. As such we can write the postsynaptic current in terms of the reversal potential and membrane voltage as:

$$\boxed{I(t)=g(t-t_{in})\big(V_E-V(t)\big)}$$

To take a time-average of the conductance, we will treat it as a constant g. The equation for the membrane voltage is then:

$$\boxed{C\frac{dV(t)}{dt}+\frac{V(t)}{R}=g\left(V_{syn}-V(t)\right)}$$

With the process of spike generation being the same as before:

$$\boxed{S(t)=\begin{cases}\delta(t) \text{ if } V(t)\geq V_{\text{th}}\\ 0 \quad \text{ if } V(t)<V_{\text{th}}\end{cases}}$$

This is solved as follows:

$$\begin{aligned}C\frac{dV(t)}{dt}+\frac{V(t)}{R}&=g\left(V_{syn}-V(t)\right)\\ V'+\frac{1}{RC}V+\frac{g}{C}V&=gV_{syn}\\ V'+\left(\frac{1+Rg}{RC}\right)V&=\frac{g}{C}V_{syn}\end{aligned}$$

Which is the same as the leaky integrator neuron with time constant:

$$\tau_c=\frac{RC}{1+Rg}$$

Solve using the integrating factor $e^{\int\frac{1}{\tau}dt}$:

$$\begin{aligned}V'e^{\int\frac{1}{\tau}dt}+\left(\frac{1}{\tau}\right)Ve^{\int\frac{1}{\tau}dt}&=\frac{g}{C}V_{syn}e^{\int\frac{1}{\tau}dt}\\ \frac{d}{dt}\left[Ve^{\left(\frac{t}{\tau}\right)}\right]&=\frac{g}{C}V_{syn}e^{\left(\frac{t}{\tau}\right)}\\ \int_0^{t'}\frac{d}{dt}\left[Ve^{\left(\frac{t}{\tau}\right)}\right]dt&=\frac{g}{C}V_{syn}\int_0^{t'}e^{\left(\frac{t}{\tau}\right)}dt\\ \left[Ve^{\left(\frac{t}{\tau}\right)}\right]_0^{t'}&=\frac{g}{C}\tau V_{syn}\left[e^{\left(\frac{t}{\tau}\right)}\right]_0^{t'}\\ Ve^{\left(\frac{t}{\tau}\right)}-V_i(t=0)&=\frac{g}{C}\tau V_{syn}\left[e^{\left(\frac{t}{\tau}\right)}-1\right]\\ Ve^{\left(\frac{t}{\tau}\right)}&=\frac{g}{C}\tau V_{syn}\left[e^{\left(\frac{t}{\tau}\right)}-1\right]+V_0\\ V(t)&=\frac{g}{C}\tau V_{syn}\left[1-e^{(-t/\tau)}\right]+V_0e^{(-t/\tau)}\end{aligned}$$

Assuming $V_0=0$, we can compare this to the solution for non-conductance-based spiking models:

$$\begin{aligned}V(t)&=\frac{\tau_c}{C}gV_{syn}\left[1-e^{(-t/\tau)}\right]\\ V(t)&=\frac{\tau}{C}I\left[1-e^{(-t/\tau)}\right]\end{aligned}$$

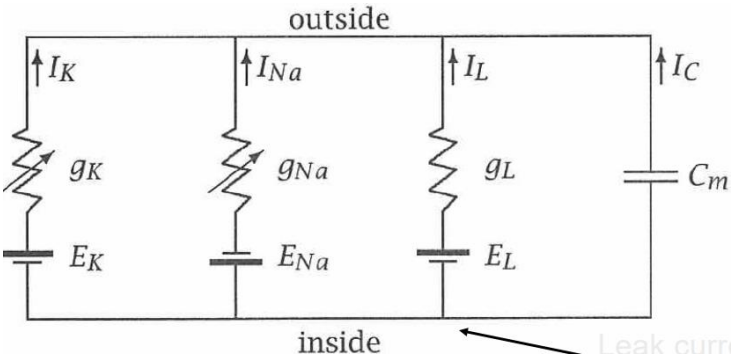
We see that these equations are equivalent (given that $V_{syn}=I/g$), apart from the different value of τ. For large conductance values, the effective membrane time constant is much smaller than the passive membrane time constant of RC.

Lecture 11

Hodgkin-Huxley Neurons

The Hodgkin-Huxley neuron model is a conductance-based model that incorporates more complex changes of g over time. The current equation is:

$$I_m = C_m \frac{dV}{dt} + I_K + I_{Na} + I_L$$



Substituting in the expression used in lecture 10 for a synaptic current (in this case one each for K, Na, and leak current), we have:

$$I_m = C_m \frac{dV}{dt} + I_K + I_{Na} + I_L$$
$$I_m = C_m \frac{dV}{dt} + g(V,t)(V - V_K) + g(V,t)(V - V_{Na}) + g_L(V - V_L)$$

Note here that the K and Na conductances are voltage and time dependent, while the leakage conductance is a constant. Hodgkin and Huxley parameterised the K and Na conductances using the following equations:

$$g_K(V,t) = G_K n^4$$
$$g_{Na}(V,t) = G_{Na} m^3 h$$
$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n$$
$$\frac{dm}{dt} = \alpha_m(1 - m) - \beta_m m$$
$$\frac{dh}{dt} = \alpha_h(1 - h) - \beta_h h$$

FitzHugh-Nagumo Neurons

A simpler way to look at Hodgkin-Huxley-style models is to examine the abstract model developed by van der Pol, Bonhoeffer, FitzHugh, Nagumo, Arimoto and Yoshizawa. This involves only two equations:

$$\frac{dV}{dt} = V - \frac{V^3}{3} - W + I$$
$$\frac{dW}{dt} = \phi(V + a - bW)$$

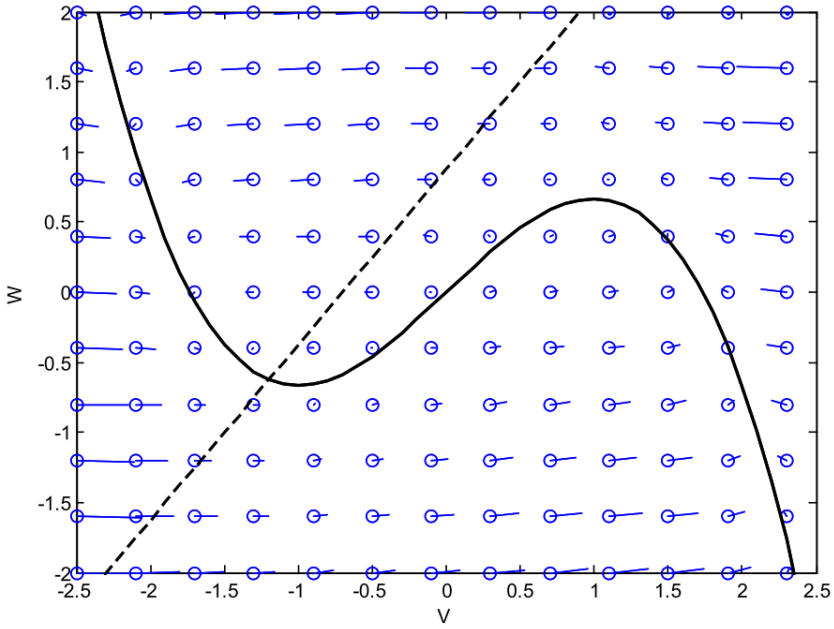
Where the parameters are set to $a = 0.7, b = 0.8$, and $\phi = 0.08$. W is an abstract parameter which represents the synaptic conductance changing over time.

This is called a singularly perturbed system, where one variable (V) changes much faster than the other variable (W). We can represent the trajectory of the system over time using a phase plane plot. On this we can plot the nullclines, the series of points where one derivative is equal to zero:

$$0 = V - \frac{V^3}{3} - W + I$$
$$0 = \phi(V + a - bW)$$

$$W = V - \frac{V^3}{3} + I$$
$$W = \frac{1}{b}(V + a)$$

These nullcline equations are shown on the following plot:



If the system is on the W nullcline, then the trajectory can only be horizontal since only V can change. If the system is on the V nullcline, then the trajectory can only be vertical since only W can change.

Zero current

Setting $I = 0$, we can solve for the stationary equilibrium point (intersection of the nullclines) as:

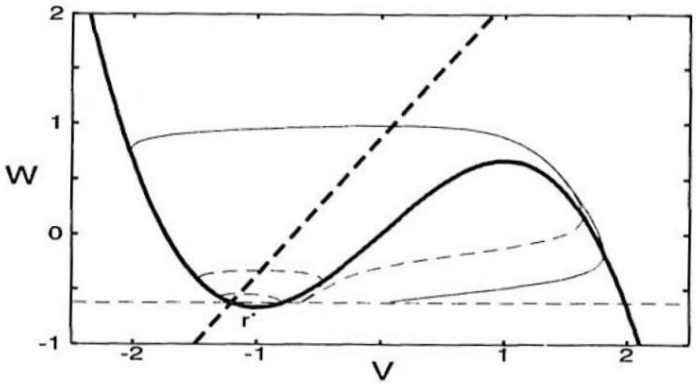
$$V - \frac{V^3}{3} = \frac{1}{b}(V + a)$$
$$V - \frac{V^3}{3} = \frac{1}{b}V + \frac{a}{b}$$
$$V - \frac{1}{b}V - \frac{V^3}{3} - \frac{a}{b} = 0$$
$$-\frac{1}{3}V^3 + \left(1 - \frac{1}{b}\right)V - \frac{a}{b} = 0$$
$$V^3 - 3\left(1 - \frac{1}{0.8}\right)V + 3\frac{0.7}{0.8} = 0$$
$$V^3 - 3(-0.25)V + 2.625 = 0$$

$$V^3 + 0.75V + 2.625 = 0$$

This has a real solution of $V = -1.2$. The system will decay to this stable point in a spiral pattern.

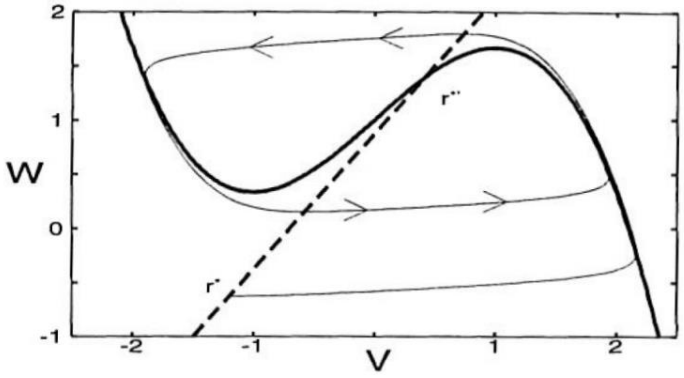
Delta current

Now instead setting $I(t) = Q\delta(t)$, we find that the value of V will jump by a unit Q. For small values of Q, such as 0.4 or 0.55, V will quickly decrease once again, and after a small cycle the system will return to equilibrium. By contrast, for large values of Q, such as 0.56 or 1.2, V will rapidly increase further, tracing out a large sweeping path before eventually returning to equilibrium. This corresponds to depolarisation triggering an action potential.



Constant current

With a steady input current, the V nullcline is shifted upward, while the W nullcline remains the same. This changes the position of the equilibrium point. The new equilibrium point is unstable, and the system will trace out a large cycle called a stable limit cycle. This corresponds to a regular train of action potentials.



A useful rule is that the equilibrium point is unstable <u>whenever the W nullcline meets the V nullcline where the V nullcline has a positive slope</u>. In such cases, small deviations from the equilibrium will push the system to the stable limit cycle.

Channel Model Neurons

These incorporate stochastic ion channel opening and closing as being the cause of the changes in conductances. Individual this is probabilistic, but when the numbers of channels are large, behaviour becomes predictable.

Multi-Compartment Neurons

These are an extension of conductance-based models which involve modelling a neuron as a series of connected compartments, each of which has a cable equation that is solved separately, with connected boundary conditions. The cable equation is given by:

$$C\frac{dT}{dt} = \frac{1}{2ar_L}\frac{d}{dx}\left(\frac{a^2 dV}{dx}\right) - i_{syn} + i_e$$

Where a is the radius of the cylindrical cable segment, r_L is the intracellular resistivity, i_{syn} is the synaptic current per unit area, and i_e is the electrode current per unit area.

Neural Learning

Lecture 12

The Synaptic Basis of Learning

Since neurotransmitter release is probabilistic, synaptic transmission itself is stochastic. Synaptic weight is affected by the number of neurotransmitter release sites n, the probability of release per site p, and the density of receptors q. The first two are presynaptic variables, while the last is postsynaptic. The simplest expression for the average synaptic transmission is simply the product:

$$R = npq$$

Type	Duration	Location	Occurrence	Cause
Paired-pulse facilitation	100 msec	Presynaptic	Response to second single presynaptic stimulus is greater than the first single stimulus.	Increased presynaptic [Ca2+] leads to a greater release of synaptic vesicles (higher p).
Augmentation	10 sec	Presynaptic	Repetitive stimulation increases the postsynaptic response.	Increased presynaptic [Ca2+] leads to a greater release of synaptic vesicles (higher p).
Post-tetanic potentiation	1 min	Presynaptic	Brief, high-frequency stimulus increases the postsynaptic response.	Increased presynaptic [Ca2+] leads to a greater release of synaptic vesicles (higher p).
Long-term potentiation	Hours+	Pre- and postsynaptic	Simultaneous presynaptic and postsynaptic activity with high intracellular [Ca2+] produces sustained increase in postsynaptic activity.	Mediated by NMDA receptors.
Long-term depression	Hours+	Pre- and postsynaptic	Simultaneous presynaptic and postsynaptic activity with low intracellular [Ca2+] produces sustained decrease in postsynaptic activity.	May be mediated by NMDA receptors
Spike-timing dependent	Hours+	Pre- and postsynaptic	Upregulation of postsynaptic firing rate when presynaptic	

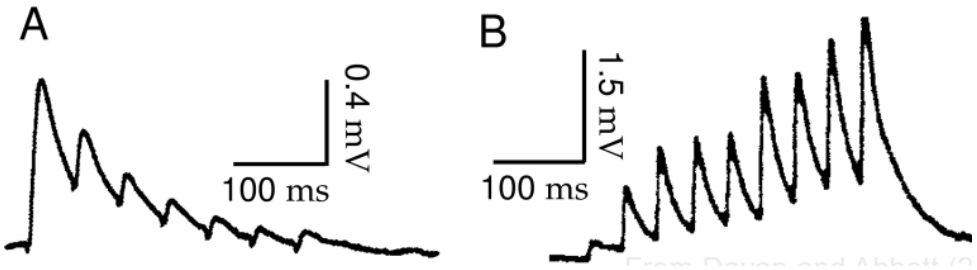
plasticity			activity precedes postsynaptic activity; downregulation in the opposite case.	
Activity-Dependent Synaptic Scaling	Hours+	Postsynaptic	Chronically elevated activity leads to reduction in synaptic weight.	Regulation of AMPA receptors.

Short-term Synaptic Plasticity

In short-term synaptic plasticity, changes persist on the order of milliseconds to seconds. All forms of short-term plasticity appear to only depend on the presynaptic terminal, and are all thought to relate to temporary build-up of calcium ions in the presynaptic cytoplasm.

Paired-pulse facilitation

In paired-pulse facilitation, the synaptic response of a single stimulus modifies the response to the next stimulus by increasing the probability of neurotransmitter release (i.e., increasing p). This probably then decays back to baseline after a few hundred milliseconds.



This can be modelled using a simple exponential:

$$p(t) = p_0 + (p_f - p_0)e^{-\frac{t}{\tau_f}}$$

where p_0 and p_f are the probabilities before and after facilitation, respectively, and τ_f is the characteristic decay time of facilitation.

Augmentation

Augmentation is a form of short-term synaptic plasticity which also increases the probability of releasing synaptic vesicles during and after repetitive stimulation. The main difference between paired-pulse facilitation and augmentation is that augmentation occurs in response to multiple spikes, whereas paired-pulse facilitation is relevant to spikes in response to a single stimulus.

Augmentation can be modelled using a similar exponential equation:

$$p(t) = p_0 + (p_a - p_0)e^{-\frac{t}{\tau_a}}$$

Post-tetanic potentiation

Post-tetanic potentiation occurs following a tetanic stimulus, which is a brief high-frequency stimulus. It is essentially the same as augmentation, except that it only occurs following a very high frequency stimulus, and tends to have a much longer time constant.

It too is modelled using an exponential equation:

$$p(t) = p_0 + (p_{PTP} - p_0)e^{-\frac{t}{\tau_{PTP}}}$$

Paired-pulse depression

This temporary reduction in the synaptic strength following presynaptic activity is thought to be due to the depletion in presynaptic vesicles.

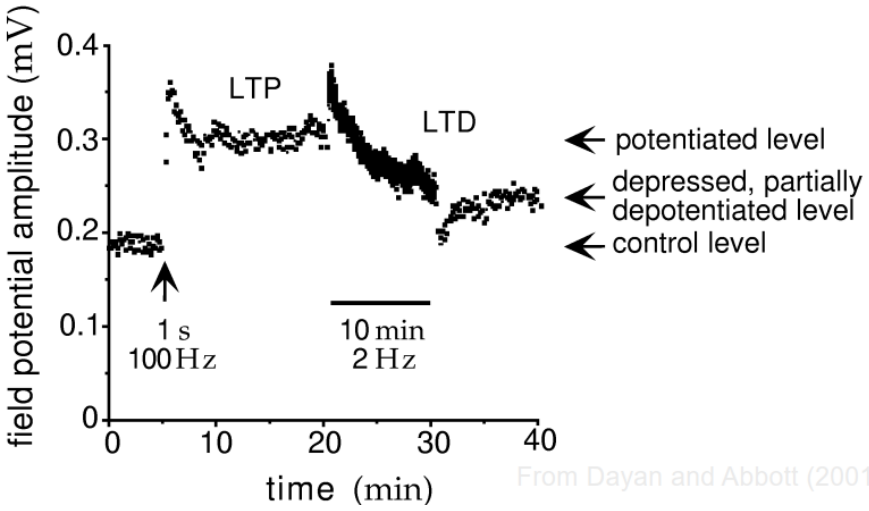
Long-Term Synaptic Plasticity

Apart from lasting for much longer, long-term plasticity usually depends on both presynaptic and postsynaptic activity. These forms also appear to require protein synthesis.

Long-term potentiation

Long-term potentiation (LTP) is a rapid and sustained increase in synaptic strength following a brief but potent stimulus. LTP can last for hours, days, weeks or longer. Interest in LTP is in part due to it providing a possible model for learning and memory.

LTP is known to occur following simultaneous presynaptic neurotransmitter release and postsynaptic polarisation. It is thought to be mediated by NMDA receptors, which are directly gated by both membrane voltage and neurotransmitters, meaning that they pass current only when the membrane is depolarised sufficiently to relieve a block by magnesium ions.



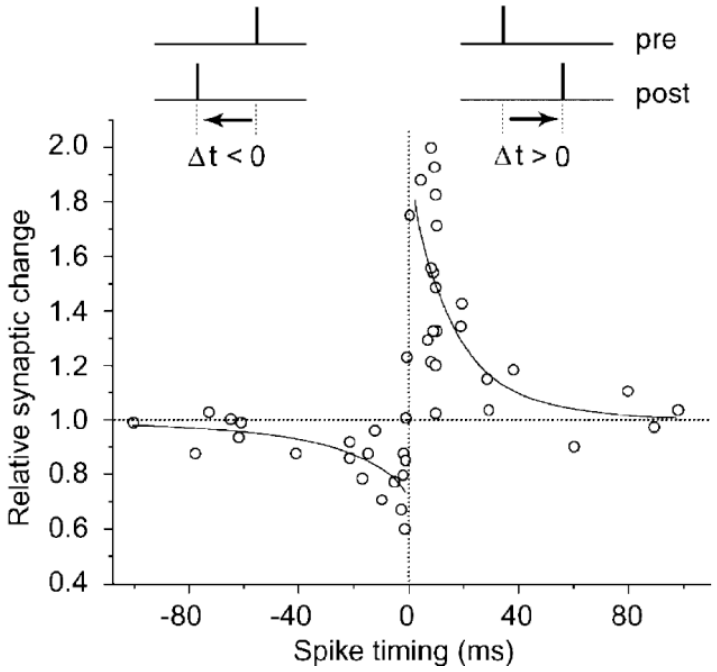
From Dayan and Abbott (2001

Long-term depression

Long-term depression is induced in a similar way to LTP. Some forms require NMDA receptor activation. Whether LTD or LTP is induced depends on a threshold of free intracellular calcium concentration. Above the threshold, LTP is induced, whereas below the threshold, LTD is induced.

Spike-timing dependent plasticity

This process adjusts the connection strengths based on the relative timing of a particular neuron's output and input action potentials. If an input spike to a neuron tends, on average, to occur immediately before that neuron's output spike, then that input is made somewhat stronger. If an input spike tends, on average, to occur immediately after an output spike, then that input is made somewhat weaker.



Activity-Dependent Synaptic Scaling

Activity-dependent synaptic scaling (ADSS) is a homeostatic mechanism in which the brain responds to chronically elevated activity in a neural circuit with negative feedback, allowing individual neurons to reduce their overall action potential firing rate. ADSS is a postsynaptic mechanism and does not seem to depend on presynaptic spike activity. Instead, it modifies each synapse by the same multiplicative factor, thereby preserving the relative weights.

Lecture 13

Hebbian Learning

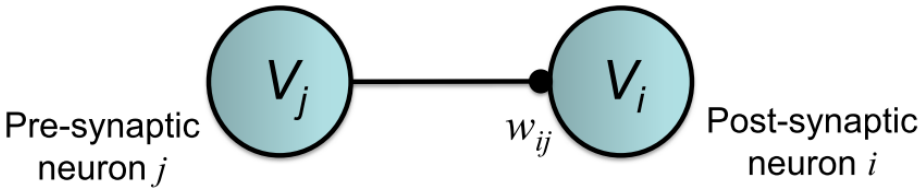
Table for forms of Hebbian learning:

Rule	Equation	Effects
Basic Hebb rule	$\tau_w \frac{d\tilde{w}}{d\tau} = \langle v\tilde{u} \rangle$	Unstable. Projects output vector parallel to principal eigenvector of the correlation matrix of input.
Covariance rule	$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (v - \theta_v)\tilde{u} \rangle$	Unstable. Projects output vector parallel to principal eigenvector of the covariance matrix of input.
Oja's rule	$\tau_w \frac{d\tilde{w}}{dt} = \langle v\tilde{u} \rangle - \alpha v^2 \tilde{w}$	Stable. Local multiplicative normalisation form of basic Hebb rule.

Subtractive normalisation	$\tau_w \frac{d\tilde{w}}{dt} = \langle v\tilde{u} \rangle - \frac{v}{N_\mu}(\tilde{n} \cdot \tilde{u})\tilde{n}$	Stable. Non-local subtractive normalisation form of basic Hebb rule.

The Hebbian learning rule is a local and cooperate learning rule, meaning that pre- and postsynaptic neurons must both be simultaneously active for a synaptic change to occur, and that information for synaptic changes is only information locally available to the synapse. Hebb's rule can be written mathematically as:

$$\Delta w_{ij} = v_i v_j$$



The postsynaptic activity v of a single firing-rate neuron is described using the equation:

$$\tau_r \frac{dv}{d\tau} = -v + \sum_{i=1}^{N} w_i u_i$$

Here N is the number of presynaptic neurons and w_i is the weight between neuron i and the postsynaptic neuron.

To simplify the analysis, we can make the adiabatic assumptions:

1. The process of synaptic plasticity is much slower than the firing-rate dynamics of the model.
2. Stimuli are presented slowly enough to allow the network to obtain its steady-state.

This leads us to set the time derivative to zero, hence yielding the steady-state equation:

$$v = \tilde{w} \cdot \tilde{u}$$

With this setup, the Hebbian learning rule can be expressed as:

$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle v\tilde{u} \rangle$$

Here τ_w is the time constant that controls the rate of synaptic change. Given this equation, the expression $v\tilde{u}$ may be interpreted as a measure of the probability that the pre- and postsynaptic neurons both fire spikes during a short time interval.

To understand the process, we can take the average over the set of input patterns, and also substitute the expression above for the output $v = \tilde{w} \cdot \tilde{u}$ to find (using this property):

$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle v\tilde{u} \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (\tilde{w}^T \tilde{u})\tilde{u} \rangle$$

$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle \tilde{u}(\tilde{w}^T \tilde{u}) \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle \tilde{u}(\tilde{u}^T \tilde{w}) \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle \tilde{u}\tilde{u}^T \rangle \tilde{w}$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = Q\tilde{w}$$

Here Q is the input correlation matrix.

Stability of Hebbian Learning

To find the derivative of the magnitude of the weights, we differentiate as follows:

$$\tau_w \frac{d(\tilde{w} \cdot \tilde{w})}{d\tau} = \tau_w \frac{d(\tilde{w})}{d\tau} \cdot \tilde{w} + \tau_w \tilde{w} \cdot \frac{d(\tilde{w})}{d\tau}$$
$$\tau_w \frac{d|\tilde{w}|^2}{d\tau} = 2\tau_w \frac{d(\tilde{w})}{d\tau} \cdot \tilde{w}$$
$$\tau_w \frac{d|\tilde{w}|^2}{d\tau} = 2(v\tilde{u}) \cdot \tilde{w}$$
$$\tau_w \frac{d|\tilde{w}|^2}{d\tau} = 2v(\tilde{u} \cdot \tilde{w})$$
$$\tau_w \frac{d|\tilde{w}|^2}{d\tau} = 2v^2$$

Notice that this is always positive, meaning that the magnitude of the weight vector continuously grows. To avoid this, we need to impose bounds on the weights.

Learning Dynamics

To consider the learning dynamics of the Hebbian learning equation, define the eigenvectors e_μ of Q:

$$Q\tilde{e}_\mu = \lambda_\mu \tilde{e}_\mu$$

Since the eigenvectors form a complete basis for the activity space, any weight vector can be represented as a weighted sum of eigenvectors:

$$\tilde{w}(t) = \sum_{\mu=1}^{N} c_\mu(t)\tilde{e}_\mu$$

Substituting this into the weight update equation we have:

$$\tau_w \frac{d\tilde{w}}{dt} = Q\tilde{w}$$
$$\tau_w \frac{d}{dt} \sum_{\mu=1}^{N} c_\mu(t)\tilde{e}_\mu = Q \sum_{\mu=1}^{N} c_\mu(t)\tilde{e}_\mu$$
$$\tau_w \frac{d}{dt} \sum_{\mu=1}^{N} c_\mu(t)\tilde{e}_\mu = \sum_{\mu=1}^{N} c_\mu(t) Q\tilde{e}_\mu$$

$$\tau_w \frac{d}{dt} \sum_{\mu=1}^{N} c_\mu(t) \tilde{e}_\mu = \sum_{\mu=1}^{N} c_\mu(t) \lambda_\mu \tilde{e}_\mu$$

Hence for each eigenvalue μ:

$$\frac{d}{dt} c_\mu(t) = \frac{\lambda_\mu}{\tau_w} c_\mu(t)$$

$$c_\mu(t) = c_\mu(0) \exp\left(\frac{\lambda_\mu}{\tau_w} t\right)$$

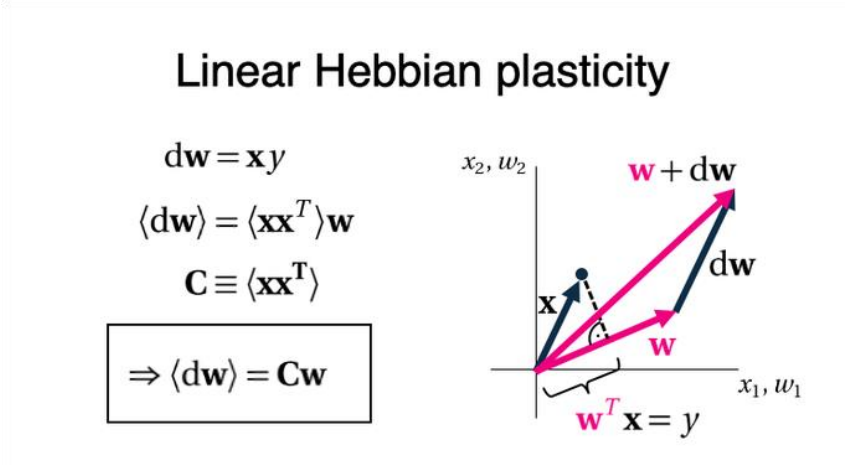
The weights then become:

$$\tilde{w}(t) = \sum_{\mu=1}^{N} \left(\tilde{w}(0) \cdot \tilde{e}_\mu\right) \exp\left(\frac{\lambda_\mu}{\tau_w} t\right) \tilde{e}_\mu$$

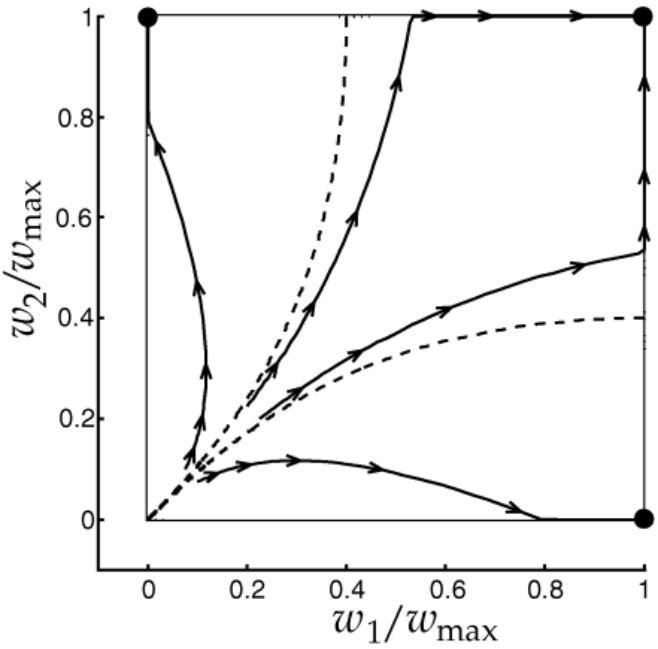
The exponential term grows over time because the eigenvalues are all non-negative. For large t, the term with the largest eigenvalue becomes much larger than any of the other terms and dominates the sum for w. The corresponding eigenvector $\tilde{e}_1$ is called the principal eigenvector. As such, after training the response to (most) arbitrary input vectors is well-approximated by:

$$v \propto \tilde{e}_1 \cdot \tilde{u}$$

Hence, we can regard Hebbian plasticity as performing a projection of the input vector onto the principal eigenvector of the correlation matrix of the inputs used during training.



The trajectories of training are shown in the diagram below. For the initial weight combinations below the dotted lines, the weights tend to the corners. Within the dotted lines, it converges to the top corner.



Here the final state does not converge to being parallel with the principal eigenvector, as the weights hit the saturation boundary first.

Oja's Rule

Oja's rule is a modification of the Hebb rule that provides stability and only requires information that is local to the modified synapse. However, it is based on theoretical arguments and not experimental data.

$$\tau_w \frac{d\tilde{w}}{dt} = v\tilde{u} - \alpha v^2 \tilde{w}$$

As before, we can consider how the magnitude of the weights change over time:

$$\begin{aligned}
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= \tau_w \frac{d(\tilde{w})}{d\tau} \cdot \tilde{w} + \tau_w \tilde{w} \cdot \frac{d(\tilde{w})}{d\tau} \\
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= (v\tilde{u} - \alpha v^2 \tilde{w}) \cdot \tilde{w} + \tilde{w} \cdot (v\tilde{u} - \alpha v^2 \tilde{w}) \\
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= 2(v\tilde{u} - \alpha v^2 \tilde{w}) \cdot \tilde{w} \\
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= 2(v\tilde{u} \cdot \tilde{w} - \alpha v^2 \tilde{w} \cdot \tilde{w}) \\
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= 2(v^2 - \alpha v^2 |\tilde{w}|^2) \\
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= 2v^2(1 - \alpha |\tilde{w}|^2)
\end{aligned}$$

We can find the steady state by setting the derivative to zero:

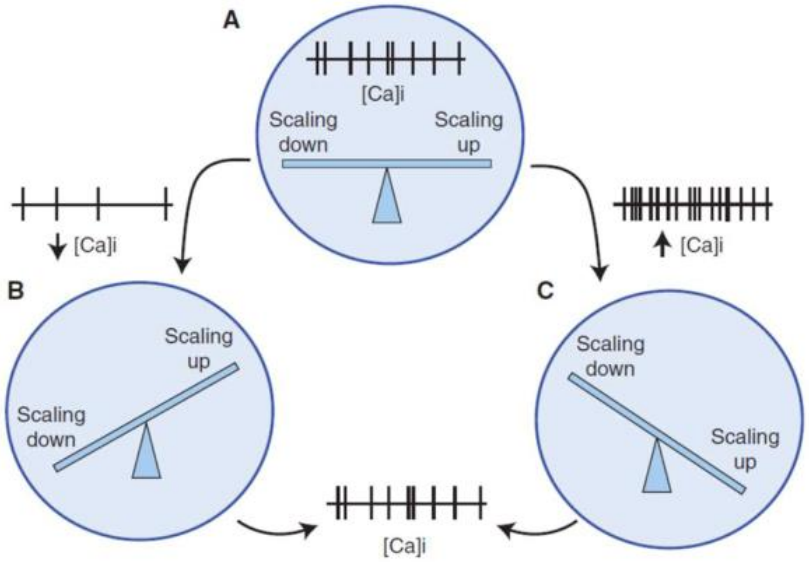
$$\begin{aligned}
0 &= 2v^2(1 - \alpha |\tilde{w}|^2) \\
0 &= 1 - \alpha |\tilde{w}|^2 \\
|\tilde{w}|^2 &= \frac{1}{\alpha}
\end{aligned}$$

Hence, we see that weights do not grow without bound.

Lecture 14

Multiplicative Normalisation

Activity Dependent Synaptic Scaling (ADSS) is a mechanism that adjusts the synaptic weights during learning to regulate postsynaptic activity. Synaptic scaling involves neurons detecting changes in their own firing rates through a set of calcium-dependent sensors that regulate the number of glutamate receptors on the cell membrane. Higher levels of activity lead to higher calcium concentrations, leading to scaling down of the number of glutamate receptors and hence a reduction in activity. The reverse occurs with low levels of activity.



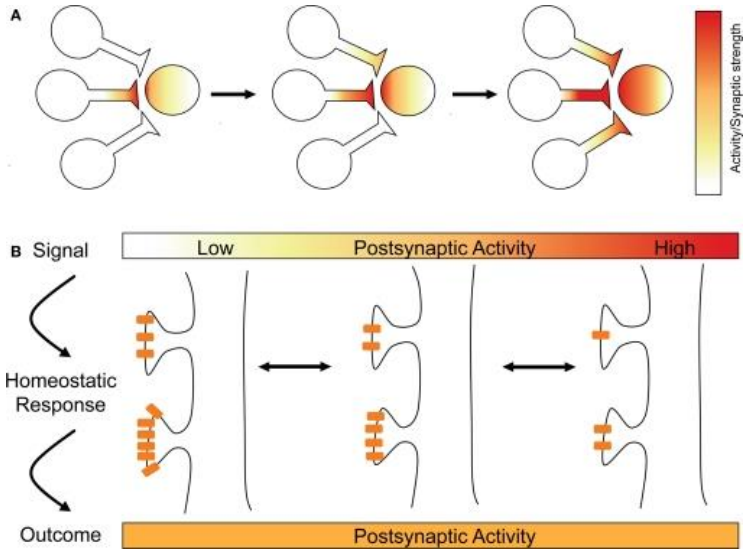
The postsynaptic activity of a cell can be represented by a slowly varying metric $a(t)$:

$$\tau \frac{da(t)}{dt} = -a(t) + \sum_i \delta(t - t_i)$$

In multiplicative ADSS, the weights are updated by a common factor β:

$$\frac{dw(t)}{dt} = \beta w(t)\left[a_{goal} - a(t)\right]$$

While Hebbian plasticity mechanisms modify neural synaptic connections selectively, synaptic scaling normalizes all neural synaptic connections by decreasing the strength of each synapse by the same factor (multiplicative change), so that the relative synaptic weighting of each synapse is preserved. Oja's rule is a form of multiplicative normalisation.



Subtractive Normalisation

Hebbian learning can be made stable by applying subtractive normalisation, where the same amount is subtracted from all weights regardless of their magnitude. This is a non-local operation, as it requires that the sum of all input activity be available to each individual synapse.

The modified Hebbian update rule takes the form:

$$\tau_w \frac{d\tilde{w}}{dt} = v\tilde{u} - \frac{v}{N_\mu}(\tilde{n} \cdot \tilde{u})\tilde{n}$$

Where $\tilde{n}$ is an N_μ-dimensional vector with all components equal to 1.

Subtractive normalisation must be augmented by a saturation constraint to prevent weights from becoming negative.

Ocular Dominance

Ocular dominance refers to the phenomenon where neurons in V1 tend to respond primarily to neurons from one eye over the other. We can understand ocular dominance as a manifestation of subtractive normalisation.

Consider a single neuron in V1 which receives two inputs from the LGN, one associated with the right eye with activity u_R and the other associated with the left eye with activity u_L. The output of the V1 neuron is then given by:

$$v = w_R u_R + w_L u_L$$

The input correlation matrix is given by:

$$\begin{aligned} Q &= \langle \tilde{u} \cdot \tilde{u} \rangle \\ &= \begin{bmatrix} \langle u_R u_R \rangle & \langle u_R u_L \rangle \\ \langle u_L u_R \rangle & \langle u_L u_L \rangle \end{bmatrix} \\ Q &= \begin{bmatrix} q_S & q_D \\ q_D & q_S \end{bmatrix} \end{aligned}$$

To solve for the equilibrium behaviour of the network after Hebbian learning, we need to find the eigenvalues and eigenvectors of this matrix.

To find the eigenvalues:

$$\begin{vmatrix} q_S - \lambda & q_D \\ q_D & q_S - \lambda \end{vmatrix} = 0$$
$$0 = (q_S - \lambda)^2 - q_D^2$$
$$0 = q_S^2 - 2q_S\lambda + \lambda^2 - q_D^2$$
$$0 = \lambda^2 - 2q_S\lambda + (q_S^2 - q_D^2)$$
$$0 = \big(\lambda + (q_S + q_D)\big)\big(\lambda + (q_S - q_D)\big)$$
$$\lambda_1 = q_S + q_D, \qquad \lambda_2 = q_S - q_D$$

To find the eigenvectors:

$$Q\tilde{e}_1 = \lambda_1\tilde{e}_1$$
$$(Q - I\lambda_1)\tilde{e}_1 = 0$$
$$\begin{bmatrix} -q_D & q_D \\ q_D & -q_D \end{bmatrix}\begin{bmatrix} e_1^1 \\ e_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} e_1^1 \\ e_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$e_1^1 = e_1^2$$
$$\therefore \tilde{e}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$Q\tilde{e}_2 = \lambda_2\tilde{e}_2$$
$$(Q - I\lambda_2)\tilde{e}_2 = 0$$
$$\begin{bmatrix} q_D & q_D \\ q_D & q_D \end{bmatrix}\begin{bmatrix} e_2^1 \\ e_2^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} e_2^1 \\ e_2^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$e_2^1 = -e_2^2$$
$$\therefore \tilde{e}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Since the input from the two eyes are likely to be correlated, $q_D > 0$, and hence λ_1 will be the principal eigenvector. As we saw before, in the long term this means the weight vector will become parallel to the principal eigenvector. The weights then will become:

$$\tilde{w}(t) = \left(\tilde{w}(0) \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\exp\left(\frac{q_S + q_D}{\tau_w}t\right)\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$\tilde{w}(t) = (\tilde{w}(0) \cdot \tilde{e}_1)\exp\left(\frac{q_S + q_D}{\tau_w}t\right)\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Hence we see that $w_R = w_L$, which does not result in ocular dominance.

Now instead consider what happens when we use the subtractive normalisation rule:

$$\tau_w\frac{d\tilde{w}}{dt} = v\tilde{u} - \frac{v}{N_\mu}(\tilde{n} \cdot \tilde{u})\tilde{n}$$
$$\tau_w\frac{d\tilde{w}}{dt} = (\tilde{w}^T\tilde{u})\tilde{u} - \frac{(\tilde{w}^T\tilde{u})}{2}(\tilde{n}^T\tilde{u})\tilde{n}$$
$$\tau_w\frac{d\tilde{w}}{dt} = \tilde{u}(\tilde{u}^T\tilde{w}) - \frac{1}{2}(\tilde{w}^T\tilde{u})\tilde{n}(\tilde{n}^T\tilde{u})$$
$$\tau_w\frac{d\tilde{w}}{dt} = Q\tilde{w} - \frac{1}{2}(\tilde{w}^T\tilde{u})(\tilde{n}\tilde{n}^T)\tilde{u}$$
$$\tau_w\frac{d\tilde{w}}{dt} = \begin{bmatrix} u_Ru_R & u_Ru_L \\ u_Lu_R & u_Lu_L \end{bmatrix}\tilde{w} - \frac{1}{2}(\tilde{w}^T\tilde{u})\begin{bmatrix} u_R + u_L \\ u_R + u_L \end{bmatrix}$$
$$\tau_w\frac{d\tilde{w}}{dt} = \begin{bmatrix} u_Ru_R & u_Ru_L \\ u_Lu_R & u_Lu_L \end{bmatrix}\tilde{w} - \frac{1}{2}\begin{bmatrix} u_R \\ u_L \end{bmatrix}[u_R + u_L \quad u_R + u_L]\tilde{w}$$

$$\tau_w \frac{d\tilde{w}}{dt} = \begin{bmatrix} u_R u_R & u_R u_L \\ u_L u_R & u_L u_L \end{bmatrix} \tilde{w} - \frac{1}{2} \begin{bmatrix} u_R u_R + u_R u_L & u_R u_R + u_R u_L \\ u_R u_L + u_L u_L & u_R u_L + u_L u_L \end{bmatrix} \tilde{w}$$
$$\tau_w \frac{d\tilde{w}}{dt} = \frac{1}{2} \begin{bmatrix} u_R u_R - u_R u_L & u_R u_L - u_R u_R \\ u_L u_R - u_L u_L & u_L u_L - u_R u_L \end{bmatrix} \tilde{w}$$
$$\tau_w \frac{d\tilde{w}}{dt} = \frac{1}{2} \begin{bmatrix} q_S - q_D & q_D - q_S \\ q_D - q_S & q_S - q_D \end{bmatrix} \tilde{w}$$
$$\tau_w \frac{d\tilde{w}}{dt} = \frac{1}{2} \begin{bmatrix} a & -a \\ -a & a \end{bmatrix} \tilde{w}$$

To find the eigenvalues of the matrix we use:

$$\frac{1}{2}\left(\left(\frac{a}{2} - \lambda\right)^2 - \left(\frac{a}{2}\right)^2\right) = 0$$
$$\frac{a^2}{4} - a\lambda + \lambda^2 - \frac{a^2}{4} = 0$$
$$\lambda^2 - a\lambda = 0$$
$$\lambda(\lambda - a) = 0$$
$$\lambda_1 = a, \qquad \lambda_2 = 0$$

Now solving for the eigenvectors we have:

$$Q\tilde{e}_1 = \lambda_1 \tilde{e}_1$$
$$(Q - I\lambda_1)\tilde{e}_1 = 0$$
$$\begin{bmatrix} \frac{a}{2} - a & -\frac{a}{2} \\ -\frac{a}{2} & \frac{a}{2} - a \end{bmatrix} \begin{bmatrix} e_1^1 \\ e_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} -\frac{a}{2} & -\frac{a}{2} \\ -\frac{a}{2} & -\frac{a}{2} \end{bmatrix} \begin{bmatrix} e_1^1 \\ e_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} e_1^1 \\ e_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$e_1^1 = -e_1^2$$
$$\therefore \tilde{e}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Substituting this into the equation for the weights we find:

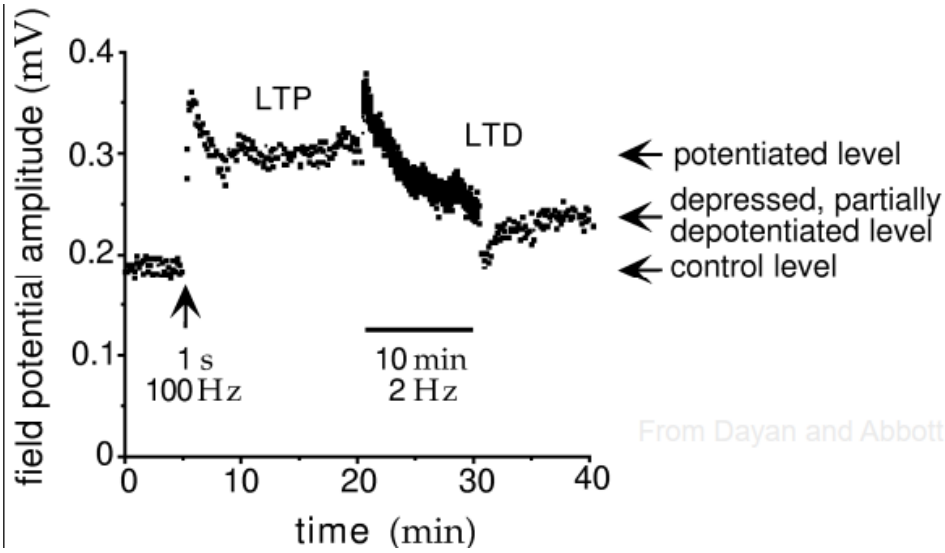
$$\tilde{w}(t) = (\tilde{w}(0) \cdot \tilde{e}_1) \exp\left(\frac{q_S - q_D}{\tau_w} t\right) \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Note that now the opposite eigenvector to before is growing and becomes dominant. The direction of growth will depend on the initial condition of the weights. If $\tilde{w}(0) \cdot \tilde{e}_1$ is positive, w_R increases and w_L decreases, while if it is negative, then w_L increases and w_R decreases. Either way, one weight will grow while the other will approach zero, thereby achieving ocular dominance.

The Covariance Rule

Experimental results indicate that when cells are exposed to high rates of activity sufficient to generate postsynaptic activity, the synaptic weight is increased. Conversely, when cells are exposed to low rates

of activity that is nonetheless sufficient to generate postsynaptic activity, the synaptic weight is decreased. These effects are shown in the diagram below.



From Dayan and Abbott

We can model these effects using a modification of the standard Hebbian learning rule. The standard rule has the form:

$$\tau_w \frac{d\tilde{w}}{d\tau} = v\tilde{u}$$

There are two possible modifications:

$$\tau_w \frac{d\tilde{w}}{d\tau} = (v - \theta_v)\tilde{u}$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = v(\tilde{u} - \hat{\theta}_u)$$

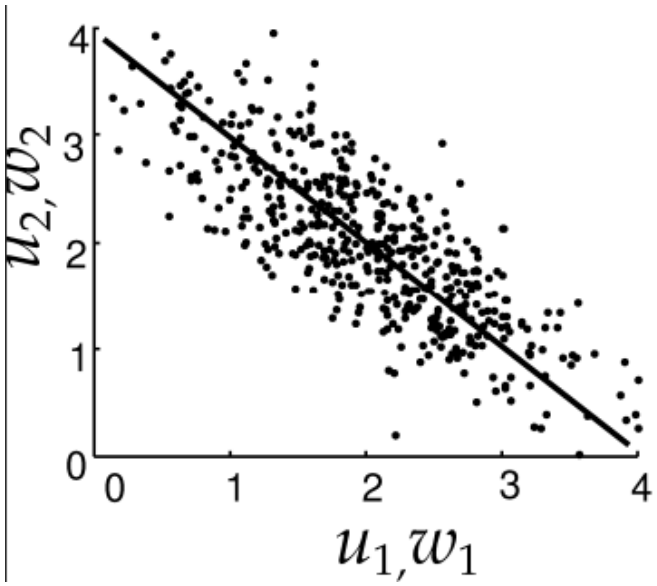
Both involve introduction of a threshold parameter that determines the level of presynaptic or postsynaptic activity at which LTD becomes LTP. The threshold is usually set to the average presynaptic or postsynaptic activity over the training data.

Substituting the average presynaptic activity into the parameter of the first rule, we have:

$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (v - \theta_v)\tilde{u} \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (\tilde{u} \cdot \tilde{w} - \langle \tilde{u} \cdot \tilde{w} \rangle)\tilde{u} \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (\tilde{u} - \langle \tilde{u} \rangle) \cdot \tilde{w}\tilde{u} \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (\tilde{u} - \langle \tilde{u} \rangle)\tilde{w}^T\tilde{u} \rangle$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (\tilde{u} - \langle \tilde{u} \rangle)\tilde{u}^T \rangle \tilde{w}$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = (\langle \tilde{u}\tilde{u}^T \rangle - \langle \tilde{u} \rangle^2)\tilde{w}$$
$$\tau_w \frac{d\tilde{w}}{d\tau} = \langle (\tilde{u} - \langle \tilde{u} \rangle)(\tilde{u} - \langle \tilde{u} \rangle)^T \rangle \tilde{w}$$

$$\tau_w \frac{d\tilde{w}}{d\tau} = C\tilde{w}$$

Where C is the input correlation matrix. Thus, covariance learning produces weight vectors which are parallel to the principal eigenvector of the covariance matrix, instead of the correlation matrix as for the standard Hebbian learning. This is shown in the diagram below.



Note that the covariance learning rule is still unstable, as the weight magnitude changes as:

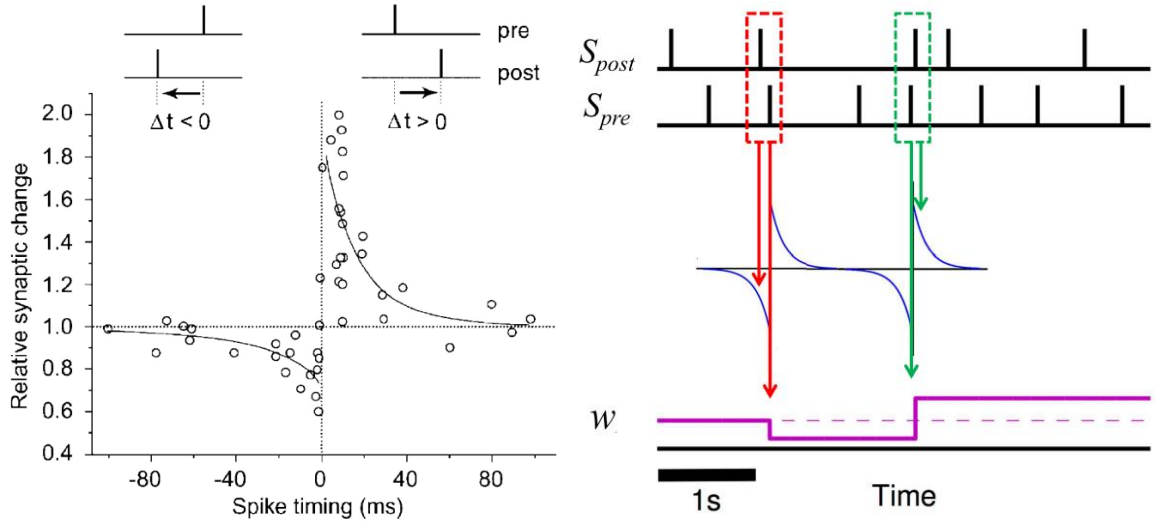
$$\begin{aligned}
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= \tau_w \frac{d(\tilde{w})}{d\tau} \cdot \tilde{w} + \tau_w \tilde{w} \cdot \frac{d(\tilde{w})}{d\tau} \\
&= (v - \langle v \rangle)\tilde{u} \cdot \tilde{w} + \tilde{w} \cdot (v - \langle v \rangle)\tilde{u} \\
&= (v - \langle v \rangle)\tilde{u} \cdot \tilde{w} + (v - \langle v \rangle)\tilde{u} \cdot \tilde{w} \\
&= 2(v - \langle v \rangle)v \\
\tau_w \frac{d|\tilde{w}|^2}{d\tau} &= 2v(v - \langle v \rangle)
\end{aligned}$$

We see that the magnitude of weights increases so long as activities are non-zero.

Lecture 15

Spike-Timing Dependent Plasticity

In Hebbian learning, a synapse is strengthened if pre- and postsynaptic neurons are simultaneously active. However when we consider spiking models, we need to consider the relative timing of spikes in addition to the overall rate. This leads to the idea of Spike-Timing Dependent Plasticity (STDP), in which the timing of the presynaptic relative to the postsynaptic spike determines the change in synaptic weight. Experimental data for this effect is shown below.



STDP Learning Rule

The most general equation for the weight changes up to the quadratic level is:

$$\Delta w_{ij} = c_{ij} v_i v_j + c_{ii} v_i^2 + c_{jj} v_j^2 + b_i v_i + b_j v_j + a$$

Where v_i is the presynaptic activity rate, v_j is the postsynaptic activity rate, and c, b, a are constants.

The change in weight after an experiment of duration T is then given by:

$$\Delta w_{ij} = \int_0^T \int_0^T c_{ij} S_i(t) S_j(t')\, dt\, dt' + \int_0^T b_i S_i(t)\, dt + \int_0^T b_j S_j(t')\, dt' + \int_0^T a\, dt$$

In setting a value for c_{ij}, we assume that a change in synaptic weights only occurs if both spikes occur within some time interval:

$$|t_i - t_j| < s$$

We then define a window function $c_{ij} = W(s)$ that determines how the synaptic weight changes as a function of s. The shape of W defines how the synaptic change depends on the time difference.

Substituting in this function for c_{ij} and dividing through by T we have:

$$\frac{\Delta w_{ij}}{T} = \frac{1}{T}\int_0^T \int_0^T W(t - t') S_i(t) S_j(t')\, dt\, dt' + \frac{b_i}{T}\int_0^T S_i(t)\, dt + \frac{b_j}{T}\int_0^T S_j(t')\, dt' + a$$

$$\frac{\Delta w_{ij}}{T} = \frac{1}{T}\int_0^T \int_{-t'}^{T-t'} W(s) S_i(s + t') S_j(t')\, ds\, dt' + \frac{b_i}{T}\int_0^T S_i(t)\, dt + \frac{b_j}{T}\int_0^T S_j(t')\, dt' + a$$

$$\frac{1}{T}\int_0^T \frac{dw_{ij}(t)}{dt} dt = \frac{1}{T}\int_0^T \int_{-t}^{T-t} W(s) S_i(t + s) S_j(t)\, ds\, dt + \frac{b_i}{T}\int_0^T S_i(t)\, dt + \frac{b_j}{T}\int_0^T S_j(t)\, dt + a$$

If $T \gg s$, we can replace the upper integral terminal with T. Then extending the integral out to infinity and negative infinity, which is valid since far outside the window (in the negative direction in particular) no weight changes will occur anyway, we can write:

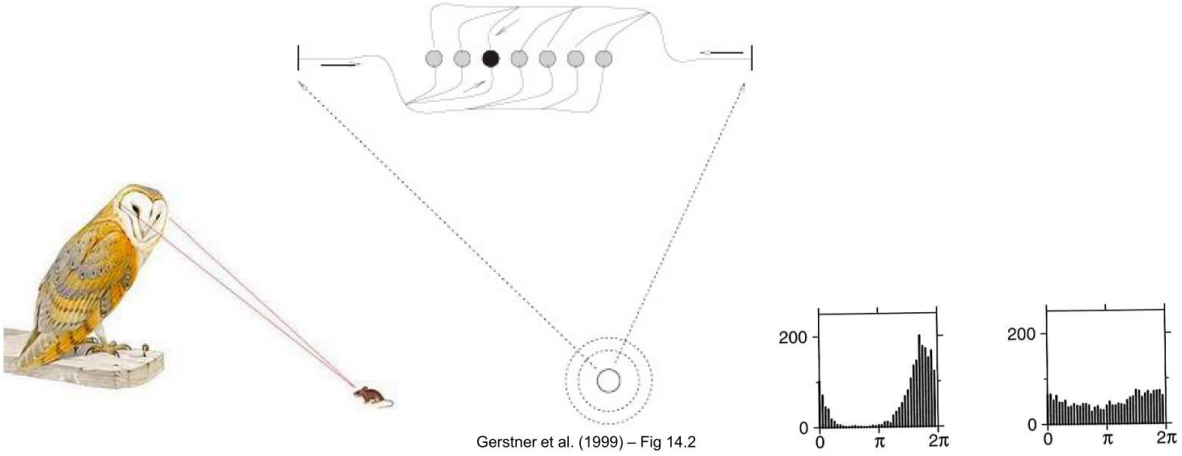
$$\frac{1}{T}\int_0^T \frac{dw_{ij}(t)}{dt}\,dt = \frac{1}{T}\int_{-\infty}^{\infty}\int_{t'}^{T-t'} W(s)S_i(t+s)S_j(t)\,ds\,dt + \frac{b_i}{T}\int_0^T S_i(t)\,dt + \frac{b_j}{T}\int_0^T S_j(t)\,dt + a$$

$$\left\langle \frac{d}{dt}w_{ij}(t)\right\rangle = \int_{-\infty}^{\infty} W(s)\left\langle S_i(t+s)S_j(t)\right\rangle ds + b_i\langle S_i(t)\rangle + b_j\left\langle S_j(t)\right\rangle + a$$

We see that learning is driven now by temporal correlations in the spiking times of the presynaptic and postsynaptic neurons.

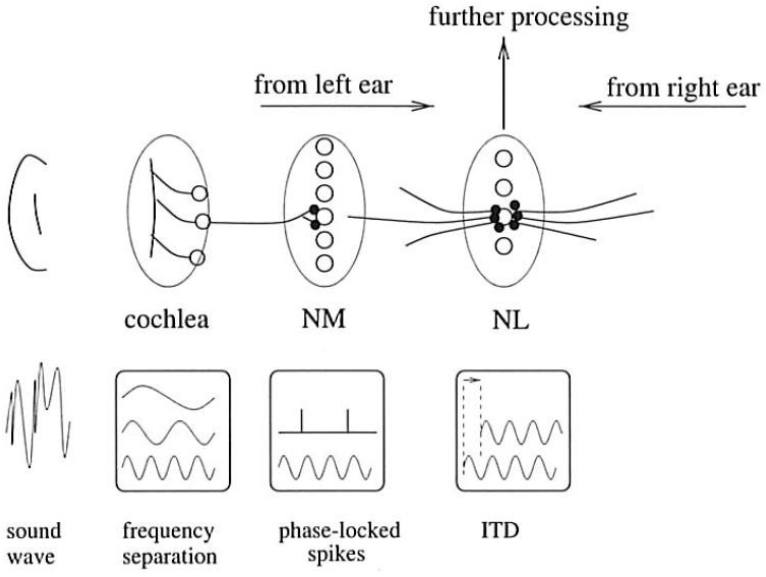
Sound Localisation

Barn owls can use interaural time differences (ITDs) for sound source localisation to locate prey even in complete darkness. By this method, they can localise sound to within 1-2 degrees, which corresponds to temporal arrival differences between the ears of less then $5\mu s$. This capability is not innate, but learned by connections being strengthened and pruned during development.



Gerstner et al. (1999) – Fig 14.2

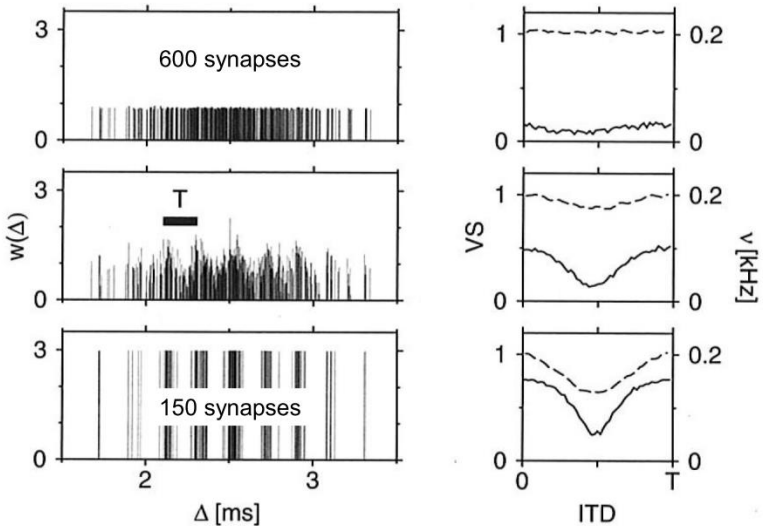
At each ear, sounds are separated into their frequency components through phase locking. This is facilitated by unusually low membrane time constants of the neurons involved, around $\tau = 0.1ms$ compared to more typical values of $\tau = 10ms$. Signals are then passed through the nucleus magnocellularis (NM) to the nuclear laminaris (NL), where interaural time differences are computed, which for a simple sinusoidal signal will be equal to the phase difference.



A neural model of this process has been developed by Gerstner et al, based on an integrate-and-fire neuron with exponential current input as a function of input spikes from j input neurons, which in turn are generated from a Poisson process with sinusoidal firing rate. The resulting input can be approximated by a sum of Gaussians. The model then consists of the following equations:

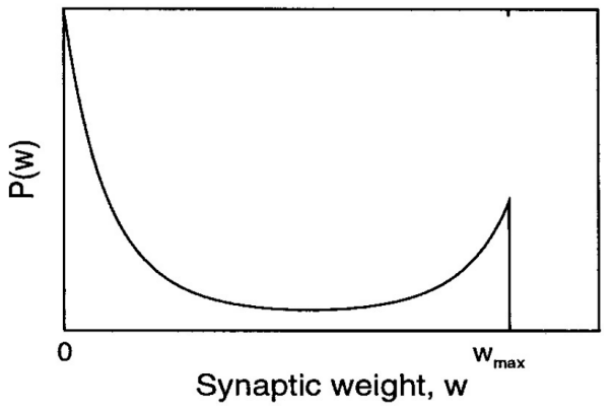
$$\tau_m \frac{du}{dt} = -u + RI(t)$$
$$I(t) = \sum_j \frac{w_j}{R/\tau_m} \sum_{t_j} \frac{1}{\tau_s} \exp\left(-\frac{t - t_j}{\tau_s}\right) H(t - t_j)$$
$$t_j \sim Po(\sin(\sigma t))$$

This model adjusted weights using STDP, and was able to learn synaptic weights that are sensitive to the periodic structure of the input. This is shown below, with the weights plotted against the interaural time difference before (top), during (middle), and after learning (bottom).The dashes lines on the right show that after learning, the neurons have the highest response rate when the time difference is zero, and lowest when it is equal to have a period. Thus they serve as a coincidence detector.



Additive vs Multiplicative STDP

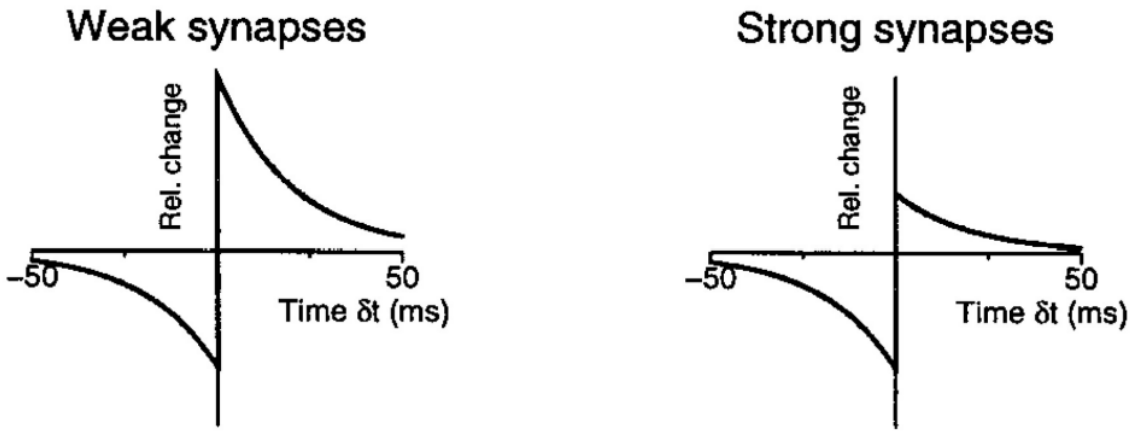
STDP is competitive, since changing the strength of one synapse will shift spike timing of the corresponding neuron, which would then affect the synaptic strengths of the other synapses. Additive STDP tends to result in a bimodal distribution of synaptic weights, with weights driven to the extremes.



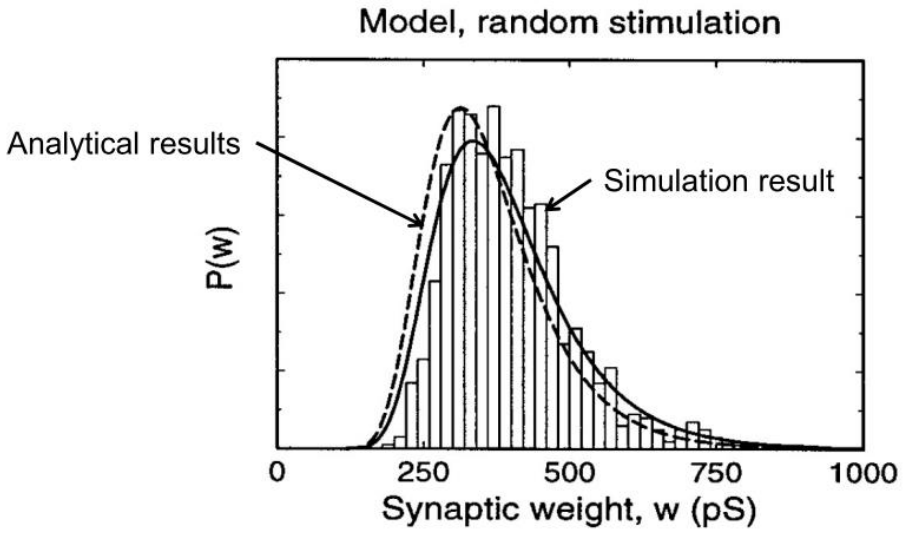
In Additive STDP, correlated groups of synapses tend to drive the output, and therefore tend to potentiate one another in comparison to uncorrelated groups of synapses. An alternative model is multiplicative STDP. The difference between additive and multiplicative STDP are shown in the equations below:

$$\Delta w_{+} = c \exp(-\delta t/\tau)$$
$$\Delta w_{\times} = cw \exp(-\delta t/\tau)$$

The multiplicative term results in a balancing of synapses, so that strong synapses are weakened, and weak synapses are strengthened. Because of this effect, neurons trained using multiplicative STDP tend to form a unimodal weight distribution.



This form of learning produces weight distributions like those observed experimentally. It is also stable without a need to set a maximum weight value. Competition can be reintroduced using a mechanism such as activity-dependent synaptic scaling.

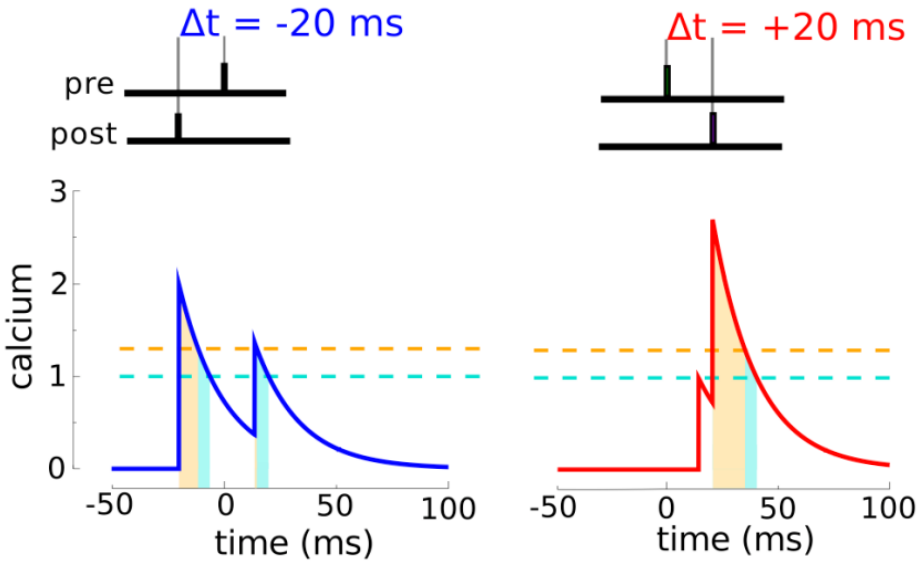


Lecture 16

Calcium Model of STDP

Biophysical models focus on the underlying mechanisms involved in producing an observed phenomenon. They are distinct from phenomenological models (e.g., STDP), which simply try to describe the observed phenomenon. Most of these are based on the role of calcium.

In this model, there is a spike in calcium whenever there is a presynaptic or postsynaptic action potential, but the spike is larger in the case of a postsynaptic action potential. In either case, the calcium concentration then decays exponentially. What matters in this model is the proportion of time that the cell is above the calcium threshold for potentiation (α_p, threshold shown in dotted orange), compared to the proportion of the time it is above the calcium threshold for depression (α_d, threshold shown in dotted green). As shown in the bottom graph, α_d is relatively higher when the postsynaptic spike occurs before the presynaptic spike (because the large first spike bumps up the second to be in the depression range), while α_p is relatively higher in the reverse case (because the small first spike does nothing except raise up the second spike so it spends longer above the potentiation threshold). This leads to depression when the postsynaptic spike occurs first, and potentiation when the postsynaptic spike occurs second. Note that there is a delay D between a presynaptic spike and the onset of the corresponding calcium spike, but no such delay in the case of a postynaptic spike.



The calcium dynamics $c(t)$ are given by the equation:

$$\frac{dc}{dt} = -\frac{c}{\tau} + C_{pre}\sum_i \delta(t - t_i - D) + C_{post}\sum_j \delta(t - t_j)$$

There t_i are times of presynaptic spikes and t_j are times of postsynaptic spikes. This calcium behaviour gives rise to the values α_p and α_d, which are the fractions of time the calcium transient is above the potentiating and depressing thresholds θ_x, calculated over total stimulation time T as:

$$\alpha_x = \frac{1}{T}\int_0^T H(c(t) - \theta_x)\, dt$$

The behaviour of the synaptic efficiency ρ is then modelled by the equation:

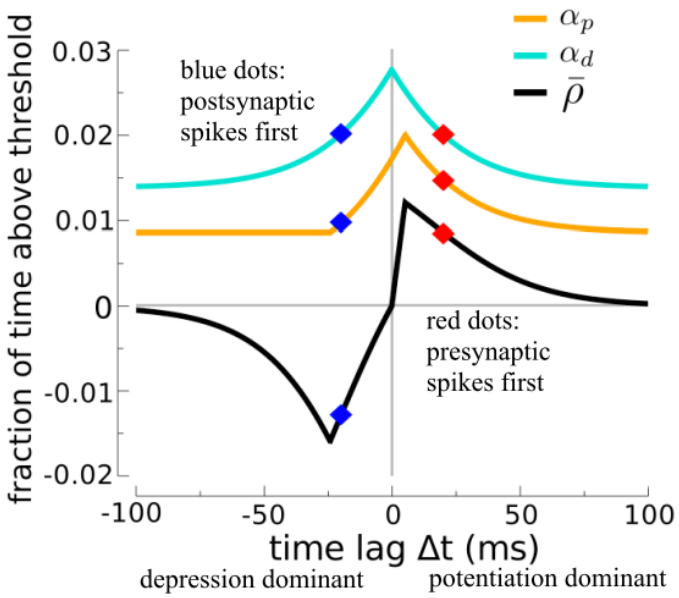
$$\tau\frac{d\rho(t)}{dt} = \alpha_p\gamma_p\big(1 - \rho(t)\big) - \alpha_d\gamma_d\rho(t) + \sqrt{\tau}\sqrt{\alpha_p + \alpha_d}z(t)$$

Where $z(t)$ is a Gaussian noise term with standard deviation σ and zero mean.

To find the long-term equilibrium for a given set of spikes $\bar{\rho}$, we ignore the noise term and set the derivative to zero:

$$\begin{aligned}
\alpha_p\gamma_p(1-\bar{\rho}) &= \alpha_d\gamma_d\bar{\rho} \\
\alpha_p\gamma_p - \alpha_p\gamma_p\bar{\rho} &= \alpha_d\gamma_d\bar{\rho} \\
\alpha_p\gamma_p &= \alpha_d\gamma_d\bar{\rho} + \alpha_p\gamma_p\bar{\rho} \\
\alpha_p\gamma_p &= \bar{\rho}(\alpha_d\gamma_d + \alpha_p\gamma_p) \\
\bar{\rho} &= \frac{\alpha_p\gamma_p}{\alpha_d\gamma_d + \alpha_p\gamma_p}
\end{aligned}$$

The effect of varying the α parameters is shown in the figure below (note that $\bar{\rho}$ is on a different scale to the axis on the left).



Note that STDP is still a simplification, as it does not incorporate effects such as:

1. Fire rate dependence
2. Spike-triplets and spike-quadruplets
3. Bursts
4. Dendritic location dependence

Unsupervised Learning

Unsupervised learning is a type of Hebbian learning that finds previously unknown patterns in a data set (without requiring pre-existing labels). It usually performs principal component analysis (or similar) through an algorithm that involves variance maximisation.

Supervised Learning

Supervised learning involves presenting labelled data to a network, which then learns weights in accordance with a mechanism that optimises some predefined error function. Effectively this involves learning to approximate some function g of the input $\tilde{x}$ with a parameterised approximation function $G(\tilde{w}, \tilde{x})$. The approximation algorithm then attempts to find the optimal weights $\tilde{w}^*$ that reduce the error until it is below some threshold ϵ, relative to some norm $|\cdot|$. This is represented as:

$$|G(\tilde{w}^*, \tilde{x}) - g(\tilde{x})| \leq \epsilon$$

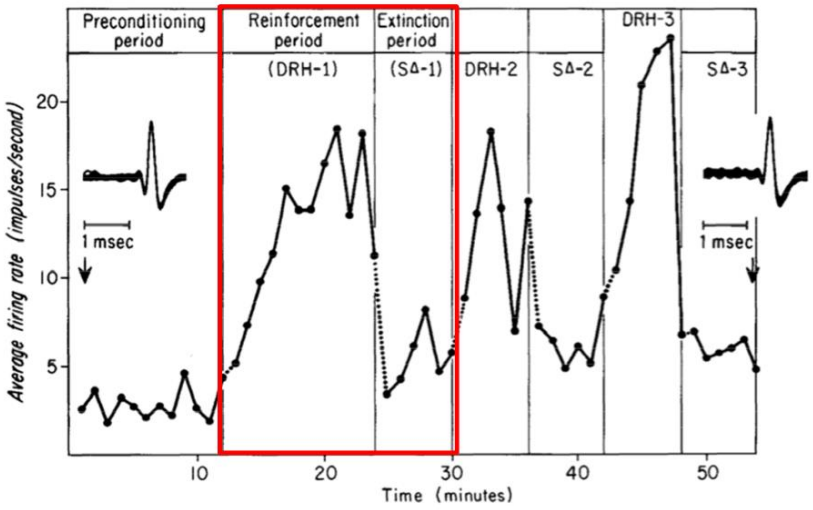
In neural network applications, the following parameterisations are often made:

- The approximation function G is often sigmoidal.
- The norm $|\cdot|$ is often Euclidean.
- The search algorithm for $\tilde{w}^*$ is typically some form of gradient descent.

For supervised learning, the approximation function needs to provide for universality, the ability of $G(\tilde{w}, \tilde{x})$ to represent $g(x)$ accurately; and generalisation, the ability of $G(\tilde{w}, \tilde{x})$ to correctly map new points not seen during the learning process.

Reinforcement Learning

Reinforcement learning involves interacting with an environment to receive a learning signal, which then drives learning. Not only animal behaviours can be reinforced, but even individual neurons can be reinforced to yield increased firing rates (see example experimental data below).

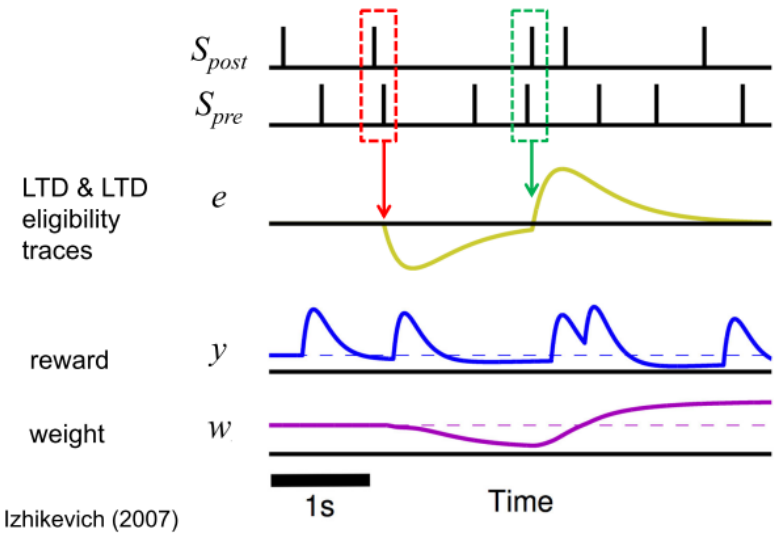


Reward-Modulated STDP

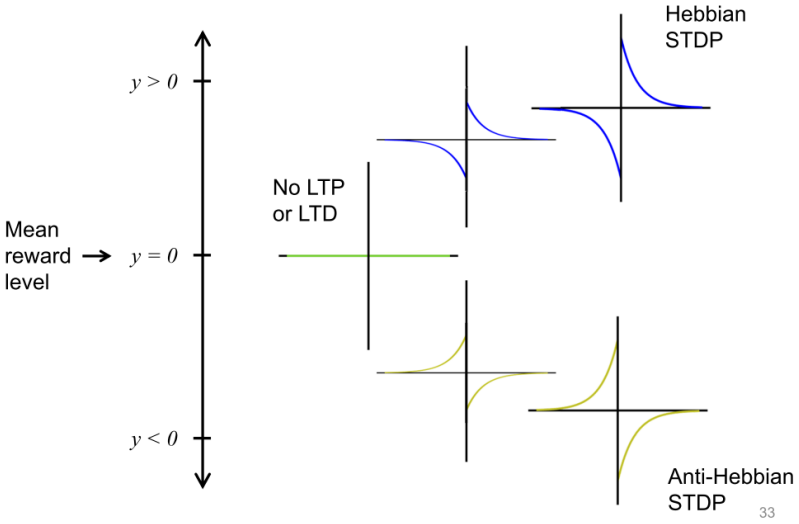
Spike time dependent plasticity in its ordinary form is an unsupervised form of learning, as synaptic changes are not based on any teaching or reward signal. As such, it would have to be modified to be used for reinforcement learning. To achieve this we need a way of solving the 'credit assignment problem', which is how to preserve a memory of the stimulus so that by the time a reward signal is received, the change in network weights can be conditioned on the stimulus.

One solution is to use a mechanism called eligibility traces. In this mechanism, the eligibility function integrates signals from the presynaptic and postsynaptic spikes, and then computes the changes in weights by multiplying the eligibility trace by the reward signal:

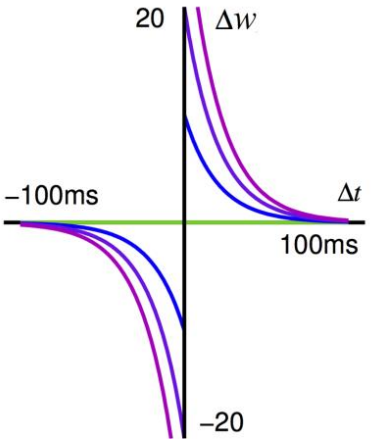
$$\Delta w = e \times y$$

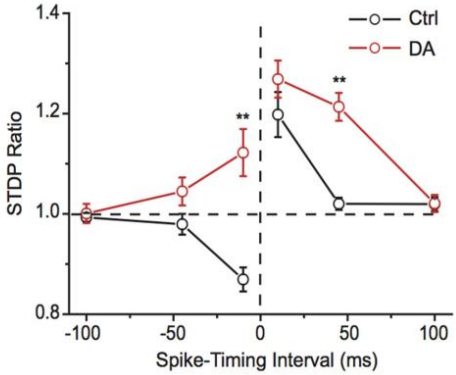


This mechanism can even generate anti-Hebbian learning when the reward signal is negative.

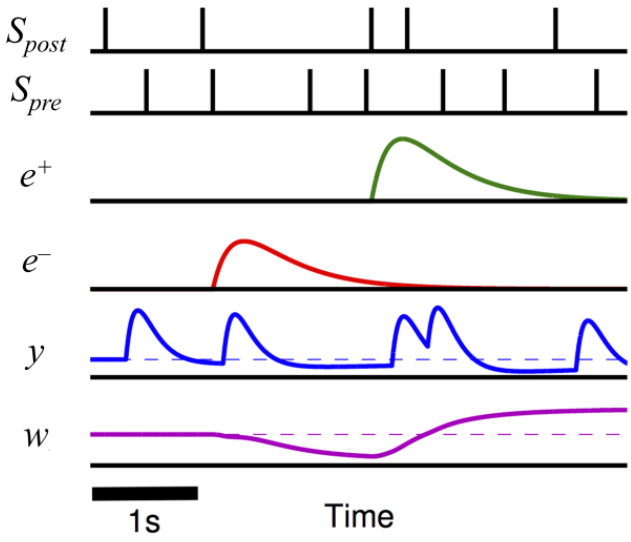


Reward signals are often carried by dopamine (DA), which functions as both a neurotransmitter and also a neuromodulator. Neurons of the ventral tegmental area (VTA) and substantia nigra respond to rewarding stimuli, projecting dopamine to many other brain regions. As shown below, this can positively modulate the extent of STDP.





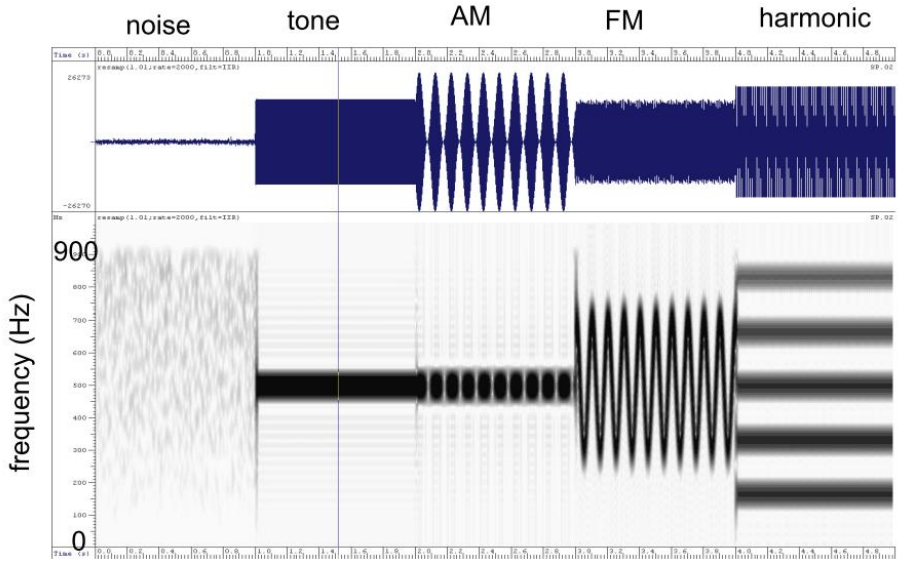
We can explain the flip in the sign of the STD depression shown in the diagram above (in the presence of dopamine (DA)) if we introduce two separate eligibility traces, one for depression and one for potentiation.



Brain Systems

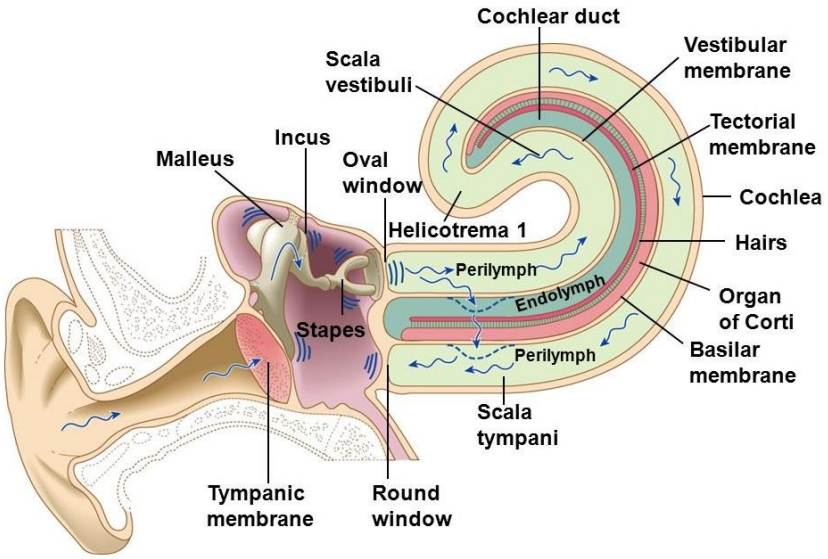
Lecture 17

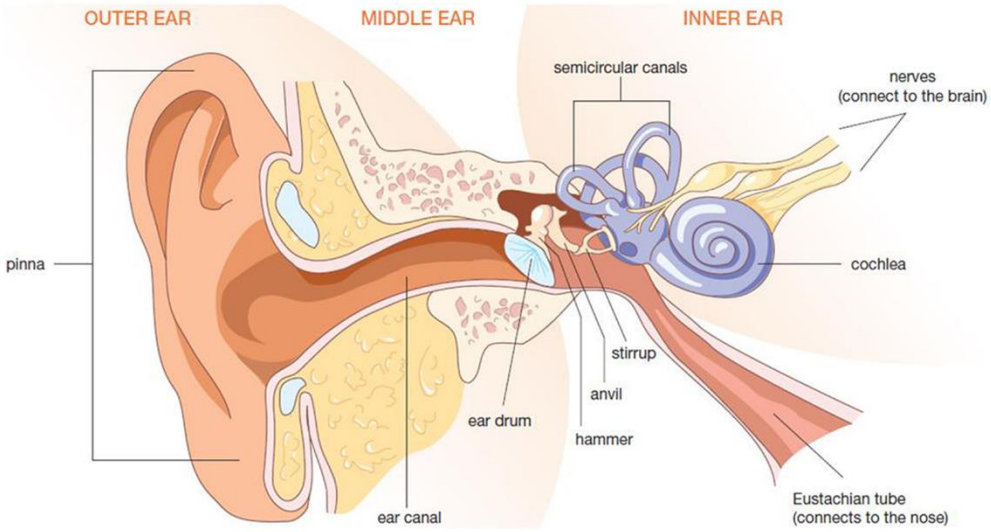
Acoustics

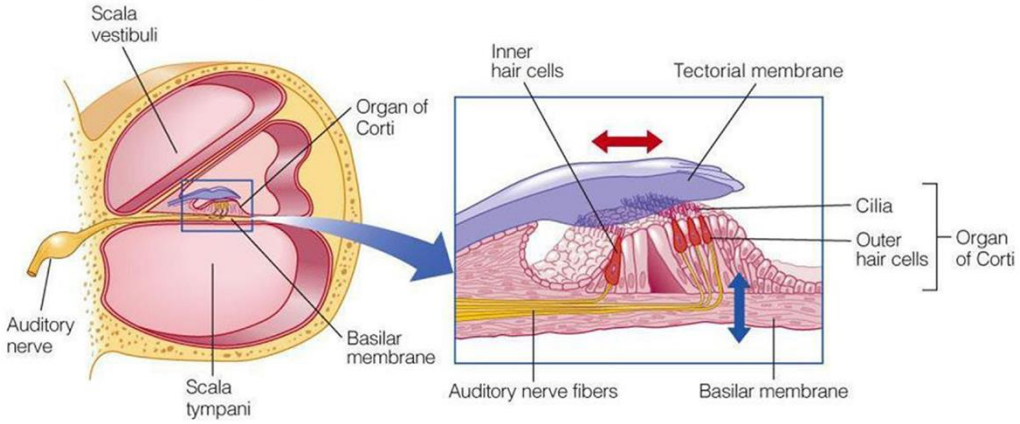


Ear Anatomy

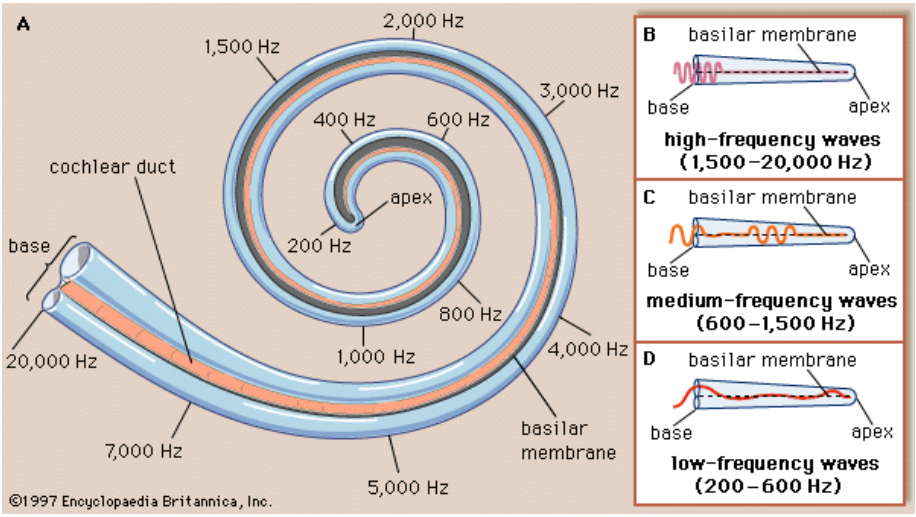
The ear is comprised of an outer, middle, and inner ear. The outer ear consists of the pinna, auditory canal, and tympanic membrane, which separates outer and middle ear. The middle ear consists of three small bones (called the malleus (hammer), incus (anvil) and stapes (stirrup)), which transmit vibrations from the tympanic membrane to the oval window (also called the ear drum). The inner ear consists of fluid filled chambers including semicircular canals (equilibrium) and cochlea (hearing). Within the cochlea is the organ of corti, which is responsible for the sound transduction process.



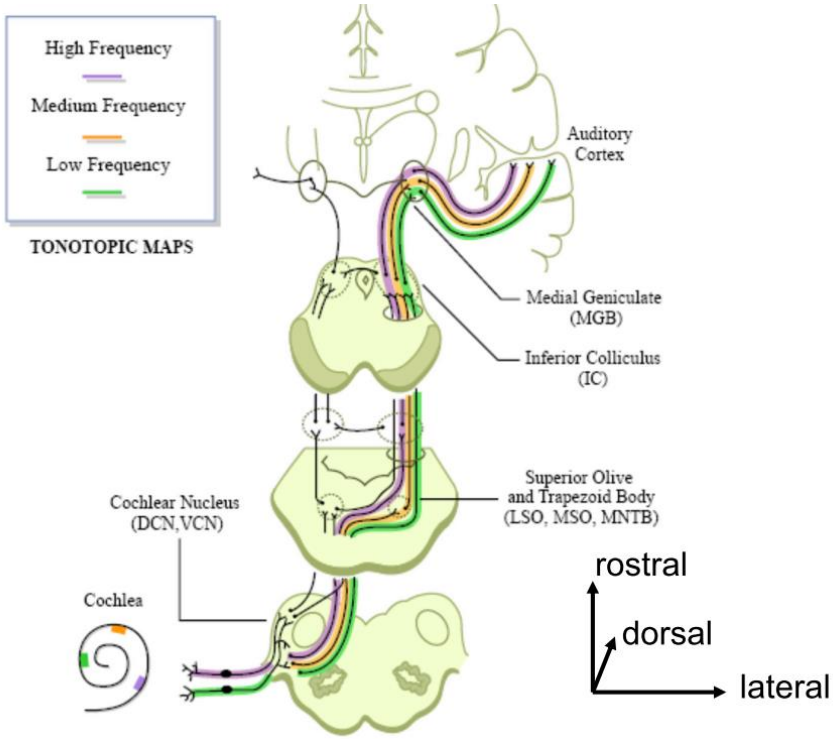




Higher frequency sounds are detected near the beast of the basilar membrane, while low frequency sounds are detected near its apex. This is a form of spatial coding.

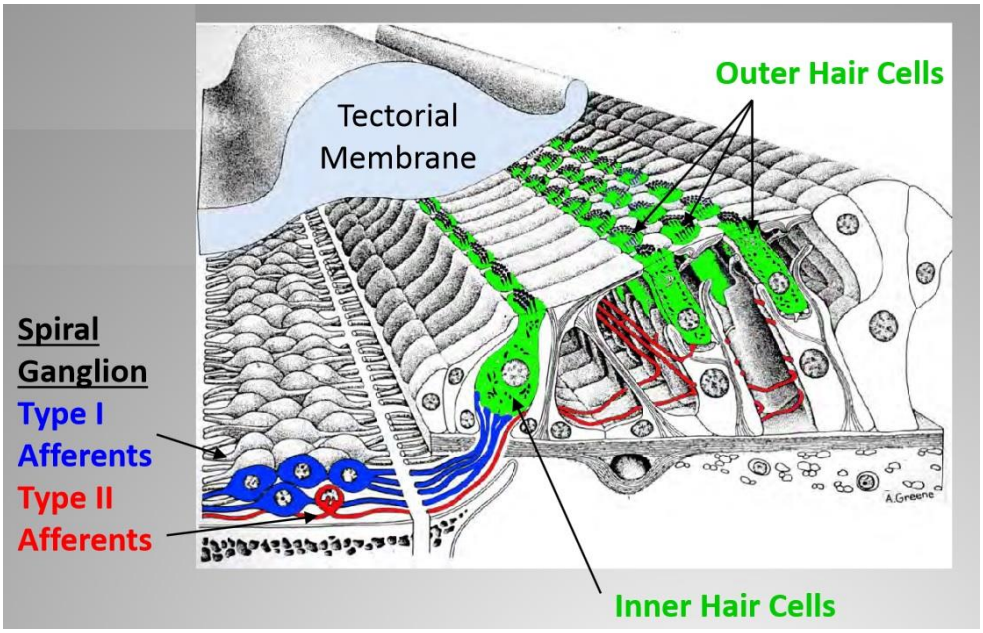


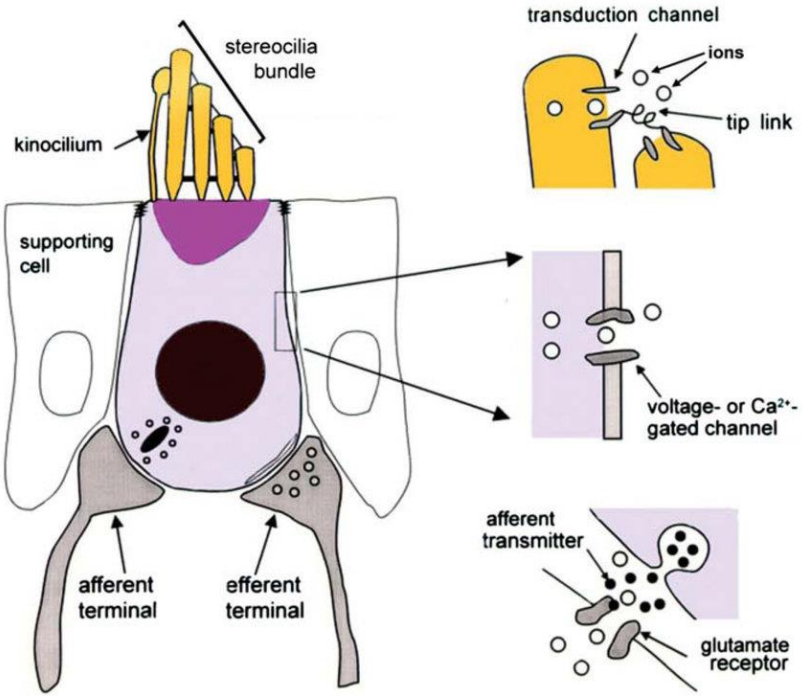
This tonotopic mapping is maintained throughout the auditory pathways.



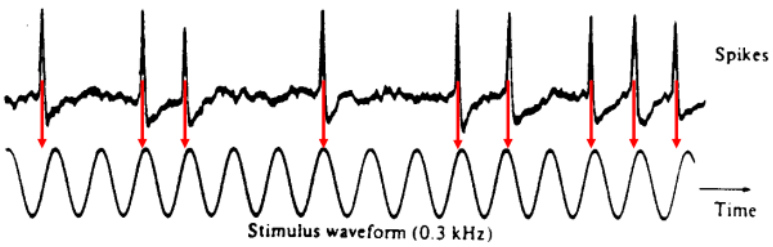
Hair Cells

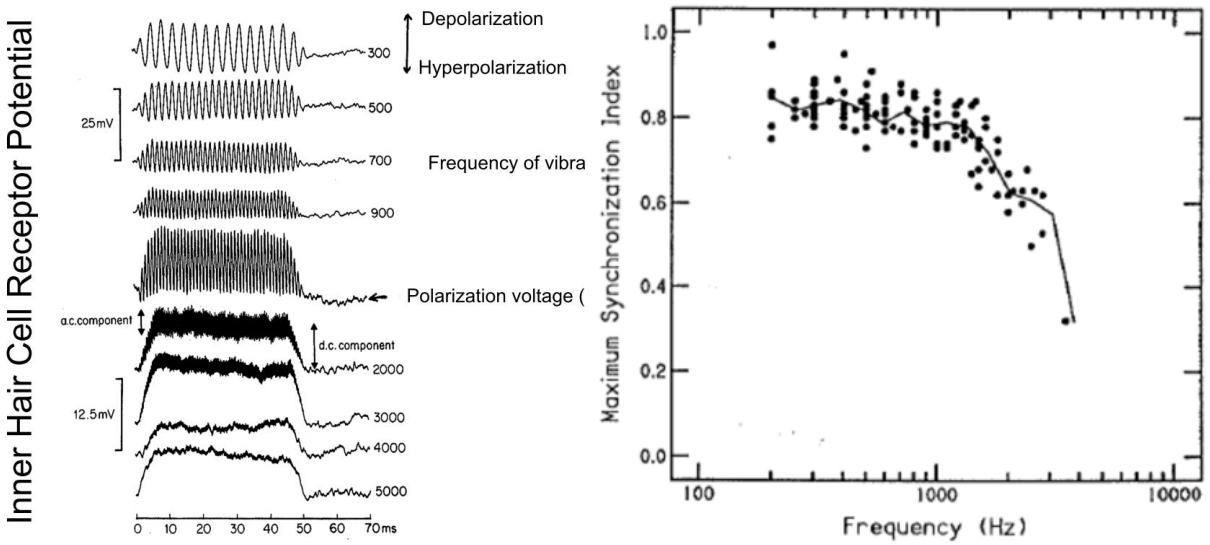
Hair cells are embedded in the basilar membrane and project small cilia into the overlying tectorial membrane. When sound vibrates the perilymph in the scala vestbuli and scala tympani, the basilar membrane vibrates relative to the tectorial membrane. This causes the cilia of the hair cells to move relative to each other, opening mechanically gated ion channels. This depolarises the membrane, which then triggers the opening of voltage or calcium gated channels, further depolarising the membrane. If threshold is reached, neurotransmitter is released, triggering action potentials in efferent neurons.





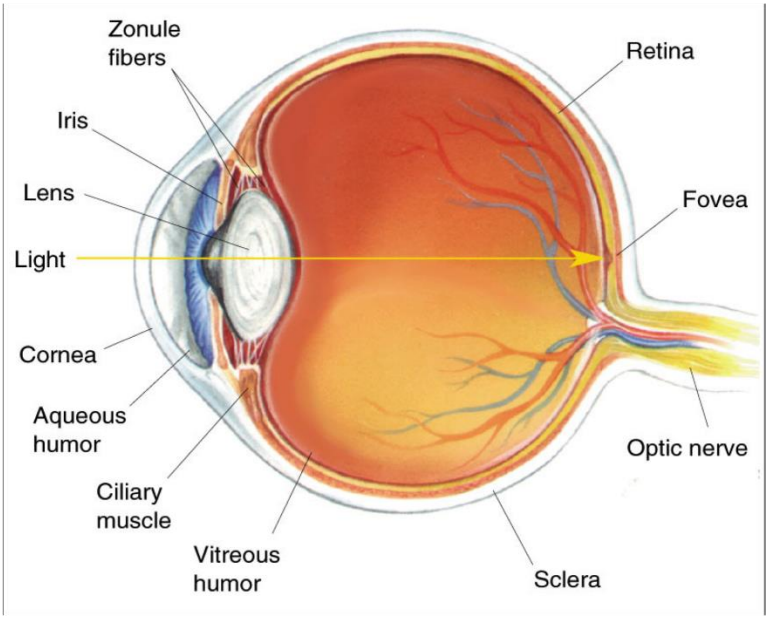
At lower frequencies, the response of hair cells is phase locked to the input stimulus. This becomes impossible at high frequencies owing to the membrane time constant.





Eye Anatomy

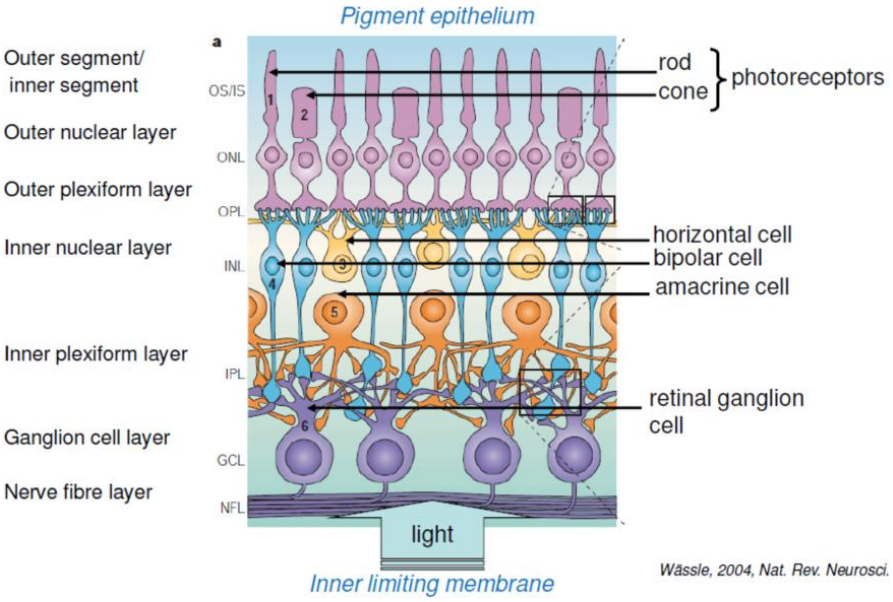
Vertebrate eyes consist of: the sclera, tough white outer connective tissue; the cornea, a clear part of sclera in the front of the eye which allows the light in and acts as a fixed lens; the iris, a pigmented inner layer that can change size to regulate the amount of light coming in; the lens itself; the retina, where the photoreceptors are located at the back of the eye; aqueous humour, a fluid that fills the anterior cavity; and vitreous humour, a jellylike fluid that fills the posterior chamber.

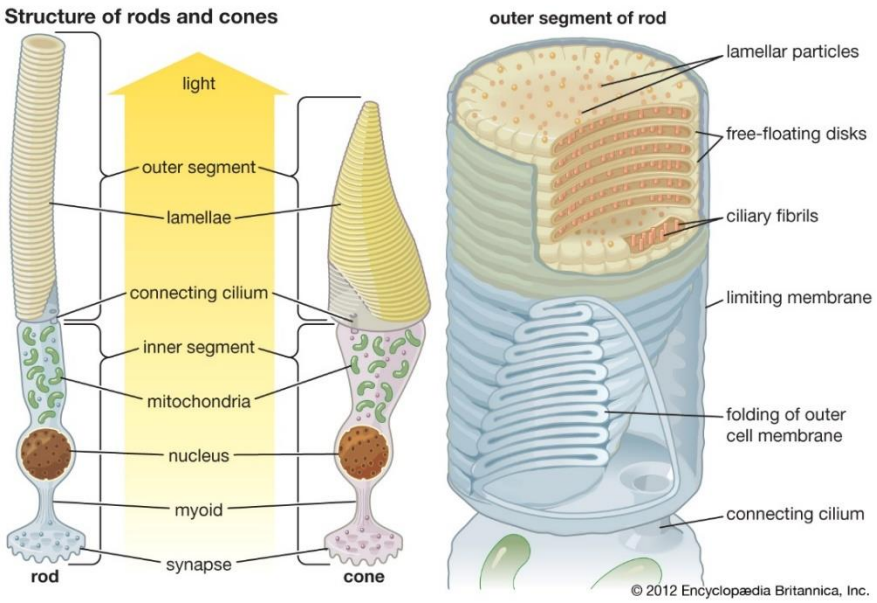


Photoreceptors

The photoreceptors lining the retina consist of rods and cones. Each has stacks of membrane discs containing rhodopsin (a vitamin A derivative + opsin). These are activated and cause sensory transduction in response to incident photons.

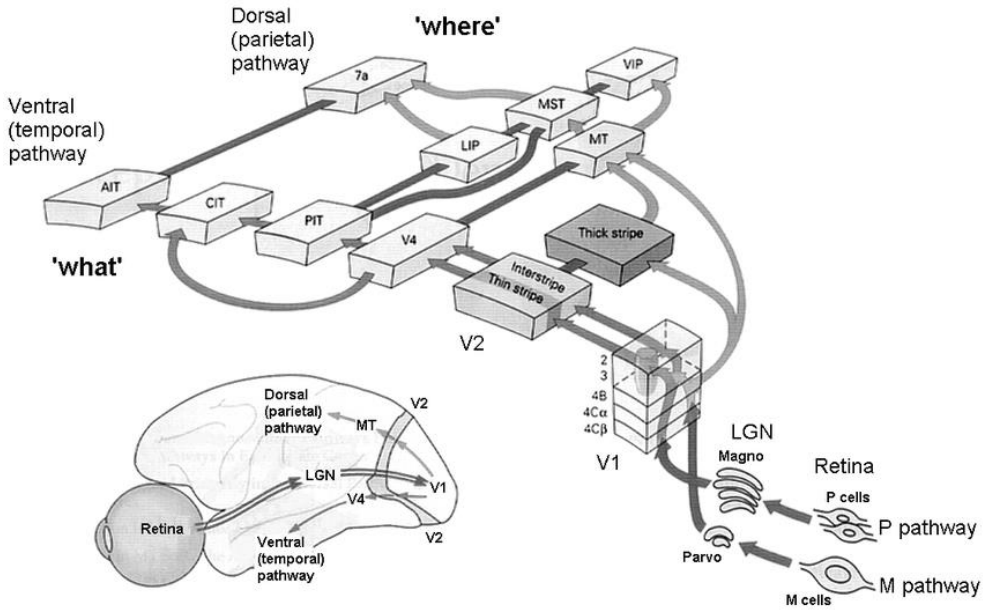
Laminar organisation





Cell type	Location	Description
Centre-surround cells	RGC and LGN	Concentric circles where the stimulus is excitatory and inhibitory respectively.
Simple cells	V1	Responds primarily to oriented edges and gratings (bars of particular orientations), in a particular location in visual field.
Complex cells	V1/V2	Responds primarily to oriented edges and gratings, regardless of exact location in visual field. Some respond optimally only to movement in a certain direction.
Hyper-complex cells	V2/V+	Responds to oriented edges and gratings regardless of exact location, but also sensitive to the length of the lines. This is called edge-stopping.

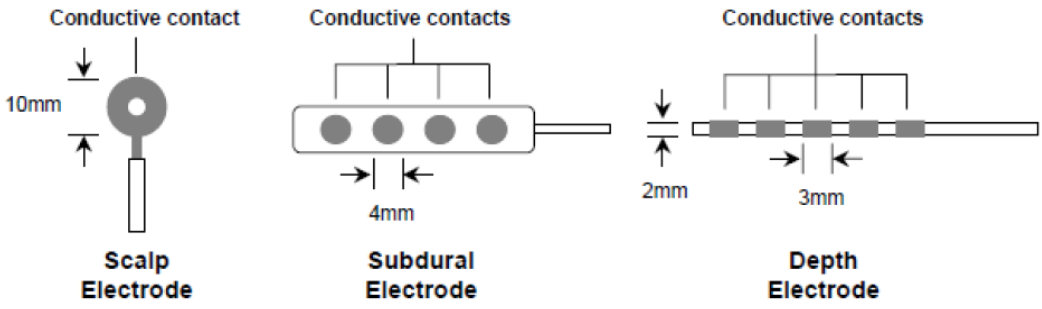
The visual system consists of two main pathways: ventral ('what') and dorsal ('where').



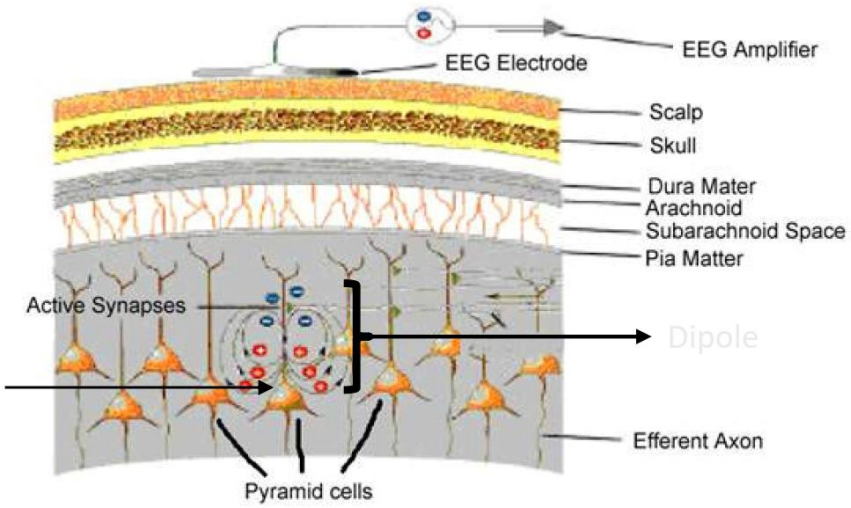
Lecture 18

Introduction to EEG

Electroencephalography (EEG) measures the time evolution of the electric potential (voltage) generated by the brain. EEG can be measured on the scape, on the surface of the cortex, or in deeper parts of the brain. The electrode size must vary in accordance with the location of the electrodes.

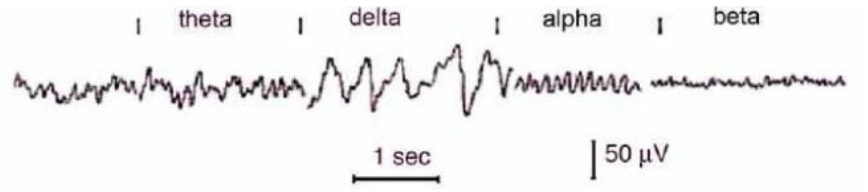


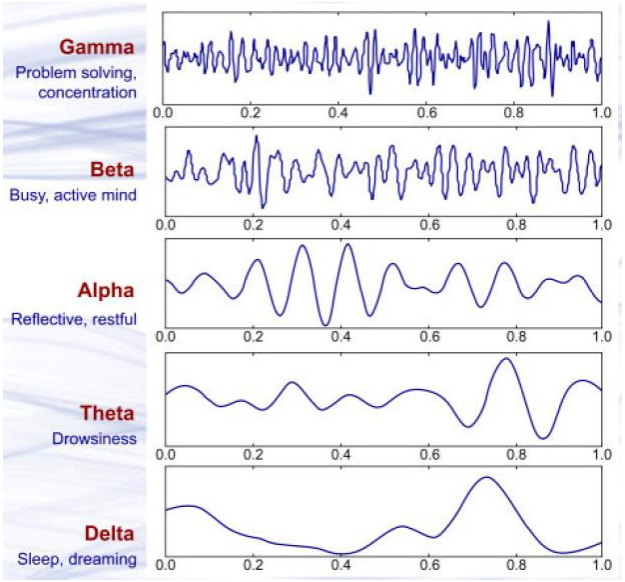
There are many layers of insulation between the electrodes and the neural tissue.



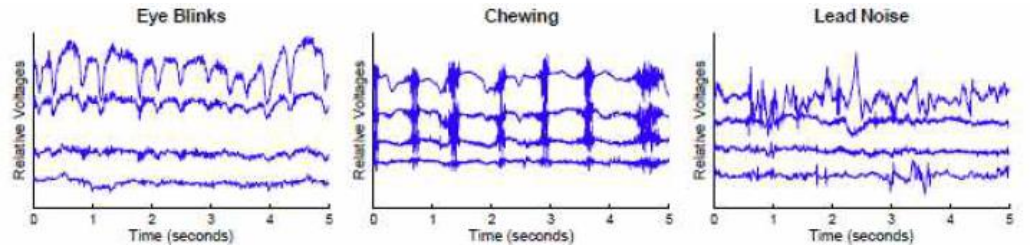
A number of different waveform patterns have been identified in EEG:

- Delta: less than 3 Hz. Dominant rhythm in infants and in stages 3 and 4 of sleep.
- Theta: 3.5 – 7.5 Hz. Slow activity.
- Alpha: 7.5 – 13 Hz. Appears when closing the eyes and relaxing.
- Beta: 14 – 20 Hz. Fast activity. Dominant rhythm when eyes are open.
- Gamma: 20-100 Hz. Faster activity.



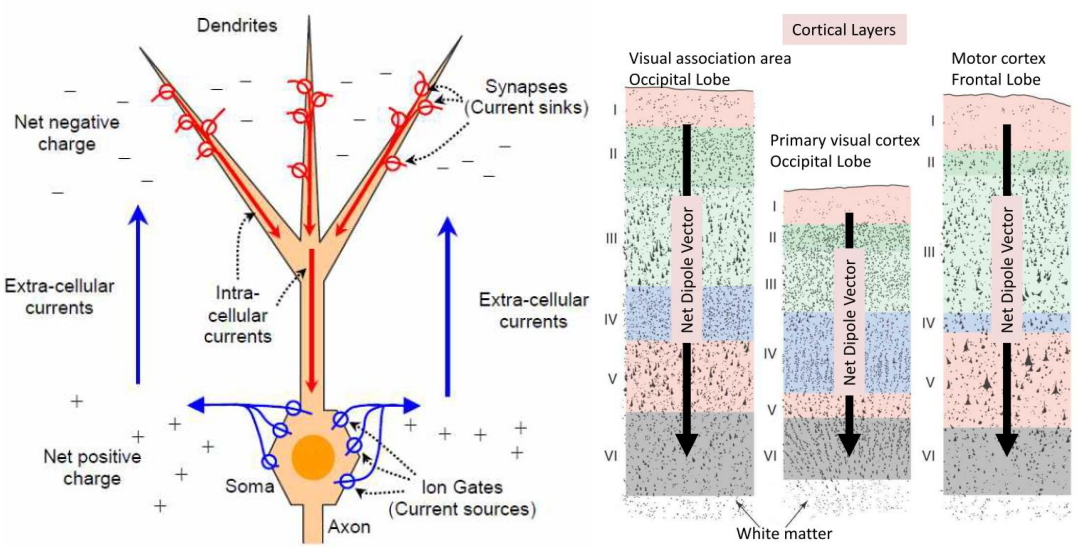


Several sources of noise make interpretation of EEG difficult. These include eye blinking, chewing, and movement of the electrical leads.



Modelling the Electric Potential

In biology, current is transmitted by ions rather than electrons. Most of these currents flow within the dendrites or axon of the cell as in action potential transmission, however since cell membranes are highly resistive, intracellular currents do not contribute to the EEG. Instead, brain electric fields are generated by chemical currents of charged ions flowing outside the cell.



The electric field generated by a particle of charge q at location r_s is:

$$E(r) = \frac{q}{4\pi\epsilon_0|r - r_s|}\tilde{r}$$

The electric potential of a single charge is:

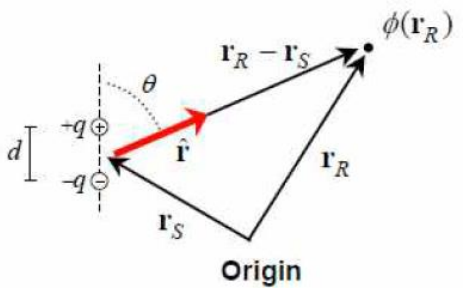
$$\phi(r) = \frac{q}{4\pi\epsilon_0 R}$$

The electric potential generated by N charges is:

$$\phi(r) = \sum_{n=1}^{N} \phi_N(r) = \frac{1}{4\pi\epsilon_0}\sum_{n=1}^{N}\frac{q_n}{R_N}$$

If there are only two charges, this equation simplifies to that of a dipole:

$$\phi(r) = \frac{q}{4\pi\epsilon_0}\frac{d\cos(\theta)}{R^2}$$



Because the measurement of EEG is far from the sources, a volume containing many charges can be approximated by a single dipole, provided there is roughly an equal number of positive and negative ions within it. This is known as an equivalent dipole and is useful in approximating the activity in small volumes of the brain. However, for equivalent electric dipoles we still need to know a rough distribution of all charges in the brain, so instead we use a current dipole, with current sources and sinks instead of positive and negative charges. The equation for a current dipole is:

$$\phi(d,\theta) \approx \frac{Id}{4\pi\sigma}\frac{\cos(\theta)}{R^2}$$

The net electrical potential at r generated by many small volumes is then approximated by the sume of the equivalent dipole representing each small volume:

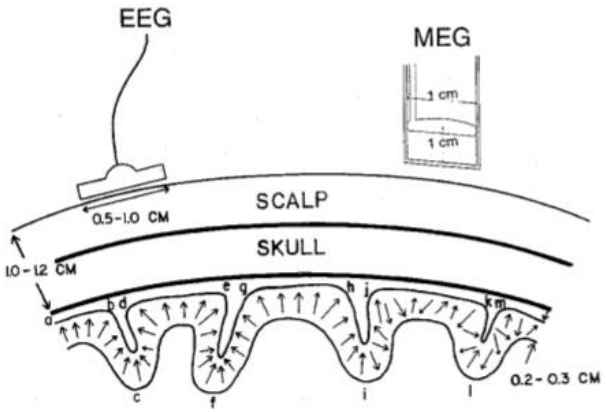
$$\phi_{TOT}(r) = \sum_i \frac{Id_i}{4\pi\sigma}\frac{\cos(\theta_i)}{R^2}G(d_i,\theta_i,r)$$

Here $G(d_i,\theta_i,r)$ or $G(r_s,r_R)$ is the Green's function, and describes the effects of the material between the source r_s and the recording location r_R.

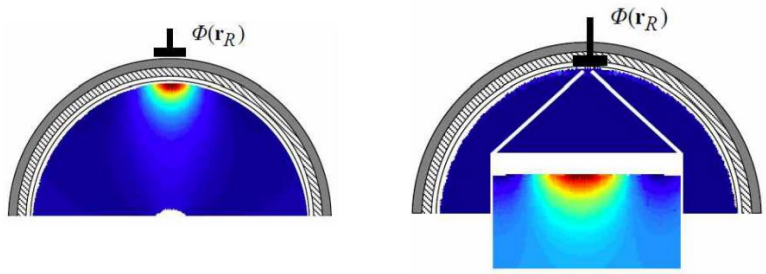
Interpreting EEG Signals

Because the human head is not spherical and is inhomogeneous, the true Green's function is very complex, and must take into account both the different materials in the brain, but also the boundaries between them.

To deal with this complexity approximations are needed. In the brain, the neurons in the cortex are aligned and often large areas are activated together, meaning that potentials in the cortex are much more influential to the EEG than those in deeper structures. Hence, an appropriate simplification is to consider a sheet of dipoles rather than a volume.



Using this model, simulations show that the measured voltages are affected by large areas of cortex. EEG is thus an ambiguous measurement, in that the same EEG signal can be generated by many different patterns or regions of brain activity.



Scalp EEG
10-20cm^2

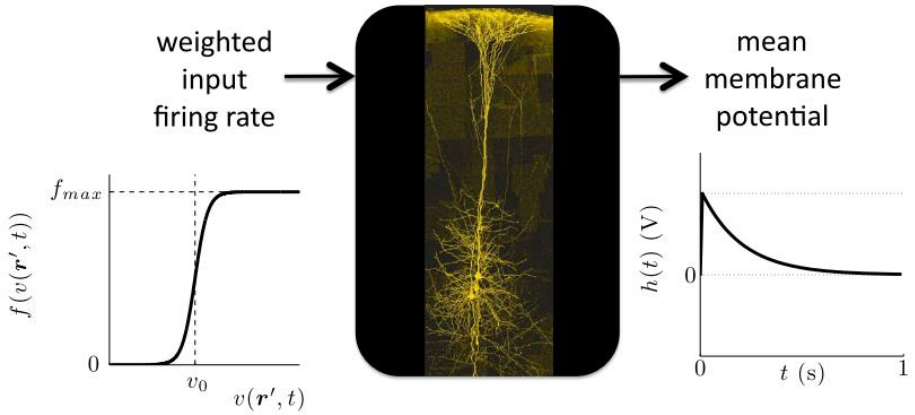
Cortical EEG
4-5mm^2

Lecture 19

Neural Mass Models

Neural mass models describe macro-columns using only one or two state variables to represent the mean activity of the whole population of neurons. This procedure, sometimes referred to as a mean-field approximation, is very efficient for determining the steady-state behaviour of neuronal systems. It is often used to model alpha rhythms.

Mass models consist of two main components: a neural output function $f(v)$, which describes the firing rate of a neuron as a function of its membrane potential v; and a synaptic kernel $h(t)$, which describes the membrane potential as a function of time after threshold v_0 has been reached.



wave to pulse

pulse to wave

Working backwards, the output firing rate of cell two $g_2(t)$ is determined by the sigmoid output function $f(v)$:

$$g_2(t) = f\big(v_2(t)\big)$$

The postsynaptic potential of cell two, $v_2(t)$, is in turn determined by convolving the synaptic kernel cell $h(t)$ with the firing rate of neuron one $g_1(t)$:

$$v_2(t) = h(t) * g_1(t) = \int_{-\infty}^{t} h(t-t')g_1(t')\,dt'$$

Many different synaptic kernel functions can be used:

$$h(t) = H(t)W\exp\left(-\frac{t}{\tau}\right)$$
$$h(t) = H(t)W\frac{t}{\tau}\exp\left(-\frac{t}{\tau}\right)$$
$$h(t) = H(t)\frac{W}{\tau_2 - \tau_1}\left[\exp\left(-\frac{t}{\tau_1}\right) - \exp\left(-\frac{t}{\tau_2}\right)\right]$$

A common form for the firing rate function is:

$$f\big(v_i(t)\big) = \frac{f_{max}}{1 + \exp\Big(a\big(v_0 - v(t)\big)\Big)}$$

Green's Function Methods

These equations can be combined into a single differential equation by taking advantage of the properties of Green's functions. Suppose we have a linear differential operator D and an equation of the form:

$$Dv(x) = f(x)$$

A Green's function of D is a function $G(x, x')$ such that:

$$DG(x, x') = \delta(x' - x)$$

In words, $G(x, x')$ describes the impulse response of the differential operator D. It turns out that function $G(x, x')$ for D is all we need to solve the original equation. To see this, multiply by $f(x')$ and then integrate both sides:

$$\int DG(x, x')f(x')\,dx' = \int \delta(x' - x)f(x')\,dx'$$
$$D\int G(x, x')f(x')\,dx' = f(x)$$

Where the second lines follows because D is linear and operates only on x, not x'. Now all we need to do is observe that:

$$v(x) = \int G(x, x')f(x')\,dx'$$

And we have a solution to the original equation:

$$Dv(x) = f(x)$$

In the case of the neural mass model, we have the following equation:

$$v(t) = \int_{-\infty}^{t} H(t)At\exp\left(-\frac{t}{\tau}\right)g(t')\,dt'$$

It turns out that we can rewrite this as a differential equation as follows. First, take derivatives of $h(t)$.

$$h(t) = H(t)At\exp\left(-\frac{t}{\tau}\right)$$
$$h'(t) = H(t)A\exp\left(-\frac{t}{\tau}\right) - \frac{1}{\tau}h(t)$$
$$h''(t) = A\delta(t) - \frac{1}{\tau}H(t)A\exp\left(-\frac{t}{\tau}\right) - \frac{1}{\tau}H(t)A\exp\left(-\frac{t}{\tau}\right) + \frac{1}{\tau^2}h(t)$$
$$= A\delta(t) - \frac{2}{\tau}H(t)A\exp\left(-\frac{t}{\tau}\right) + \frac{1}{\tau^2}h(t)$$

Now consider the linear differential operator $D = \frac{d^2}{dt^2} + \frac{2}{\tau}\frac{d}{dt} + \frac{1}{\tau^2}$:

$$Dh(t) = \left(A\delta(t) - \frac{2}{\tau}H(t)A\exp\left(-\frac{t}{\tau}\right) + \frac{1}{\tau^2}h(t)\right) + \frac{2}{\tau}\left(H(t)A\exp\left(-\frac{t}{\tau}\right) - \frac{1}{\tau}h(t)\right) + \left(\frac{1}{\tau^2}\right)$$
$$= A\delta(t) + \frac{1}{\tau^2}h(t) - \frac{2}{\tau^2}h(t) + \frac{1}{\tau^2}$$
$$Dh(t) = A\delta(t)$$

Hence we have established that $h(t)$ is a Green's function of D. Now let us apply the same process as above to find the full equation:

$$DG(x,x') = \delta(x'-x)$$
$$\int DG(x,x')f(x')\,dx' = \int \delta(x'-x)f(x')\,dx'$$
$$D\int G(x,x')f(x')\,dx' = f(x)$$

Substitute in the relevant functions in the present case:

$$D\int H(t)At\exp\left(-\frac{t}{\tau}\right)g(t')\,dt' = Ag(t)$$

Where we our solution:

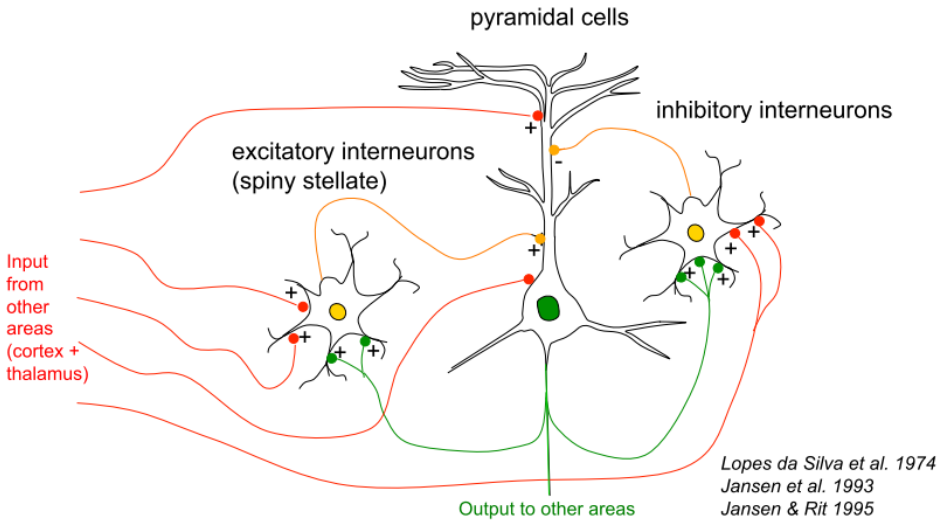
$$v(t) = \int_{-\infty}^{t} H(t)At\exp\left(-\frac{t}{\tau}\right)g(t')\,dt'$$

Thus we can write the differential equation:

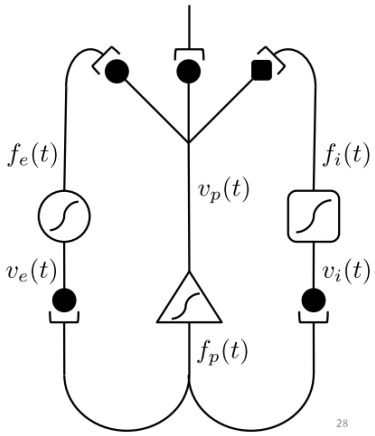
$$\frac{d^2}{dt^2}v(t) + \frac{2}{\tau}\frac{d}{dt}v(t) + \frac{1}{\tau^2}v(t) = Ag(t)$$

Modelling a Cortical Column

These methods can be applied to model a cortical column. In such models, functionally equivalent neurons are lumped together into a single node. The behaviour of the resulting circuitry should be prototypical for individual units in the underlying neural network.



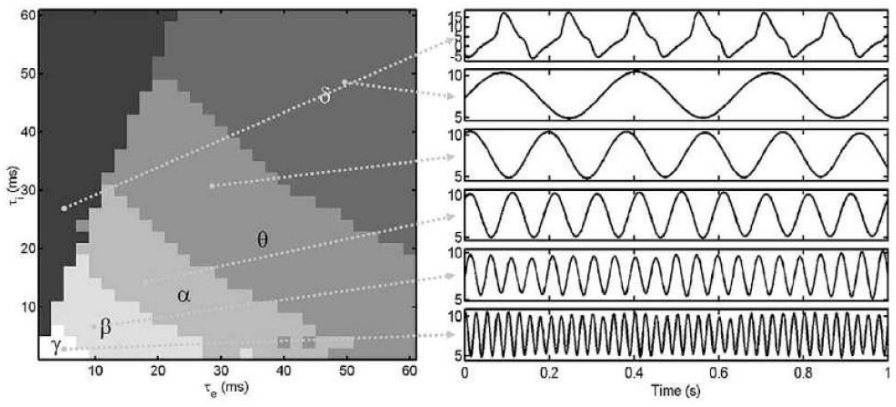
In one canonical circuit model, there is one excitatory neuron, one inhibitory neuron, and one excitatory pyramidal neuron.



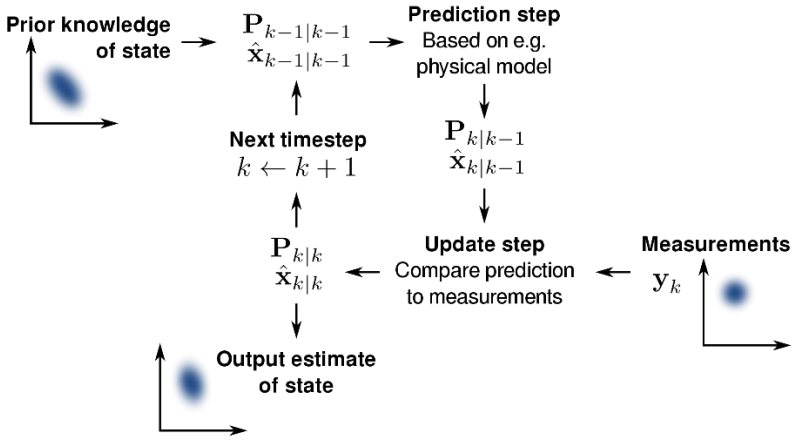
Parameters in these models include the time constants, synaptic gains, maximum firing rate, and connection weights. These are derived from experimental data from rodents, monkeys, and cats.

EEG Generation

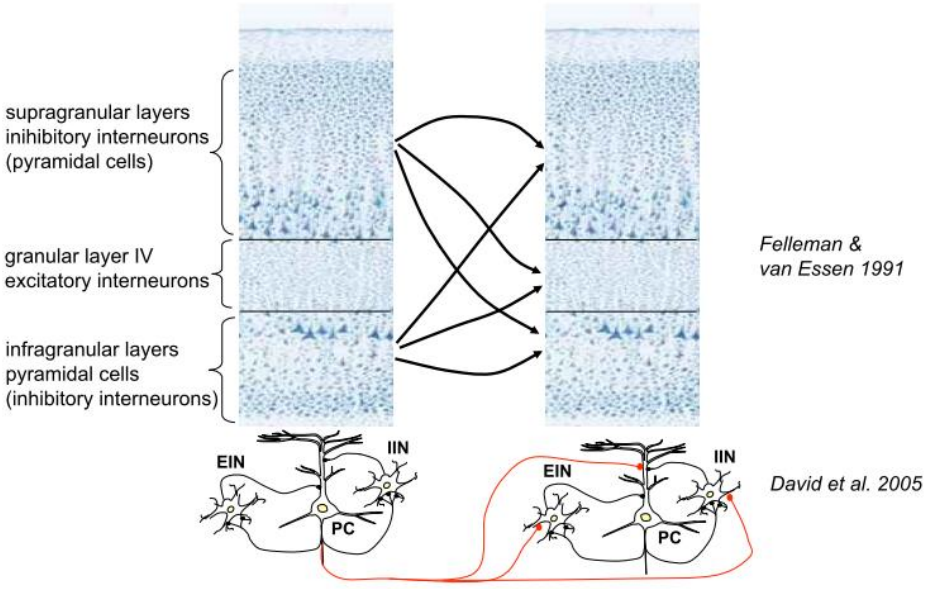
The value of neural mass models can be analysed by studying how parameters affect the simulated EEG and comparing to experimentally recorded EEGs. Such comparisons have found that altering the time constants reproduces oscillation patterns analogous to those found in various EEG waveforms.



The parameters of EEG models can be learned from dynamic data feeds using a technique called Kalman filtering. Kalman filters work by comparing the predicted state variables from a stochastic model of a dynamical system, and then updating these predictions based on the observed values that are made in that period.



Basic neural mass models can be extended by coupling multiple such models together as input-output pairs.

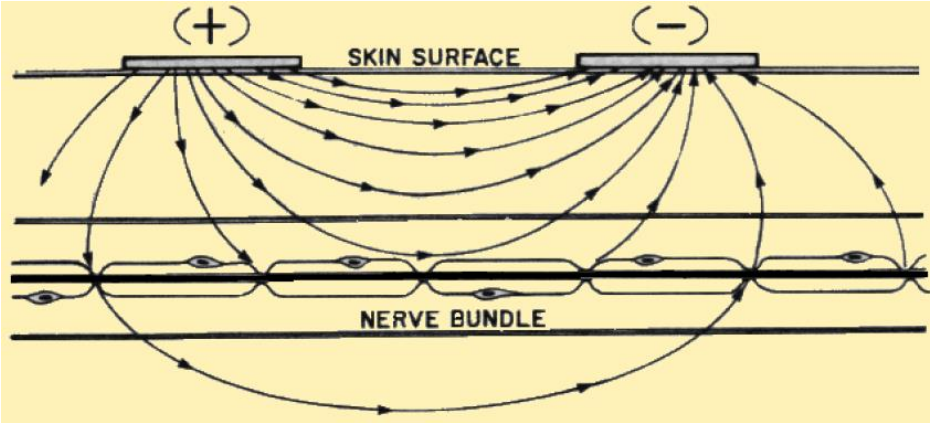


Neural Interfaces

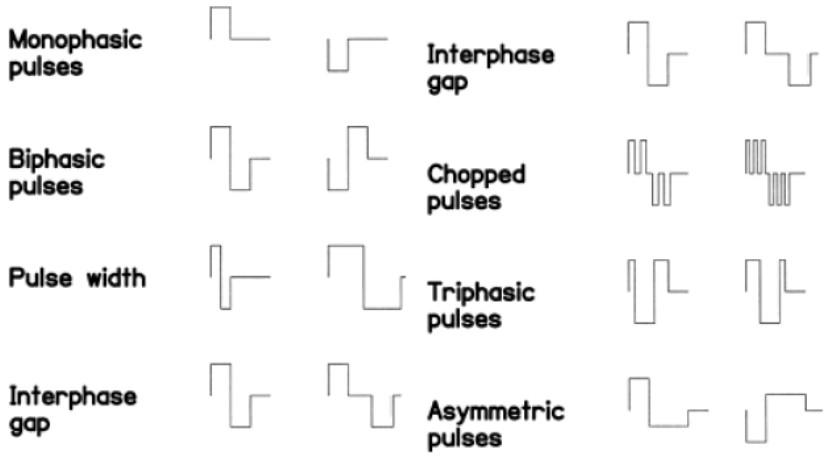
Lecture 20

Electrical Stimulation

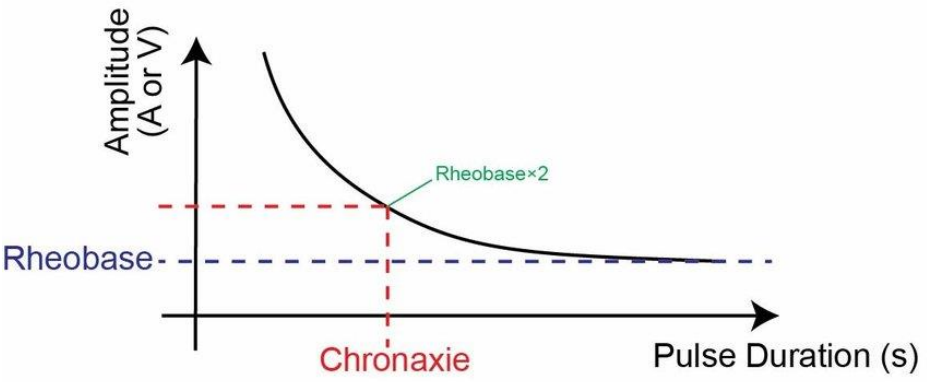
Stimulating electrodes causes currents of ions to flow in tissue. This elicits action potentials in surrounding neurons.



Different forms of stimulation can be used, but typically charge-balanced biphasic phases are used, where both phases are symmetric so as to ensure no net charge enters or leaves the tissue. Interphases gaps or chopped phases can be devised to allow the charge to persist for longer.



The membrane properties of neural tissue can be measured by constructing a strength-duration curve, in which the pulse duration is plotted against the minimum current needed to elicit action potentials, and typically is a hyperbolic shape. The Rheobase current is the threshold current level as pulse duration approaches infinity, while the Chronaxie is the pulse width at a current level twice rheobase.



Stimulation Parameters

Pulse rate

- An increase in the pulse rate will increase the stimulation of the excitable tissue.
- Each individual nerve (or muscle fibre) has an upper limit on the rate at which it can respond, due to refractory effects.
- However, other (more distant) nerves or fibres will respond as the pulse rate increases.
- The response of the population of fibres tens to lose synchronisation as the pulse rate increases.

Pulse amplitude

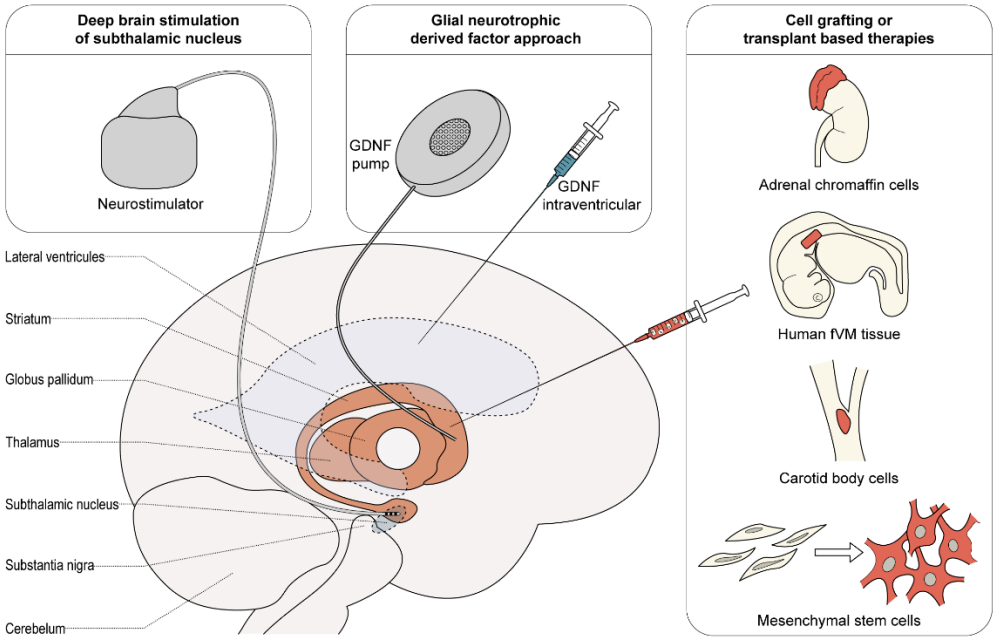
- An increase in the pulse amplitude will increase the stimulation of the excitable tissue.
- With larger amplitude the current will spread more broadly thus exciting more (distant) nerves.
- In general the current amplitude will be set to remain between the threshold of activation and some upper bound (often called the maximum comfortable level).

Pulse duration

- An increase in the pulse duration lowers the threshold of activation of nerve fibres.
- The same level of stimulation can be achieved by using shorter pulses with higher amplitude or longer pulses with lower amplitude.

Parkinson's Disease

Deep brain stimulation can be used to treat the symptoms of Parkins's disease. The targets of this stimulation are the subthalamic nuclei and globus pallidus, all targets of the dopaminergic system. Deep brain stimulation involves the implantation of a neurostimulator, which sends electrical impulses to specific parts of the brain. It is recommended for people who have motor fluctuations and tremor inadequately controlled by medication, or to those who are intolerant to medication.

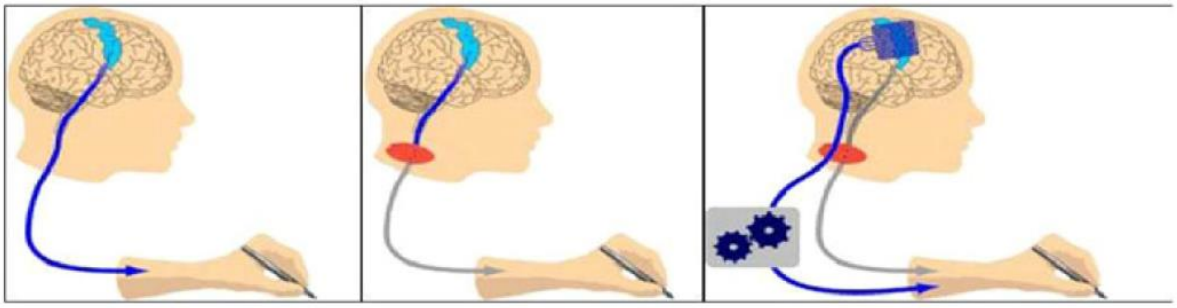


The substantia nigra (SN) is a basal ganglia structure located in the midbrain that plays an important role in reward and movement. Parkinson's disease is characterized by the loss of dopaminergic neurons in the substantia nigra pars compacta.

Lecture 21

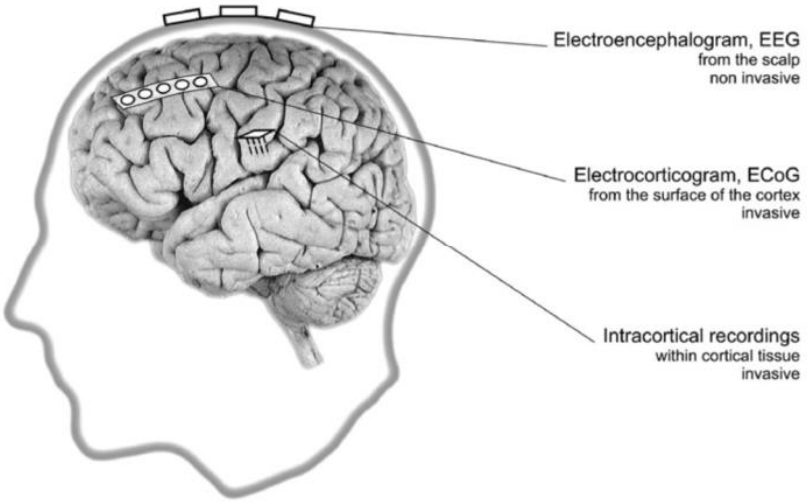
Brain-Computer Interfaces

Brain-Computer Interfaces (BCI) or Brain-Machine Interface (BMI) is a technology to communicate between the human brain directly to a computer without any physical contact. The idea is to bypass damaged or removed neural connections between a brain region (such as the motor cortex) and the effectors (such as muscles) by connecting both to an external computer.

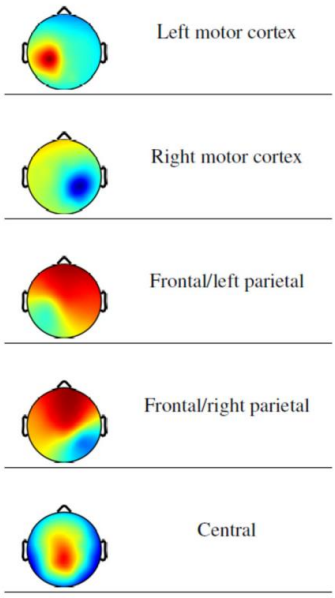


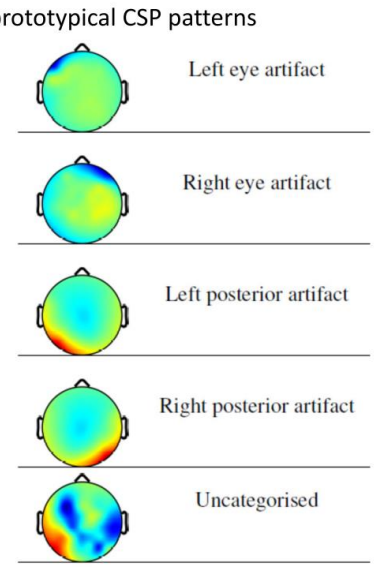
There are three main methods of detecting brain activity. The more invasive methods can use smaller electrodes, and hence there is higher resolution and less noise.

- Electroencephalogram from the scalp (non-invasive), includes EEG, MEG, and fMRI.
- Electrocorticogram from the cortex surface (invasive), includes stentrodes and microarrays.
- Intracortical recordings from electrodes deep in the brain (highly invasive).



BCIs need to measure and interpret neural activity so as to produce useful motor responses. Volitional changes in oscillatory activity near the sensorimotor cortex, known as sensorimotor rhythms (SMRs), can be measured, and localised using Common Spatial Patterns (CSPs). This helps to linearly combine information from multiple EEG electrodes to accentuate SMR activity.

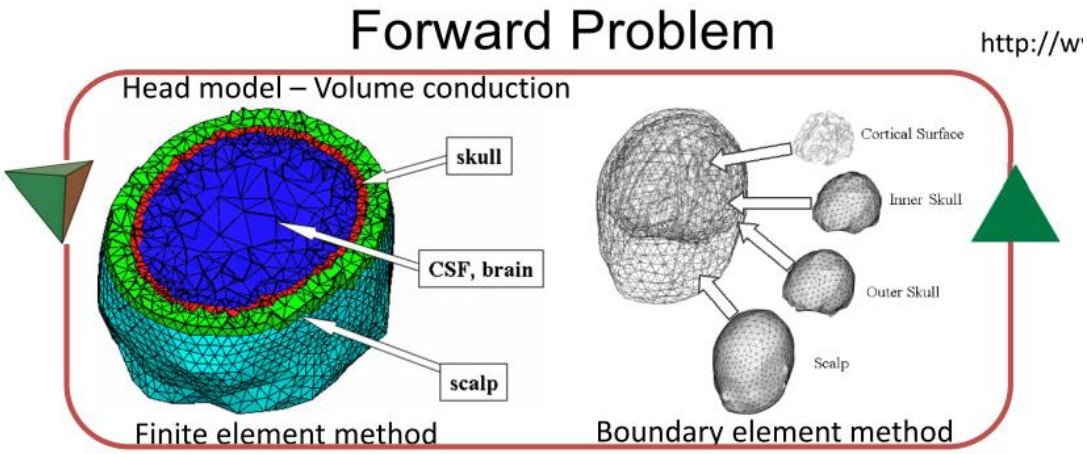




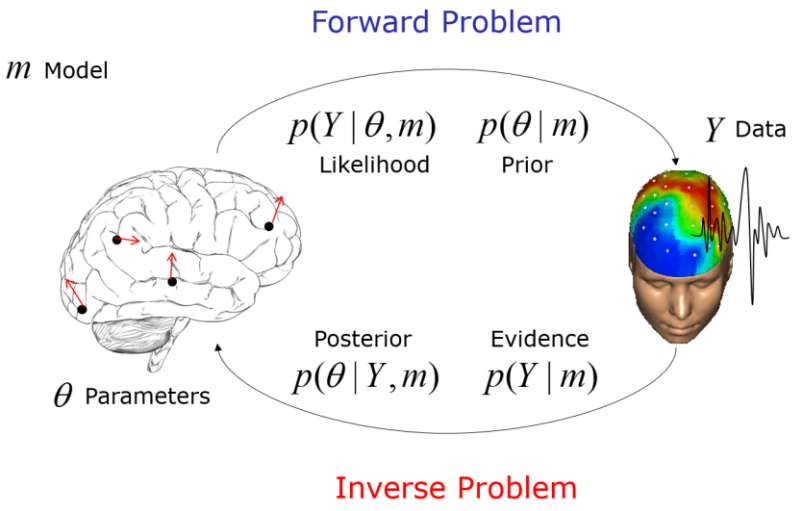
EEG and MEG

A forward model involves computing the scalp potentials at a finite set of electrode or sensor locations and orientations (called channel configuration) for a given predefined set of source positions and orientations (source space). There are four components of a forward model:

1. A head model: need to know how the electric currents generated at the source spread throughout the volume conductor (head). Typically four-layered concentric circle model is used, corresponding to the brain, the CSF, the skull, and the scalp.
2. A sensor description: need to know where the sensors are that pick up the activity coming from the sources.
3. A source model: need to know where the sources are within the brain.
4. A lead field: for each source we calculate the electric potential vector at each sensor (electrode).

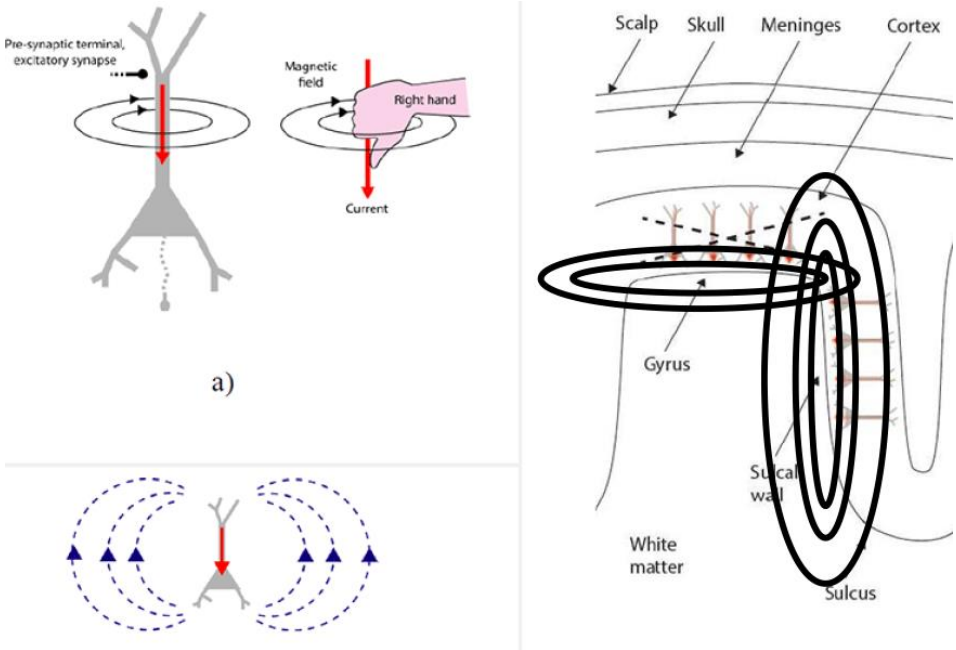


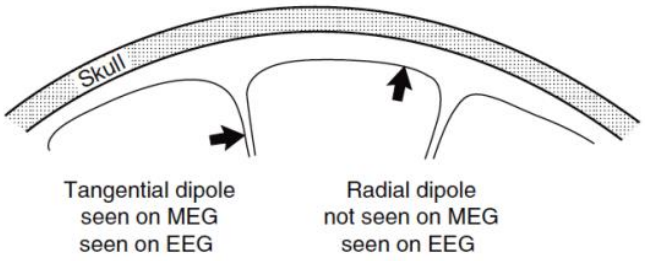
The inverse problem involves using the forward model to fit parameters of the model against the EEG data.



Pros and cons of the two techniques:

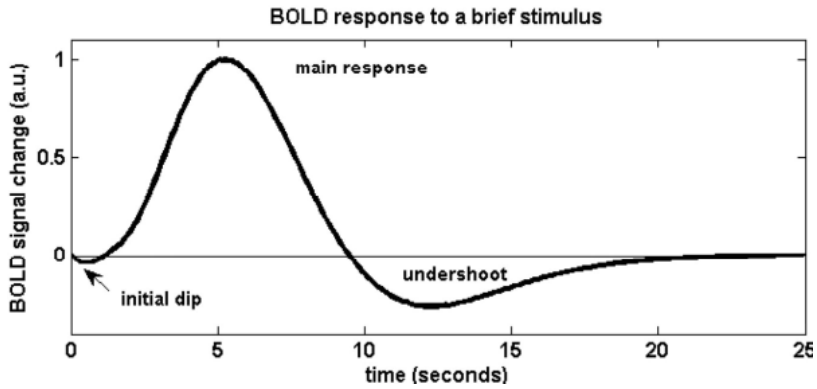
- MEG is more sensitive to currents tangential to the surface of the scalp, EEG is sensitive to tangential and radial neuronal activities.
- Magnetic fields are not distorted by the tissue the scalp, skull, cerebrospinal fluid, and brain.
- MEG provides better spatial resolution of source localization (2-3 mm) than EEG (7-10 mm).
- MEG hardware is costlier.
- Patient setup is shorter with MEG compared to traditional EEG.



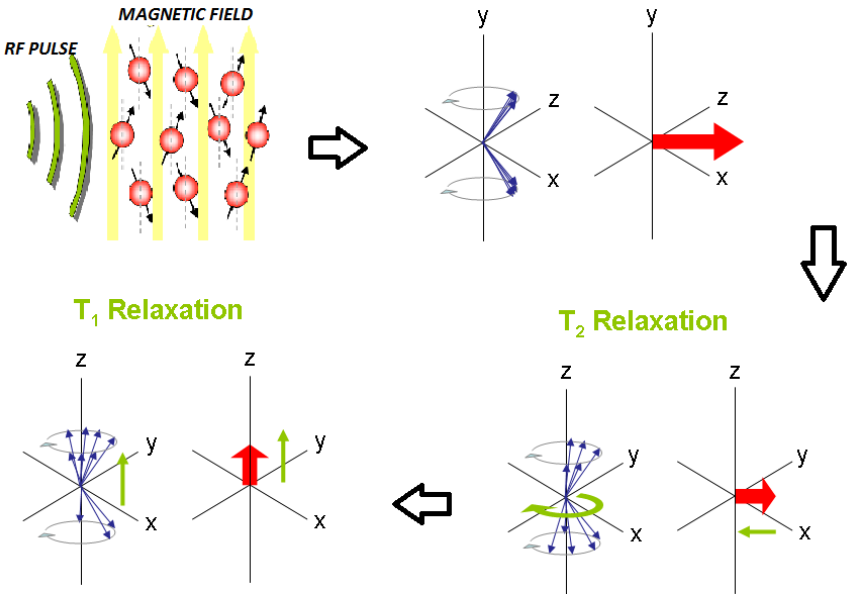


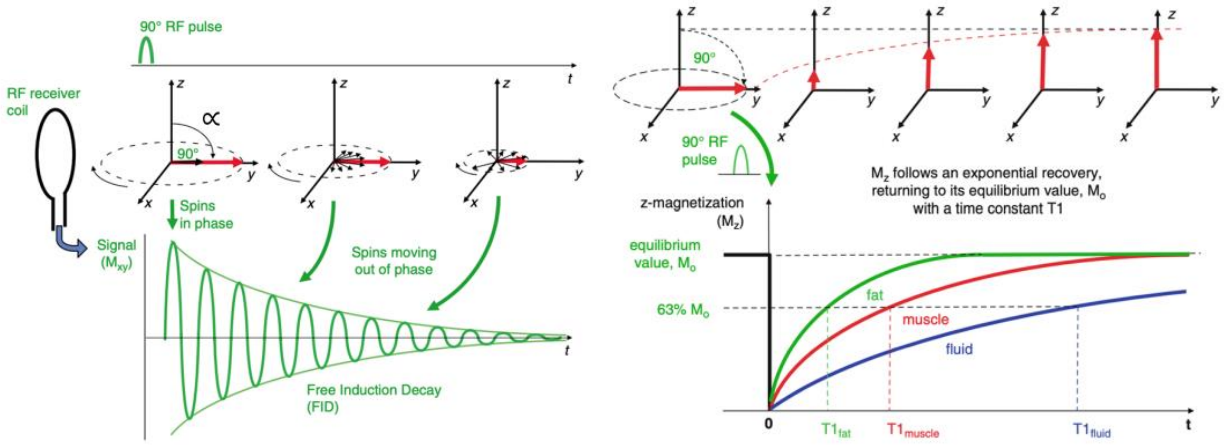
Functional magnetic resonance imaging

Increased neural activity leads to a delayed increased provision of bloodflow, as shown in the diagram below. This is called the BOLD response.

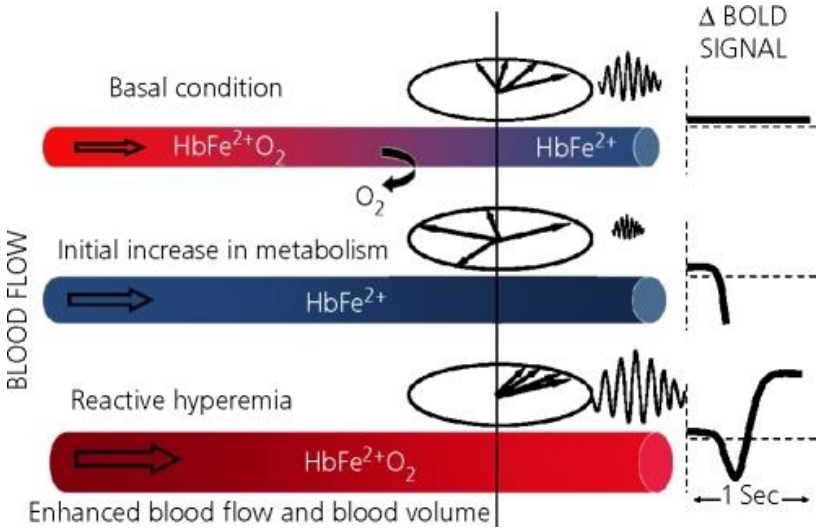


Hemoglobin differs in how it responds to magnetic fields, depending on whether it has a bound oxygen molecule. Deoxygenated hemoglobin (dHb) is more magnetic (paramagnetic) than oxygenated hemoglobin (Hb), which is virtually resistant to magnetism (diamagnetic). Magnetic spins of proton nuclei are aligned using a strong external magnet, and then a brief RF pulse applied to the field, which causes the nuclei spins to align in phase. However this is not an equilibrium, so the system will gradually relax back to a distribution of phases. This relaxation time is called the spin-spin transverse relaxation time, T_2^*.





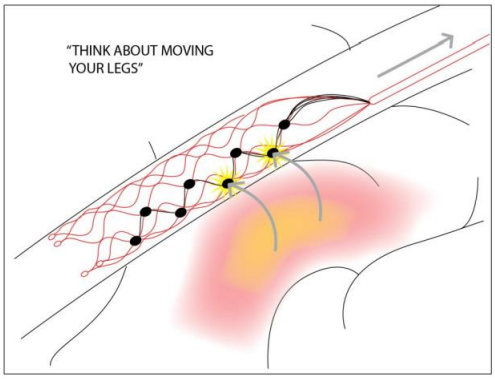
This decay occurs more rapidly in dHb compared to Hb because of the former's greater magnetic interaction and the resulting larger local field inhomogeneities. Hence, the more oxygen in the tissue, the more slowly the magnetic signal decays, and hence the larger the BOLD signal.

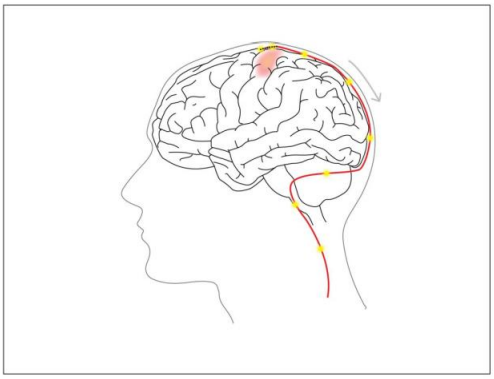


fMRI is non-invasive and has high spatial resolution, with minimal setup difficulty. However, it is a measure of metabolic input rather than neural processing directly. It also has poor temporal resolution, and can be affected by various factors such as drugs, pathology, age, and attention.

Stentrodes and Microarrays

Stentrode (Stent-electrode recording array) is a small stent-mounted electrode array permanently implanted into a blood vessel in the brain, without the need for open brain surgery. It is in clinical trials as a brain–computer interface (BCI) for people with paralyzed or missing limbs, who will use their neural signals or thoughts to control external devices.

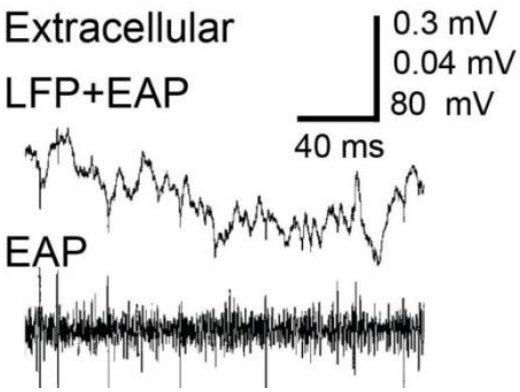


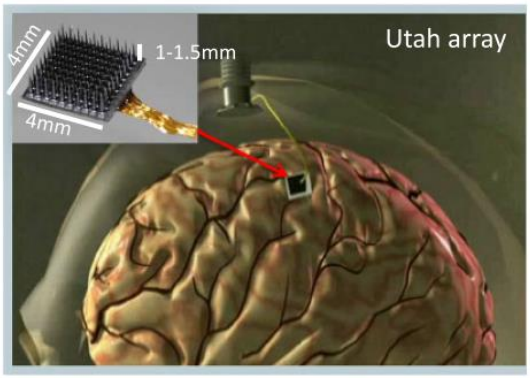


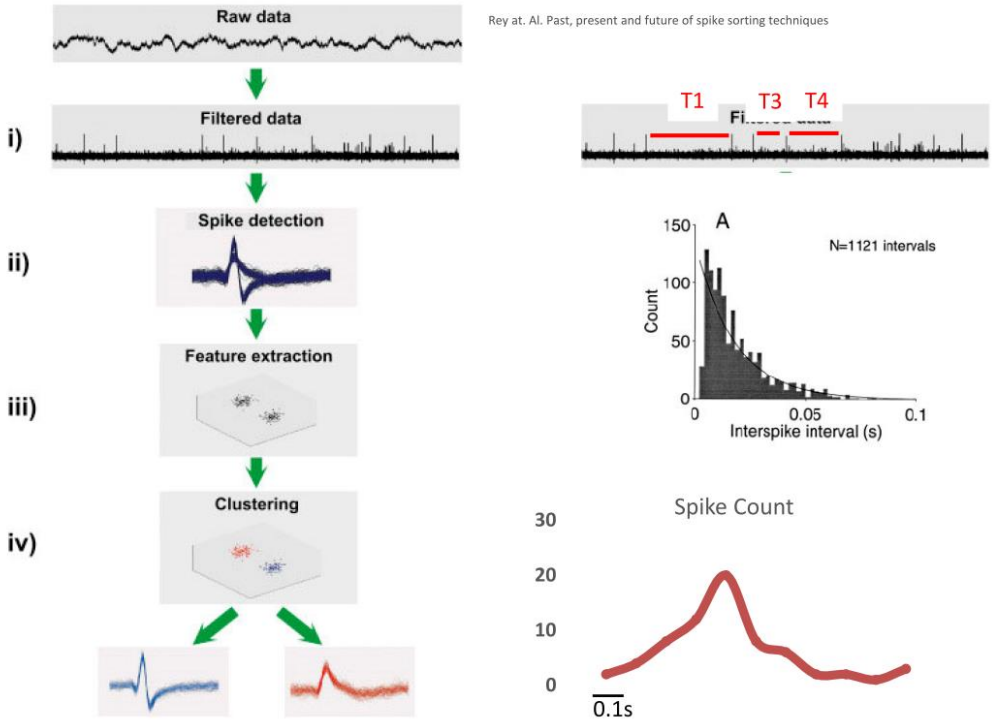
Ask patient to think about moving

Acquire and transmit brain signal

A microelectrode array is a more invasive devise implanted on the surface of the brain. It can measure extracellular potentials from a specific area using its array of electrodes. Significant processing is required to extract individual action potentials.



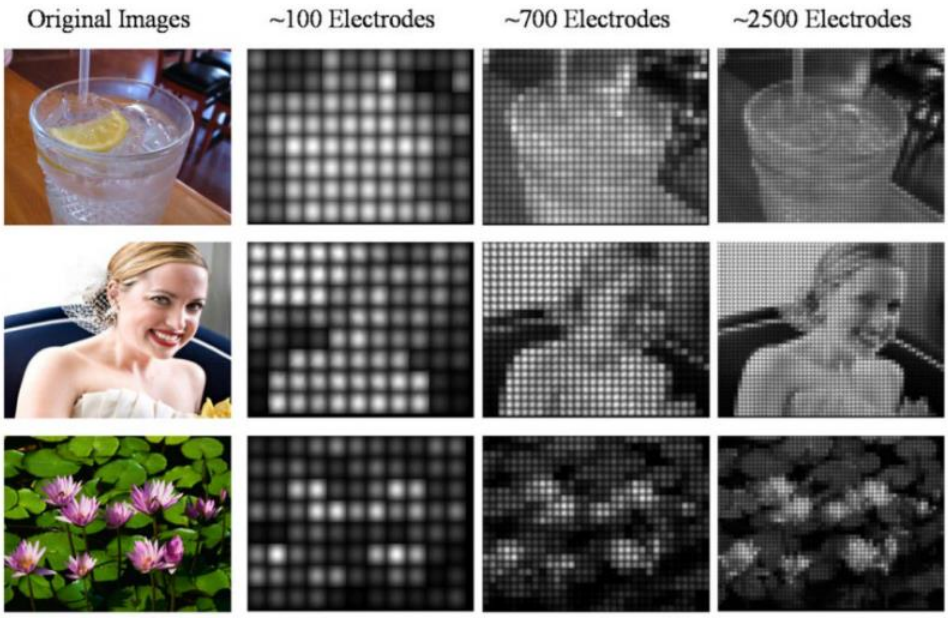




Lecture 22

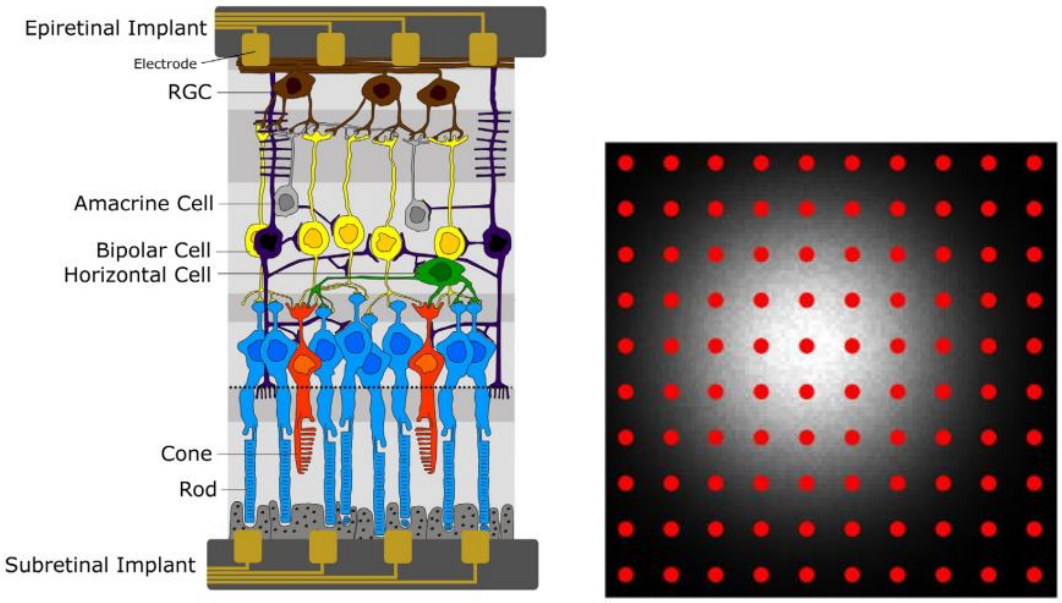
Standard Methods

A phosphene is the phenomenon of seeing light without light entering the eye. The fundamental idea of constructing retinal implants is to find a mapping from patterns of electrode activation to patterns of retinal activation, which will then map into perceptual experiences. As shown in the figure below, adding more electrodes should increase the resolution, and hence allow sharper depiction of images. However, this analysis assumes that phosphenes do not to overlap, even at high electrode densities.



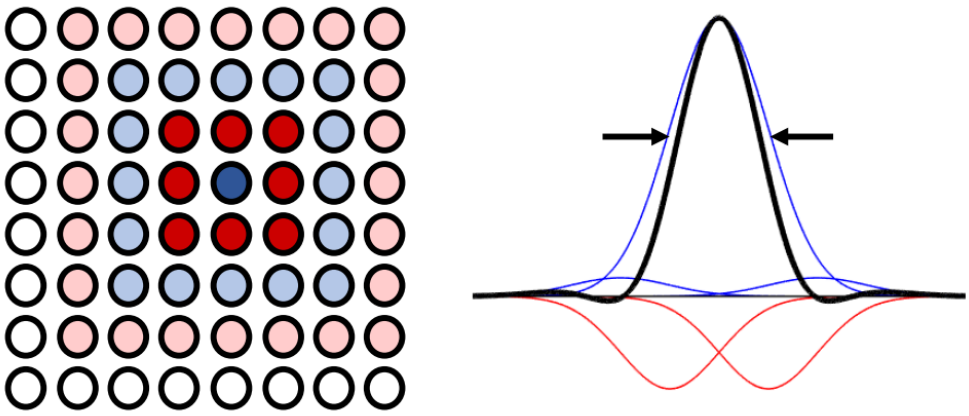
Unfortunately this assumption is not accurate. Even with small electrodes, the distance between the electrodes on the surface of the retina and the retinal ganglion cells behind them, is far more important.

There is a great deal of ‘activity spreading’ from each electrode (red dot) to reginal activation (white phosphene), so the simple approach of making electrodes smaller won’t work.

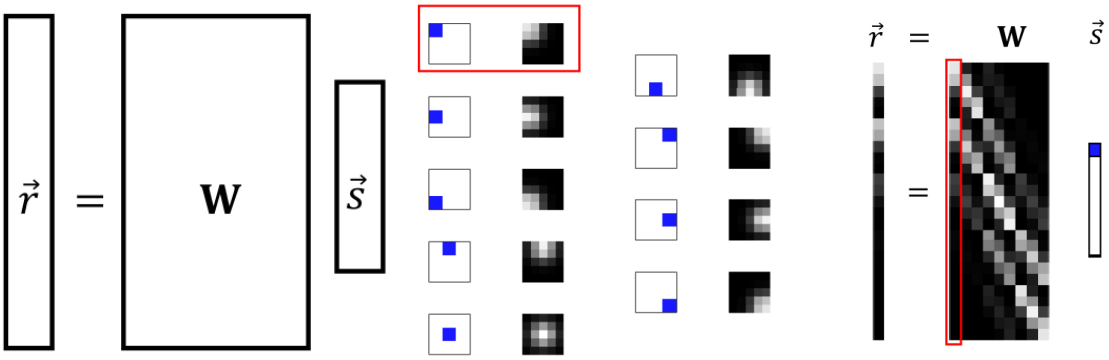


Linear Current Steering

A more sophisticated approach to avoid these limitations is known as current steering. Current steering involves simultaneous stimulation of several electrodes at once, so as to produce an overall activation pattern on the retinal ganglion cells that is more useful than one produced by activating electrodes one at a time. For example, the grid below shows a combination of positive (pink) and negative (red) current injected into various electrodes, which produce an overall pattern of ganglion activation (black curve) which is more tightly localised than is possible by activating just a single electrode at the corresponding location.



This basic approach can be refined to construct a method in which arbitrary patterns of ganglion activation can be produced by the right combination of electrode activations. The first step is to define a forward model, specifying the retinal ganglion activations $\tilde{r}$ that result from a given set of electrode activations $\tilde{s}$. The matrix W describes all the current spreads for each electrode.



Now we simply invert the forward model to solve for the required activations needed to produce a desired target image. Note that in practise the matrix W is seldom invertible, so a pseudo-inverse must be used instead, via Singular Value Decomposition:

$$W = PDQ^T$$
$$W^+ = QD^{-1}P^T$$

Where Q and P contain the orthogonal eigenvectors of WW^T and W^TW respectively.

Forward Model: $\vec{r} = \mathbf{W}\vec{s}$

Inverse Model: $\vec{s} = \mathbf{W^{-1}}\vec{r}^*$

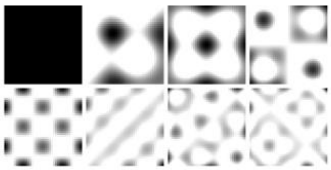
$\mathbf{W^{-1}}$ is the inverse of $\mathbf{W}$
$\vec{r}^*$ is a target retinal activation pattern

Inverse Model $\vec{s}$ = $\mathbf{W^{-1}}$ $\vec{r}^*$

Desired Retinal Activation Pattern

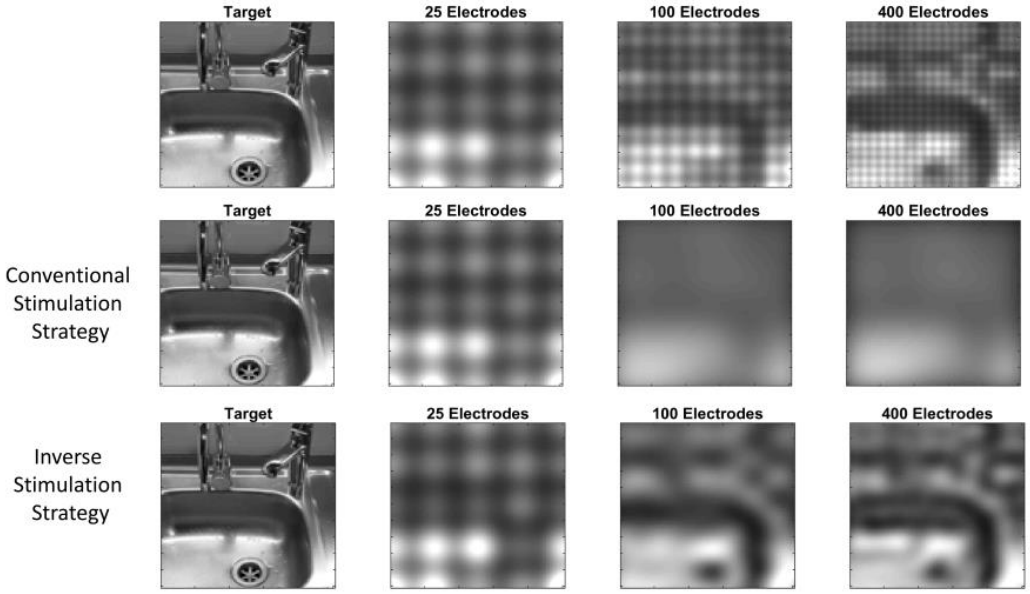
$\vec{s} = \mathbf{QD^{-1}P}^T \vec{r}^*$

P – retinal pattern



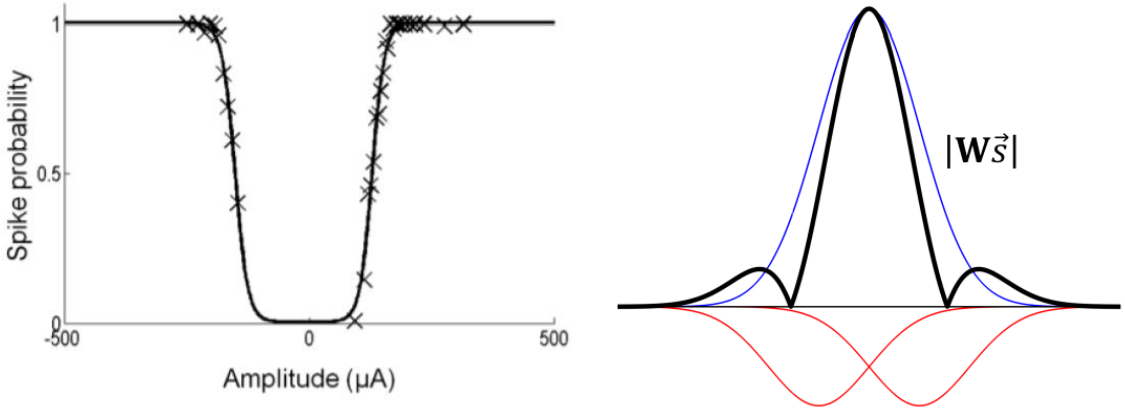
Q – electrode pattern

This approach results in a much superior image quality compared to conventional methods.

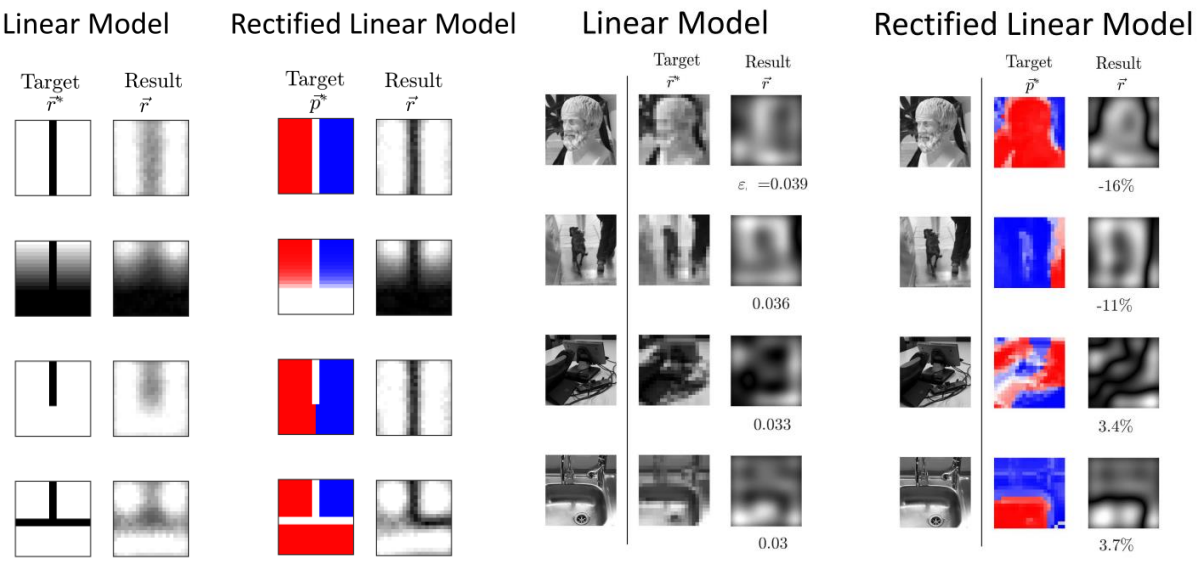


Rectified Current Steering

There is a further aspect to the story, however, since experiments show that only the absolute value of the electrode current matters, meaning that positive and negative electrode currents both yield the same activity in retinal ganglion cells. This represents an opportunity to 'utilise' high-contrast lines that are not available in the pure linear model. This approach is called a rectified linear model.



Whereas before we calculated the activity using $\tilde{r} = W\tilde{s}$, now we use the formula $\tilde{r} = |W\tilde{s}|$. Previously, negative currents were only used to attenuate and shape positive current. Now, total negative values of $W\tilde{s}$ are used to directly create retinal activation. Unfortunatley we now cannot use pseudo-inverse matrix methods, and must solve for $\tilde{s}$ using numerical methods. This is NP-hard, and is done using simulated annealing and other techniques.



Lecture 23

Cochlea Implants

Cochlea implants are beneficial for people with profound sensorineural hearing loss, who do not benefit from conventional hearing aids. This means that the damage exists at the level of the cochlea or hair cells, while the auditory pathway from the cochlea to the brain is intact. The technology consists of two main components: one implanted internally and the other worn externally.

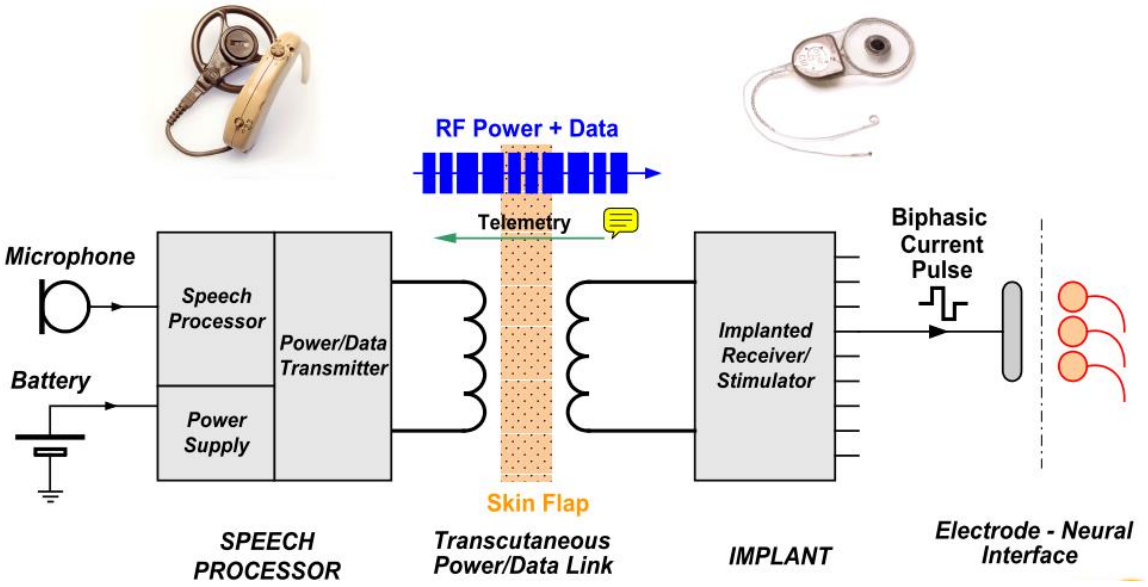
Externally worn components:

- Microphone
- Speech Processor
- Power (batteries)
- Transmitter

Implanted surgically:

- Receiver (in a drilled indentation on the temporal bone)
- Electrode Array (inserted in the cochlea)

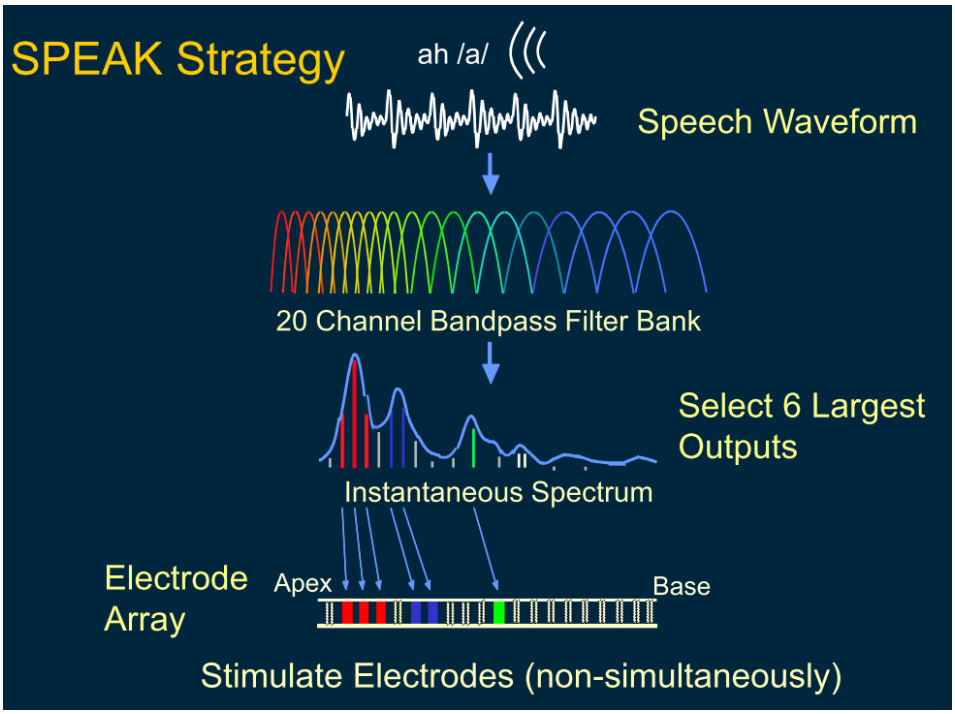
The device works as follows. First, sound is detected by microphone and converted into electrical signal, which is then sent to the speech processor. The input signal is analysed and relevant features are extracted and encoded. The coded signal is then sent via the transmitting coil, through the skin, to the receiver as an RF signal. This is to avoid long-term implantations protruding through the skin, which are very difficult to maintain without infection. The implanted receiver then decodes the RF signal to determine the electrode number, stimulation level and stimulation rate. This signal is transformed into the appropriate electrical pulses in the electrode array in the cochlea, which stimulates the nerves inside the cochlea. The receive also sends telemetry data back to the external unit, for diagnostic and recording purposes.



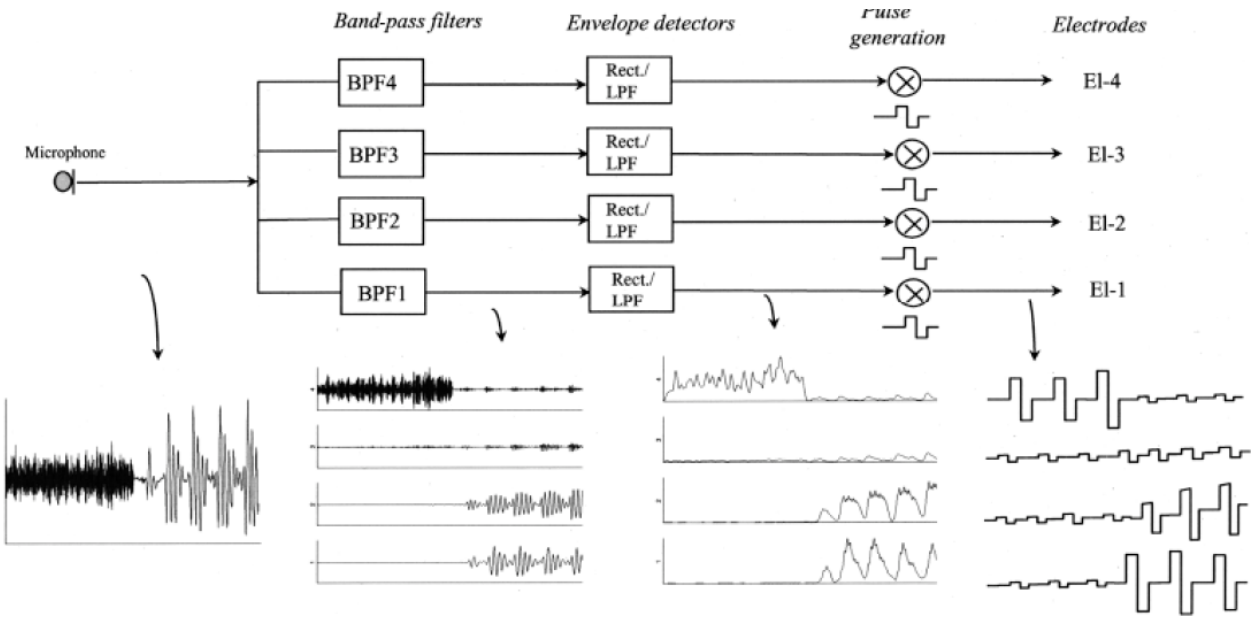
The electrode array is wound into a flexible coil to fit inside the cochlea. It must be strong but inert, with platinum used for the electrodes. The implanted component must be non-corrosive, totally sealed, and resistant to mechanical stresses and vibration. This is critical because it typically remains implanted for the life of the patient, and circuitry will be rendered useless if any conductive fluids seep through. A magnet on the receiver holds the externally worn coil in place on the patient's skin and ensures that the transcutaneous link is reliable and efficient.

Speech Processing

The key bottleneck in the Cochlea design is the signal transmitted from the transmitter to the implanted receiver. As such, the speech processor must be carefully programmed to extract and send only the most important components of the recorded sound. The electrical encoding exploits the place mechanism for coding frequencies, with high frequencies near base of cochlea, and low frequencies at the more apical portion of cochlea.



The diagram below summarises some of the key stages of the processing pathway. The signal first goes through a set of bandpass filters that divide the acoustic waveform into six channels. The envelopes of the bandpassed waveforms are then detected by rectification and low-pass filtering. Current pulses are generated with amplitudes proportional to the envelopes of each channel and transmitted to the electrodes through a radio-frequency link.



Important calibration tests include:

- T-Levels: the level at which the patient first identifies sound sensations. Determined by passing the person's hearing threshold using an ascending method.
- C-Levels: the maximum stimulation level that doesn't produce an uncomfortable loudness sensation for the patient.

Performance

Experience with the cochlea is generally good when implanted later in life after acquired hearing loss. For congenital deafness, the performance depends on the age of implantation, with earlier ages of implantation showing the best outcomes. However, years of training are still required to obtain optimal performance. Performance also depends on the condition of the surviving hair cells, the health of the auditory neve fibres, and the central auditory neurons.