

# Contents

<b>1</b>	<b>Mathematical Foundations</b>	<b>4</b>
1.1	Common Maclaurin series . . . . .	4
1.2	Differential equations . . . . .	4
1.3	Statistics . . . . .	7
<b>2</b>	<b>Biophysics</b>	<b>9</b>
2.1	Ionic current flow . . . . .	9
2.2	Passive membrane potential dynamics . . . . .	12
2.3	Active membrane potential dynamics . . . . .	14
2.3.1	Action Potentials . . . . .	14
2.3.2	Hodgkin-Huxley Model . . . . .	16
2.3.3	Leaky Integrate and Fire Model . . . . .	19
2.4	Synapses . . . . .	22
<b>3</b>	<b>Spike patterns</b>	<b>25</b>

3.1	Spike statistics . . . . .	25
3.2	The role of noise . . . . .	26
<b>4</b>	<b>The visual system</b>	<b>28</b>
4.1	Retina . . . . .	28
4.1.1	Photoreceptors . . . . .	28
4.1.2	Receptive fields . . . . .	29
4.2	Thalamus . . . . .	31
4.3	Primary visual cortex . . . . .	32
<b>5</b>	<b>Coding of information</b>	<b>34</b>
5.1	Information theory . . . . .	34
5.2	Efficient coding in the visual system . . . . .	36
5.2.1	Retina: Whitening . . . . .	36
5.2.2	V1 and beyond: Generative models . . . . .	37
5.2.3	V1: Sparse coding models . . . . .	38
5.2.4	V1 and beyond: Predictive coding . . . . .	40

5.3	Population coding . . . . .	41
5.3.1	Population decoders . . . . .	41
5.3.2	Fisher information . . . . .	44
5.3.3	Noise correlations . . . . .	45
<b>6</b>	<b>Networks of neurons</b>	<b>47</b>
6.1	E-I networks . . . . .	47
6.2	The ring network . . . . .	49
6.3	Network stability . . . . .	50
<b>7</b>	<b>Plasticity</b>	<b>52</b>
7.1	Single neuron plasticity . . . . .	52
7.2	Short-term synaptic plasticity . . . . .	53
7.3	Long-term synaptic plasticity . . . . .	54
7.4	Learning . . . . .	56

# 1 Mathematical Foundations

## 1.1 Common Maclaurin series

- $e^x = 1 + x + \dots$
- $\ln(1 + x) = 0 + x + \dots$
- $(1 + x)^a = 1 + ax + \dots$

## 1.2 Differential equations

- Proposition (first order homogeneous ode):

$$\frac{dx}{dt} = ax \Rightarrow x(t) = x_0 e^{at} \text{ where } x_0 = x(0)$$

$$\text{Proof: } \frac{dx}{dt} = ax \Rightarrow \frac{1}{x} dx = a dt \Rightarrow \ln |x| = at + C \Rightarrow x = e^C e^{at}.$$

$$\text{Finally, } x(0) = e^C e^0 = e^C$$

- Proposition (Euler's method for first order ODE):** Let  $\frac{dx}{dt} = f(t, x)$ . Then,  $x(t + h) \approx x(t) + f(t, x(t))h$ . This provides an iterative process starting from  $x(0)$  (or any known point)

Intuition: Taking a step of size  $h$  along the tangent line is an approximation to staying on the curve

Proof: Consider the Taylor series for  $x$  around  $t$ :

$$x(u) = x(t) + f(t, x(t))(u - t) + \dots \approx x(t) + f(t, x(t))(u - t).$$

Thus,  $x(t + h) \approx x(t) + f(t, x(t))(h)$
- Proposition (first order non-homogeneous ode) (worth memorising):** Let  $\frac{dx}{dt} = ax + f(t)$ . Then,

$$x(t) = x_0 e^{at} + e^{at} \int_0^t e^{-a\tau} f(\tau) d\tau.$$

However, if this integral does not have a nice solution then we are back to Euler's method

Proof: First introduce the integrating factor,  $\frac{dx}{dt} = ax + f(t) \Rightarrow \exp(-at) \frac{dx}{dt} - \exp(-at)ax = \exp(-at)f(t)$ .

Then note that  $\frac{d}{dt} (\exp(-at)x) = -ax \exp(-at) + \exp(-at) \frac{dx}{dt}$ . Thus (simultaneously doing a change of variable for later clarity),

$$\begin{aligned}
& \frac{d}{d\tau} (\exp(-a\tau)x) = \exp(-a\tau)f(\tau) \Rightarrow \\
& \int_0^t \frac{d}{d\tau} (\exp(-a\tau)x) d\tau = \int_0^t \exp(-a\tau)f(\tau) d\tau \Rightarrow \\
& [\exp(-a\tau)x]_0^t = \int_0^t \exp(-a\tau)f(\tau) d\tau \Rightarrow \\
& \exp(-at)x = x_0 + \int_0^t \exp(-a\tau)f(\tau) d\tau \Rightarrow \\
& x = \exp(at)x_0 + \exp(at) \int_0^t \exp(-a\tau)f(\tau) d\tau \text{ as required}
\end{aligned}$$

- Proposition (first order non-homogeneous ode with matrix coefficients):

Let  $x, u(t)$  be vectors and  $W$  be a matrix. Then,

$$\frac{dx}{dt} = Wx + u(t) \Rightarrow x(t) = x_0 e^{Wt} + e^{Wt} \int_0^t e^{-W\tau} u(\tau) d\tau \text{ where}$$

$$e^{Wt} = \sum_{i=0}^{\infty} \frac{W^i t^i}{i!}$$

Proof sketch: By McLaurin series,  $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$

## 1.3 Statistics

- $\langle \mathbf{X} \rangle = \mathbf{E}[\mathbf{X}]$
- The Poisson distribution describes the number of events  $N$  per time window  $T$  from a process that generates events independently at random with a constant rate  $r$ .  $\Pr(N = k) = \frac{(rT)^k \exp(-rT)}{k!}$
- For a Poisson distribution,  $rT = \langle \mathbf{N} \rangle = \mathbf{Var}(\mathbf{N})$ . Moreover, as  $rT$  increases, the shape of Poisson approaches that of Gaussian (with  $\mu = \sigma^2 = rT$ )
- Fano factor  $(F) = \frac{\langle [X - \langle X \rangle]^2 \rangle}{\langle X \rangle} = \frac{\mathbf{Var}}{\mu}$ .  $F = 1$  is necessary but not sufficient for follows the Poisson distribution

- For a Poisson process with  $T$  divided into sufficiently small bins of uniform size  $\Delta t \ll \frac{1}{r}$ , the number of events in any bin follows a Bernoulli distribution (i.e. 0/1-valued with  $\Pr(1) = p$ ) with  $p = r\Delta t \exp(-r\Delta t) \approx r\Delta t$
- **An inhomogeneous Poisson process extends a Poisson process with a time-varying rate  $r(t)$** , but still has the events at each point in time independent of each other. Everything works the same as before at each point in time

## 2 Biophysics

### 2.1 Ionic current flow

- **Reversal potential** = the voltage at which the chemical driving force (flow from high concentration to low concentration) and electrical driving force (opposite charges repel) are in equilibrium
- **At the reversal potential, there is no net current flow** — within each ionic species there is likely a net current flow, it is merely the case the current flows of all the ions cancel out overall
- **Nernst equation:** Let  $E$  = reversal potential,  $R$  = universal gas constant,  $F$  = Faraday's constant,  $T$  = temperature of system,  $z$  = charge per ion. Then,  $E = \frac{RT}{zF} \ln\left(\frac{\#[\text{ion outside the cell}]}{\#[\text{ion inside the cell}]}\right)$ .  
That is that for a system at a fixed temperature (and naturally with a fixed charge per ion),  
 $E \propto \ln(\text{ratio of } \# \text{ ions outside the cell to } \# \text{ ions inside the cell})$

- Resting potential = reversal potential for a system where multiple species of ion are present
- Goldman-Hodgkin-Katz equation: Let  $E_m$  = resting potential,  $P_i$  = permeability of membrane to ion  $i$ ,  $R$  = universal gas constant,  $F$  = Faraday's constant,  $T$  = temperature of system. Then,  $E_m = \frac{RT}{F} \ln\left(\frac{\sum_i P_i \#[\text{ion outside the cell}]}{\sum_i P_i (\text{sgn}(z_i)) \#[\text{ion inside the cell}]}\right)$  — note that the sums are inside the fraction, summing the individual fractions would give a different result. In particular, this way we still have that  $E \propto \ln(\text{ratio of } \# \text{ ions outside the cell to } \# \text{ ions inside the cell})$  for positively charged ions (and  $E \propto -\ln(\text{ratio of } \# \text{ ions outside the cell to } \# \text{ ions inside the cell}) = \ln(\text{ratio of } \# \text{ ions inside the cell to } \# \text{ ions outside the cell})$  for negatively charged ions)

- $g = \frac{1}{R}$ , thus (as  $V = IR$ )  $I = gV$ .  $g$  is called the conductance —  $V$  is the potential difference driving the current i.e.  $V_p - E_m$  where  $E_m$  is the resting potential and  $V_p$  is the membrane potential
- Proposition: A system with multiple ions is equivalent to a system with a single pseudo ion with membrane conductance  $g_m = \sum_i g_i$  and resting potential  $E_m = \frac{\sum_i g_i E_i}{g_m}$   
 Proof: Let  $I_m$  be the net current flow through the membrane, and  $V_m$  be the membrane potential. Then,  $I_m(V_m) = \sum_i I_i(V_m) = \sum_i g_i(V_m - E_i) = \sum_i g_i V_m - \sum_i g_i E_i$ .  
 $E_m$ : At  $V_m = E_m$ ,  $I_m(E_m) = 0$  and so  $E_m = \frac{\sum_i g_i E_i}{\sum_i g_i}$ .  
 $g_m$ :  $I_m = \sum_i g_i V_m - \sum_i g_i E_i = \sum_i g_i V_m - (\sum_i g_i) E_m = (\sum_i g_i) (V_m - E_m) = g_m (V_m - E_m)$
- Limitation of this model:  $V = IR$  is only a linear approximation, real-life systems are non-ohmic!

## 2.2 Passive membrane potential dynamics

- The membrane potential controls the membrane currents, but also the membrane currents affect the membrane potential
- The cell can be modelled as a capacitor as there are charges (ions) either side of an insulator (cell membrane) that flow between the sides (membrane currents) until an equilibrium is reached
- As  $Q = CV$  and  $I = \frac{dQ}{dt}$ ,  $\mathbf{I} = \mathbf{C}\frac{d\mathbf{V}}{dt}$  (under the assumption that  $C$  does not vary with time)
- Membrane potential dynamics equation: Let  $C_m$  be the capacitance of the cell membrane,  $V$  be the membrane potential, and  $I_{ext}$  be current corresponding to inputs from other neurons (or experimentally induced currents). Then,  $C_m \frac{dV}{dt} = -I_m = -g_m(V - E_m) + I_{ext}$
- $V_\infty = \text{steady state solution} = \text{value of } V \text{ (and } t) \text{ for which } \frac{dV}{dt} = 0$

- Proposition: The membrane potential dynamics ODE has an analytical solution, namely

$$V(t) = E_m + \exp\left(-\frac{t}{\tau_m}\right) (V(0) - E_m) + \frac{1}{g_m \tau_m} \int_0^t \exp\left(\frac{t'-t}{\tau_m}\right) I_{ext}(t') dt'$$

where  $\tau_m = \frac{C_m}{g_m}$

Proof: Exercise (use formula for first order non-homogeneous ode)

- Proposition: The steady state solution to the membrane potential dynamics ODE (for constant  $I_{ext}$ ) is  $V_\infty = E_m + \frac{I_{ext}}{g_m}$

Proof:  $0 = -g_m(V_\infty - E_m) + I_{ext} \Rightarrow V_\infty = E_m + \frac{I_{ext}}{g_m}$

- This passive membrane model is a reasonable model for the small fluctuations in membrane potential caused by small input currents, but does not capture the fast dynamics seen in action potentials

## 2.3 Active membrane potential dynamics

### 2.3.1 Action Potentials

- In an action potential, the opening and closing of ion channels alters the membrane permeability and so the conductance and so the resting potential. However, membrane conductances were modelled as constant in the passive membrane model

- **Action potential:** Ion channels are voltage gated. Sodium channels and potassium channels are both closed at the resting potential. If the membrane potential is increased enough, then the sodium channels will open creating positive feedback allowing  $Na^+$  ions to chemically diffuse into the cell and so the membrane potential to further increase. Once sufficient positive feedback has occurred, the potassium channels will open allowing  $K^+$  ions to chemically diffuse out of the cell (negative feedback) and the sodium channels will re-close (stopping the **positive feedback**). The membrane potential will then fall back to the resting potential with a slight overshoot (*due to complicated aspects to do with potassium channels*) into being more negative than the membrane potential before decaying back up to the membrane potential

- Although the percentage change in membrane potential (e.g. from a resting potential of  $-70mV$  to a spike of  $+40mV$ ) is large, the membrane potential is small and so only a few ions need to move to cause this change. Thus, neurons do not typically run out of ions (also sodium-potassium pumps provide active transport etc)

### 2.3.2 Hodgkin-Huxley Model

- **Hodgkin-Huxley Model:**  $C_m \frac{dV(t)}{dt} = -g_{leak} [V(t) - E_{leak}] - g_{Na}(V, t) [V(t) - E_{Na}] - g_K(V, t) [V(t) - E_K] + I_{ext}$ , where there are specific values given to  $g_{Na}(V, t)$ ,  $g_K(V, t)$  (which give rise to 2 and 1 further (coupled) ODEs respectively) but all that is detailed below
- The leak term represents the combined effect of all ion channels with fixed conductances (mostly chloride), but the other ion channels can no longer easily be rolled up into a single  $g_m$  in this way as their conductances are no longer constant but rather a function of the current system state

- We can use a Markov chain to model the opening and closing of ion channels at each voltage in order to obtain expressions for  $g_{Na}(V, t)$  and  $g_K(V, t)$ . Let  $A, A^*$  be the fractions of ion channels for some species of ion that are closed or open respectively (deduce that  $A + A^* \equiv 1$ ). Let  $\alpha, \beta$  be the probabilities (at the current value of  $V$ ) of a closed ion channel opening or an opened ion channel closing respectively.

Then, **conductance for this ion species = conductance for this ion species when all channels are open (a constant)  $\times A^*$** . Moreover,  $\frac{dA^*}{dt} = \text{closed ion channels that open} - \text{open ion channels that close} = \alpha(1 - A^*) - \beta A^*$ .

- Proposition: The ion channel Markov model has solution

$$A^*(t) = \frac{\alpha}{\alpha + \beta} + \exp((\alpha + \beta)(-t)) \left( A_0^* - \frac{\alpha}{\alpha + \beta} \right)$$

Proof: Exercise (use formula for first order non-homogeneous ode)

- In reality the probabilities vary with voltage (which varies with time), and each individual ion channel has multiple gates all of which must be open for the channel to be open. If we replace  $\alpha$ ,  $\beta$  with  $\alpha(V)$ ,  $\beta(V)$  (and so forfeit our analytical solution), then we can view our Markov model as a model of the proportion of open gates. As our proportions are really probabilities, **we can multiply the gate figures from the Markov model (as these are independent events) to obtain a new figure for the proportion of open channels**
- **For potassium there are 4 identical gates** (which close at low potential and open at high potential). **Whereas for sodium, there are 3 identical gates** (which close at low potential and open at high potential) **and one gate which works the opposite way** (opens at low potential and closes at high potential); this is what allows sodium channels to open when the potential increases but close again once the potential gets too high, as the channel is only open for the intermediate potentials at which both types of gate are open.

### 2.3.3 Leaky Integrate and Fire Model

- The Hodgkin-Huxley model has no known analytical solution and applying numerical methods can be computationally expensive. Can we create a simpler model that will be less accurate to the underlying biophysics in its implementation but give similar outputs?
- **The leaky integrate and fire model equips the passive membrane dynamics with a threshold reset rule** (if  $V(t) \geq V_{threshold}$ , then set  $V(t) = V_{reset}$  instead) in order to approximate action potentials. We can record the times  $t^{(1)}, \dots$  at which we use the reset rule, and then calculate the inter-spike interval (time between adjacent spikes)  $T$  (or equivalently the firing rate  $f = \frac{1}{T}$ ). Leaky integrate and fire gives spikes of the wrong shape (as they decay instantly), but despite this gives a good approximation of the firing rate.

- Lemma: For constant  $I_{ext}$ , the membrane potential equation (i.e. LIF without the threshold) has analytical solution

$$V(t) = E_m + \exp\left(\frac{-t}{\tau_m}\right) (V(0) - E_m) + \frac{I_{ext}}{g_m} \left(1 - \exp\left(\frac{-t}{\tau_m}\right)\right)$$

Proof: Exercise (evaluate integral of (constant)  $I_{ext}$  in analytical solution of membrane potential equation)

- Proposition: Let  $I_{ext}$  be constant.  $f = 0$  if (and only if)

$$I_{ext} \leq g_m(V_{threshold} - E_m)$$

Proof sketch:

By the lemma,  $V(t) = E_m + \frac{I_{ext}}{g_m} + \exp\left(\frac{-t}{\tau_m}\right) \left(V(0) - E_m - \frac{I_{ext}}{g_m}\right)$ . Recall that  $\forall t \geq 0$ ;  $0 < \exp\left(\frac{-t}{\tau_m}\right) \leq 1$  and is a decreasing function. Assume we are in the situation where the membrane potential is increasing over time (ie  $V(0) - E_m - \frac{I_{ext}}{g_m} < 0$  i.e.  $\frac{I_{ext}}{g_m} > V_0 - E_m$ ).

Then,  $\forall t(\geq 0)$ ;  $V(t) < E_m + \frac{I_{ext}}{g_m}$ .

Thus,  $f = 0 \Leftrightarrow \forall t(\geq 0); V(t) < V_{threshold} \Leftrightarrow I_{ext} \leq g_m(V_{threshold} - E_m)$

- Proposition: Let  $I_{ext}$  be constant. If  $I_{ext} > g_m(V_{threshold} - E_m)$  (i.e.  $f \neq 0$ ), then  $T = \tau_m \ln \left( 1 + \frac{V_{threshold} - V_{reset}}{\frac{I_{ext}}{g_m} - (V_{threshold} - E_m)} \right)$  (recall that  $\tau_m = \frac{C_m}{g_m}$ )

Proof: Consider any  $x$  for which  $V(x) = V_{reset}$ . Then,  $T$  is the value for which  $V(x + T) = V_{threshold}$ . Using the lemma,

$$\frac{V(x+T) - \left(E_m + \frac{I_{ext}}{g_m}\right)}{V(x) - \left(E_m + \frac{I_{ext}}{g_m}\right)} = \frac{\exp\left(\frac{-(x+T)}{\tau_m}\right)}{\exp\left(\frac{-x}{\tau_m}\right)} (1) = \exp\left(\frac{-T}{\tau_m}\right). \text{ Thus,}$$

$$T = -\tau_m \ln \left( \frac{V(x+T) - \left(E_m + \frac{I_{ext}}{g_m}\right)}{V(x) - \left(E_m + \frac{I_{ext}}{g_m}\right)} \right) = \tau_m \ln \left( \frac{V_{reset} - \left(E_m + \frac{I_{ext}}{g_m}\right)}{V_{threshold} - \left(E_m + \frac{I_{ext}}{g_m}\right)} \right) =$$

$$\tau_m \ln \left( 1 + \frac{V_{threshold} - V_{reset}}{\frac{I_{ext}}{g_m} - (V_{threshold} - E_m)} \right)$$

- Proposition: For large  $I_{ext}$ ,  $f \propto I_{ext}$

Proof:  $T \rightarrow \tau_m \ln \left( 1 + \frac{1}{\frac{I_{ext}}{g_m}} \right) \rightarrow \tau_m \frac{1}{\frac{I_{ext}}{g_m}} = \frac{g_m \tau_m}{I_{ext}}$ .  $f = T^{-1} \rightarrow \frac{I_{ext}}{g_m \tau_m} \propto I_{ext}$

## 2.4 Synapses

- Synapse = a junction between the axon and dendrite respectively of two neurons
- Neurons do not physically connect with each other to exchange signals electrically. Instead, **signals are propagated chemically through the release of neurotransmitters which diffuse across the synaptic cleft (synaptic gap) and bind to receptors on the cell membrane of the post-synaptic neuron causing it's ion channels to open or close** (excitatory or inhibitory neurotransmitter respectively)
- Ligand-gated (neurotransmitter-gated) ion channels are different to voltage-gated ion channels

- **Dale's principle:** Each neuron's axon either releases only excitatory or only inhibitory neurotransmitters into all the synapses it is the pre-synaptic neuron to — turned out not to always be true, but it often is true and it is convenient to pretend it always is
- Recall the membrane current equation  $I_m = g_m(V_m - E_m)$  and how we incorporated the time-varying conductances of voltage-gated ion channels. For the membrane current through ligand-gated ion channels (synaptic current) we have  $I_{syn}(t) = g_{syn}(t)(V(t) - E_{syn})$
- The simplest model for the synaptic conductance is to use the Dirac delta function to have conductance only at a single instant in time, but this is not very realistic
- The exponential model of synaptic conductance has it attain its maximum instantaneously (like Dirac) but then decay exponentially. Let  $H(t)$  be the Heaviside step function  $H(t) = \mathbb{1}[t \geq 0]$ .

$g_{syn}(t) = \bar{g}_{syn} \exp\left(\frac{-(t-t^{(i)})}{\tau_{syn}}\right) H(t - t^{(i)})$  where  $t^{(i)}$  is the time of the most recent presynaptic spike

- The exponential model can be made to fit a wider range of neurotransmitters by introducing extra parameters by breaking the falloff term into a weighted sum of two terms (“fast” component and “slow” component) and multiplying in a (single) exponential onset time as well.

$$g_{syn}(t) = \bar{g}_{syn} H(t - t^{(i)}) \left( 1 - \exp\left(\frac{-(t-t^{(i)})}{\tau_{rise}}\right) \right) \left( a \exp\left(\frac{-(t-t^{(i)})}{\tau_{fast}}\right) + (1 - a) \exp\left(\frac{-(t-t^{(i)})}{\tau_{slow}}\right) \right)$$

## 3 Spike patterns

### 3.1 Spike statistics

- Over short time windows, neurons are reasonably well described by a Poisson distribution ( $F = 1$ ). However, over large time windows, spike counts become overdispersed ( $F > 1$ )

- Proposition: If spike times follow a Poisson process, then inter-spike intervals follow an exponential distribution

Proof:

$\Pr(t_{isi} > t)$  = probability that there are zero spikes in the interval  $[0, t] = \frac{(rt)^0 \exp(-rt)}{0!} = \exp(-rt)$ . Thus,  $\Pr(t_{isi} \leq t) = 1 - \exp(-rt)$  which matches the CDF of the exponential distribution with  $\lambda = r$

- The distribution of long inter-spike intervals fit an exponential distribution well. However, (due to the refractory period of neurons (while their membrane potential has dipped below the resting potential after an action

potential they are harder to re-fire)) sufficiently short inter-spike intervals are less likely than longer ones which is not what the exponential distribution predicts — that the time between events fits an exponential distribution reasonably well is actually stronger evidence of the number of events following a Poisson process than the  $F = 1$

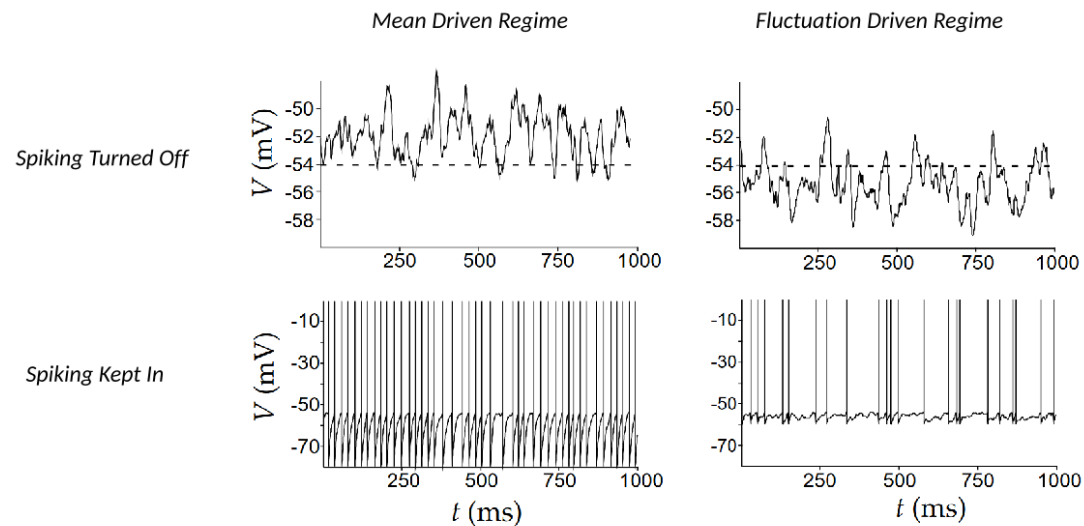
## 3.2 The role of noise

- Hodgkin-Huxley, and leaky integrate and fire are both good models of the reactions of individual biological neurons to constant current injection and both predict regular spiking. Whereas, populations of biological neurons generate Poisson spike patterns, so we would like to know how this biophysically arises

- Mean-driven regime = would spike even without noise = regular spiking with minor jitter from noise — e.g. neurons with constant current injection
- Fluctuation-driven regime = only attains threshold potential due to noise = approximately Poisson distributed spiking — e.g. neurons in the brain

### Mean vs Fluctuation Driven Regimes

---



## 4 The visual system

### 4.1 Retina

#### 4.1.1 Photoreceptors

- Light is detected by photoreceptors (rods and cones). Preliminary processing of the signal from the photoreceptors is performed by bipolar cells followed by retinal ganglion cells, before this “encoded” signal is sent down the optic nerve. However, the retina is arranged such that the photoreceptors are at the back of the eye with the bipolar cells then the ganglion cells on top of them! *The gap in the photoreceptors to allow the optic nerve to reach the ganglion cells (the optic disk) is responsible for the natural blind-spot*
- Information flow along the optic nerve is uni-directional: **the retina sends signals to the brain but do not receive feedback from the brain — this is unusual for a biological neural network**

- Rods have high sensitivity, switch off in high light, slowly switch on in dark light, and have no colour specificity. Cones have low sensitivity, rapidly switch on in high light, switch off in dark light, and are color specific (blue, green, red) — to “see” this for yourself, sit in a darkened room and note that it takes a while for your vision to adjust to see anything (rods are slow to activate) and that when it does your vision is essentially greyscale (rods have no colour specificity)

#### 4.1.2 Receptive fields

- **Retinal ganglion cells come in two types: on-centre off-surround, and off-centre on-surround.** Neuron output for on-centre, off-surround = sum of photoreceptor outputs in centre of receptive field of ganglion cell — sum of photoreceptor outputs in receptive field but outside centre of receptive field. Mutatis mutandis off-centre on-surround

- An (on-centre off-surround) receptive field  $R(x, y)$  can be modelled as a difference of Gaussians  $R(x, y) = \text{centre Gaussian} - \text{surround Gaussian} = \frac{1}{2\pi\sigma_c^2} \exp\left(-\frac{(x-c_x)^2+(y-c_y)^2}{2\sigma_c^2}\right) - \frac{1}{2\pi\sigma_s^2} \exp\left(-\frac{(x-c_x)^2+(y-c_y)^2}{2\sigma_s^2}\right)$  where  $c_x, c_y$  are the coordinates of the very centre of the receptive field and  $\sigma_c, \sigma_s$  are the centre “width” and surround “width” respectively
- Response (firing rate)  $r$  of a neuron with receptive field  $R(x, y)$ :  
 $r = \iint I(x, y) R(x, y) dx dy$  (or more accurately some activation function of the result of this integral)
- **Centre-surround receptive fields act as edge detecting (bandpass filters in the language of signal processing of frequencies) convolutional filters** — extracting the key information from the image in this way before sending it down the optic nerve is important as the optic nerve is heavily limited in the “bitrate” it can accurately transmit at without the signal being drowned out by noise

## 4.2 Thalamus

- The thalamus is the part of the brain that receives input from sensory neurons to relay (but processes prior to relaying) to specialised cortical areas for further processing
- The visual part of the thalamus is the lateral geniculate nucleus (LGN)
- Like the retina, the LGN has centre-surround difference of Gaussian receptive fields
- **Unlike the retina, the LGN receives feedback (from the cortex).**  
In fact, there are more recurrent connections from the cortex to the thalamus than connections from the thalamus to the cortex!
- *Streams of visual information are segregated: Different types of retinal ganglion cell (e.g. color sensitive vs color insensitive) connect to different parts of the LGN which in turn connect to different layers of the visual cortex*

## 4.3 Primary visual cortex

- Many neurons in V1 (primary visual cortex) exhibit orientation tuning — detecting the orientation of edges in V1 after detecting the edges themselves in retina and thalamus is directly comparable to the behaviour of the weights we see learnt by CNNs
- Simple V1 cells are selective to both orientation and phase (i.e. location). Whereas, complex cells are invariant to phase (only selective to orientation)
- A simple cell can be modelled as having receptive field of a Gabor filter (product of Gaussian and sinusoidal):  $R_{simple}(x, y) = \exp\left(-\left(\frac{(x-c_x)^2}{2\sigma_x^2} + \frac{(y-c_y)^2}{2\sigma_y^2}\right)\right) \cos(W[x \cos(\theta) + y \sin(\theta)] - \phi)$  where  $W$  is the frequency of the sinusoidal,  $\theta$  is the preferred orientation, and  $\phi$  is the phase shift

- Let  $r_{complex}, r_{simple}$  be the firing rates of complex and simple cells respectively (and so  $r_{simple}(\phi)$  be the convolution of a Gabor filter (with phase shift  $\phi$ ) over the image).  
Then,  $r_{complex} = [r_{simple}(\phi)]^2 + [(r_{simple}(\phi - \frac{\pi}{2}))]^2$  (i.e. square of response to Gabor filter with  $\cos$  + square of response to Gabor filter with  $\sin$ ) is a good model — *I don't know how squaring is meant to occur biologically*
- In some animals (e.g. most mammals) V1 exhibits a topographic map: neurons with the same orientation preference are clustered together. Whereas, in other animals (e.g. most rodents) neurons are distributed apparently randomly with respect to their orientation selectivity. It is theorised that having a topographic map is only possible if V1 is large enough compared to the retina

## 5 Coding of information

### 5.1 Information theory

- $h(p(x))$  = Surprise of an outcome  $x$  which had probability  $p(x)$   
 $= -\log(p(x))$  — the base of the logarithm does not matter as long as it is consistent but  $\log_2$  is popular for obvious reasons
- Proposition: The less likely an outcome is, the more surprising it is  
Proof:  $\log$  is an increasing function, thus  $-\log$  is a decreasing function
- Proposition: If  $x$  and  $y$  are independent observations, then  
 $h(p(x, y)) = h(p(x)) + h(p(y))$   
Proof: As  $x$  and  $y$  are independent,  $p(x, y) = p(x)p(y)$ . Thus,  
 $h(p(x, y)) = -\log(p(x)p(y)) = -(\log(p(x)) + \log(p(y)))$
- Entropy of a distribution  $p(X) = H(p(x)) =$  expectation of surprise  
 $= -\sum_x p(x) \log(p(x))$

- **Mutual information of two random variables  $X, Y = I(X; y) = H(p(Y)) - \langle H(Y|X = x) \rangle_x$**  = the amount of variation in  $Y$  that is coupled to variation in  $X$

- Proposition:  $I(X; Y) = \sum_{x,y} p(x, y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right)$  — this is called the KL-divergence between  $p(x,y)$  and  $p(x)p(y)$

Proof:

$$\begin{aligned} I(X; Y) &= \sum_y -p(y) \log(y) - \langle \sum_y -p(y|x) \log(p(y|x)) \rangle_x = \\ &= \sum_{x,y} p(x)p(y|x) \log(p(y|x)) - \sum_y p(y) \log(y) = \sum_{x,y} p(x, y) \log(p(y|x)) - \\ &= \sum_{x,y} p(x, y) \log(p(y)) = \sum_{x,y} \log \left( \frac{p(y|x)}{p(y)} \right) = \sum_{x,y} \log \left( \frac{p(x,y)}{p(x)p(y)} \right) \end{aligned}$$

Corollary: Mutual information is symmetric:  $I(y; x) = I(x; y)$

- Proposition (data processing inequality): If  $X$  influences  $Y$  (and  $Y$  does not influence  $X$ ), and  $Y$  influences  $Z$  (and  $Z$  does not influence  $Y$ ), then  $I(X; Z) \leq I(X; Y)$

*Proof: It can be shown that (chain rule for mutual information)*

$$I(X; Z) + I(X; Y|Z) = I(X; Y; Z) = I(X; Y) + I(X; Z|Y) \text{ for any}$$

$X, Y, Z$ . In the data processing context  $X$  and  $Z$  are independent given  $Y$ , so  $I(X; Z|Y) = 0$ . Thus,  $I(X; Z) + I(X; Y|Z) = I(X; Y)$ . Thus, as  $I(X; Y|Z) \geq 0$ ,  $I(X; Z) \leq I(X; Y)$  as required

## 5.2 Efficient coding in the visual system

### 5.2.1 Retina: Whitening

- A whitening filter decorrelates a correlated signal into white noise
- The retina ought to trade off whitening (encoding information more efficiently) and noise removal (not losing too much information to noise). For the distribution of natural images, whitening is a high pass filter and noise removal is a low pass filter. Indeed, the difference of Gaussian receptive field is a bandpass filter

### 5.2.2 V1 and beyond: Generative models

- The retinal ganglion cells removed second-order statistics, but higher-order correlations may remain
- We can represent an image as a matrix of pixels that has been flattened into a vector. Each neuron's receptive field then corresponds to a basis vector and its activation to the coefficient in front of that basis vector, where the image is the sum of these weighted basis vectors. An efficient code corresponds to a basis in which the activations are independent of each other
- The features learnt by PCA do not resemble Gabor filters
- A generative model  $G$  (corresponds to a set of basis vectors) models data  $u$  (an image) as being generated by a set of latent causes  $h$  (corresponds to neural activations). If we can store  $p(u|h, G)$ , then we only need to

represent stimuli as their causes  $h$  as we will be able to recover the stimulus (result of the causes)  $u$

### 5.2.3 V1: Sparse coding models

- Learning:  $G$  = Given a set of stimuli  $U$  (but not causes  $h$ ),  $G$  is learnt by  $G = \arg \max_G (\langle p(u|G) \rangle_{u \in U})$
- Inference: A given stimulus  $u$  is encoded by a given model  $p(u|h, G)$  through maximum likelihood estimation ( $h = \arg \max_h p(h|u, G)$ ). By Bayes' rule  $p(h|u, G) = \frac{p(u|h, G)p(h)}{p(u|G)}$  (and by marginalisation  $p(u|G) = \sum_{h'} p(u|h', G)p(h')$ )
- Generative models describes a family of approaches, it is necessary to chose some particular distribution for  $G$ . As firing rates are approximately exponentially distributed, sparse-coding (using a sparse distribution (such as the exponential distribution)) is a sensible choice

- Sparse-coding generative model: Sparse latent causes:  
 $p(\mathbf{h}_i) = \exp(-|\mathbf{h}_i|)$ . Independent latent causes:  
 $p(\mathbf{h}) = \prod_i p(\mathbf{h}_i)$ . Linear-Gaussian likelihood:  
 $p(\mathbf{u}|\mathbf{h}, \mathbf{B}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}|\mathbf{u} - \mathbf{B}\mathbf{h}|^2\right)$  where  $\mathbf{B}$  is a matrix and  $\mathbf{u}$ ,  $\mathbf{h}$  are vectors
- For sparse coding, inference is (somewhat) biologically plausible as it can be implemented by hill climbing:  
 $\mathbf{h}^{(t+1)} = \mathbf{h}^{(t)} + \epsilon \Delta_{\mathbf{h}} \log(p(\mathbf{h}^{(t)}|\mathbf{u}, \mathbf{B}))$ . Moreover,  
 $\log(p(\mathbf{h}^{(t)}|\mathbf{u}, \mathbf{B})) = -\frac{1}{2\sigma^2}|\mathbf{u} - \mathbf{B}\mathbf{h}|^2 + \log(-\frac{1}{2\sigma^2})$  and so  
 $\Delta_{\mathbf{h}} \log(p(\mathbf{h}^{(t)}|\mathbf{u}, \mathbf{B})) = \frac{1}{\sigma^2} \mathbf{B}^T (\mathbf{u} - \mathbf{B}\mathbf{h})$ . This can be interpreted as  $\mathbf{h}$  being the convergent neural activation of V1 in the network where LGN (with neural activation  $\mathbf{u}$ ) is connected to V1 with weights  $\mathbf{B}^T$  and there are recurrent weights between V1 neurons with weights  $-\mathbf{B}^T \mathbf{B}$

- The features learnt by sparse-coding do resemble Gabor filters iff images are whitened first (which is what the LGN does in the brain!) (otherwise it degenerates to PCA)

#### 5.2.4 V1 and beyond: Predictive coding

- A hierarchical generative model would allow us to explain why the brain has multiple layers of visual processing (V1, V2, ...)
- Intuitively we would expect that (as in a CNN) predictions pass forwards through layers (e.g. V1, V2, ...) from input to output and error signals passes backwards through layers from output to input. However, **predictive coding posits that predictions flow backwards and errors flow forwards** e.g. if layer 2 tells layer 1 to expect a long bar but level 1 finds a short bar then layer 1 will send an error signal to layer 2

- Predictive coding hierarchical generative model (note not probabilistic):  $\mathbf{r}^{(0)} = \text{input}$ ,  $\mathbf{r}^{(i)} = \mathbf{f}(\mathbf{U}^{(i)}\mathbf{r}^{(i+1)}) + \mathbf{n}^{(i)}$  where  $\mathbf{f}$  is an activation function,  $\mathbf{U}^{(i)}$  is a matrix (linear map) and  $\mathbf{n}^{(i)}$  is the error term (to be fed forward), learn  $\mathbf{U}^{(i)}$ s (or infer  $\mathbf{r}^{(n)}$ ) so as to minimize  $\mathbf{n}^{(i)}$ s (e.g. gradient decent)

## 5.3 Population coding

### 5.3.1 Population decoders

- Decoding problem: Give a population of neurons with (noisy) responses and a known model for the encoder of inputs into neural responses, predict the input from the responses
- Decision theory says that maximum likelihood estimation using Bayes' rule is the optimal decoder, but  $\arg \max$  is not very biologically plausible

- We will demonstrate population decoders through the example of decoding the orientation of a stimulus from the responses of a population of orientation-tuned neurons with known preferred orientations
- Winner takes all decoder = preferred orientation of the neuron with the highest firing rate — easily thrown off by noise, especially in small populations
- Population vector decoder =  $\arctan \left( \frac{\sum_i r_i \sin(s_i)}{\sum_i r_i \cos(s_i)} \right)$  = weighted average of orientation preferences weighted by their firing rate, making use of polar coordinates as we are working with orientations
- **The population vector models firing rates as a cos** (with peak centred at the orientation of the stimulus), **but in reality they** (typically) **are a Gaussian** (albeit with the same peak location) **so fall off faster**
- Proposition: If the population is made up of neurons with Gaussian tuning curves (with uniform width) and independent Poisson spike counts, then

the population vector decoder and the  $\arg \max$  decoder coincide (and so the population vector is an optimal decoder)

Proof sketch:

Let  $N$  be the population size. Let  $f_i$  be the probability density function of the Gaussian for neuron  $i$ . Let  $n$  be the population spike count. Let  $s$  be a stimulus.

$f_i(s) = A_i \exp \left( \frac{-(s-s_i)^2}{2\sigma^2} \right)$  where  $s_i$  is the preferred stimulus of neuron  $i$

$$p(n|s) = \prod_{i \in [N]} \frac{(f_i(s))^{n_i} \exp(-f_i(s))}{n_i!}$$

$\max \text{ likelihood} \Leftrightarrow \max \log \text{ likelihood}$

$$\log(p(n|s)) = \sum_{i \in [N]} n_i \log(f_i(s)) - \sum_i f_i(s) - \log(n_i!)$$

Approximate  $\sum_i f_i(s)$  as a constant  $k$

$$\log(p(n|s)) = \sum_{i \in [N]} n_i \log(f_i(s)) - \log(n_i!) + k =$$

$$\sum_{i \in [N]} n_i \log(A_i) - n_i \frac{(s-s_i)^2}{2\sigma^2} - \log(n_i!) + k$$

At the  $\max$  (log) likelihood,  $\frac{d}{ds} = 0$

$\frac{d}{ds} = \sum_{i \in [N]} \frac{-n_i}{2\sigma^2} \frac{d}{ds} \left( (s - s_i)^2 \right) = \sum_{i \in [N]} \frac{-n_i}{2\sigma^2} (2(s - s_i)) =$   
 $-\frac{1}{\sigma^2} \sum_{i \in [N]} s n_i - s_i n_i.$  Thus,  $\frac{d}{ds} = 0 \Rightarrow s = \frac{\sum_{i \in [N]} n_i s_i}{\sum_{i \in [N]} n_i} =$  the population  
 vector

### 5.3.2 Fisher information

- **Fisher information of a given stimulus  $s = I_F(s)$  = expectation of curvature (square of first derivate) of likelihood function =**  
 $\left\langle \left[ \frac{d}{ds} (\log(p(r|s))) \right]^2 \right\rangle = - \left\langle \frac{d^2}{ds^2} (\log(p(r|s))) \right\rangle$
- **Cramer-Rao lower bound (memorise):**  
If  $\hat{s}$  is an unbiased estimator for  $s$  ( $\langle \hat{s} \rangle - s = 0$ ), then  
 $Var(\hat{s}) \geq \frac{1}{I_F(s)}$

- Proposition: **For a population of neurons with Gaussian tuning curves, the Fisher information is inversely proportional to the tuning curve width**

Proof: Exercise

- Although for 1D stimuli narrower tuning curves are better, for 2D stimuli Fisher information is independent of tuning curve width, and for 3D+ broader tuning curves are better

### 5.3.3 Noise correlations

- Repeats of the same stimulus ought to give the same response. However, in reality there is noise. The noisy responses of a pair of neurons may be: independent, or positively correlated, or negatively correlated
- The component of noise that is parallel to the boundary line improves coding by increasing the separation between the neurons' response patterns

- The component of noise that is orthogonal to the boundary line worsens coding by increasing the overlap between the neurons' response patterns
- A noise correlation is called an information limiting correlation iff its correlation matrix is directly proportional to  $f'(s) (f')^T (s)$ . These noises have the effect of translating the tuning curve, which decoders have no choice but to view as a change in stimulus

## 6 Networks of neurons

### 6.1 E-I networks

- A firing rate model only models the average firing rate (over all neurons or over time) instead of individual spike times, this simplifies the model
- A firing rate model assumes a constant input current
- Firing rate model:  $\tau_m \frac{dV_i}{dt} = -(V_i - E_m) + \sum_j w_{ij} \phi(V_j) + u_i(t)$  where  $u_i(t) = \frac{I_{ext,i}(t)}{g_{m,i}}$  is the external input to the neuron and  $\phi$  is the transfer function. This can be rewritten (proof of equivalence is very much out of scope) in a form more like an artificial neural network:  
$$\tau_m \frac{dr_i}{dt} = -r_i + \phi \left( \sum_j w_{ij} r_j + u_i(t) \right)$$

- A synfire chain is a feed-forward spiking network consisting of multiple layers. *Whereas a deep neural network in machine learning is a feed-forward continuous network consisting of multiple layers*
- Recurrent networks are more biologically faithful than feed-forward (e.g. synfire) networks. For example, in the visual system, initial processing is feed-forward but higher-level more evolved perception uses recurrent and top-down processing
- Recurrent networks create dynamical systems with emergent behaviour. Attractors are a likely explanation for memory, oscillations are a likely explanation for communication between neurons, chaotic dynamics are a likely explanation for ability to discern fine grained differences in stimuli
- In the brain there are excitatory and inhibitory neurons (Dales' principle), inhibition allows the brain to stabilize the dynamical systems behaviour of recurrent networks

## 6.2 The ring network

- In V1, excitatory neurons with similar orientation preferences have stronger connections than those with dissimilar orientation preferences
- In V1, inhibitory neurons probably connect fairly indiscriminately (however there is some debate about this)
- Orientation tuning could be explained as a weighted sum of difference of Gaussians receptive fields (i.e. V1 neurons are a feed-forward network with inputs from LGN)
- Orientation tuning could also be explained as V1 neurons recurrently selectively amplify the preferred orientation from an LGN-derived input that is only weakly tuned to orientation

## 6.3 Network stability

- We can view network dynamics as: the firing rates  $r$  of the neurons form a vector describing a point in N-dimensional space, and the evolution  $r(t)$  produces trajectories in this N-dimensional space
- At a fixed point,  $\frac{dr}{dt} = 0$ . Thus, our typical firing rate model  $(\tau_m \frac{dr_i}{dt} = -r_i + \phi(\sum_j w_{ij} r_j + u_i(t)))$  has fixed point  $r = \phi(Wr + I_{ext})$
- If the threshold-linear transfer function  $\phi(x) = \beta[x - \theta]_+ = \beta \max(x - \theta, 0)$  is used, then our firing rate model becomes (when above the threshold) a linear dynamical system  $\frac{dr}{dt} = \frac{1}{\tau_m} [-r + \beta(Wr + I_{ext}(t) - \theta)] = Ar + u(t)$  for suitable definitions of  $A$  and  $u(t)$ . Thus, the fixed point is  $r^* = -A^{-1}u^*$

- Proposition: A linear dynamical evolves according to  $r(t) = r^* + \sum_i c_i v_i \exp(\lambda_i t)$  where  $r^*$  is the fixed point,  $v_i, \lambda_i$  are the eigenvectors and associated eigenvalues, and the  $c_i$ s are determined by  $r(0)$   
 Proof: Out of scope  
 Corollary: A linear dynamical system is stable (attracted towards its fixed point) iff every eigenvector has negative real part. A linear dynamical system is unstable (repelled away from its fixed point) iff it has an eigenvalue with positive real part
- Proposition: If  $\det(A) > 0$  and  $\text{trace}(A) < 0$ , then both eigenvalues are negative and so the network is stable  
 Proof sketch:  $\det(A) = \lambda_1 \lambda_2$  so its sign tells us whether the signs of the eigenvalues agree.  $\text{Trace}(A) = \lambda_1 + \lambda_2$  so is only negative if at least one eigenvalue is

# 7 Plasticity

## 7.1 Single neuron plasticity

- It is necessary to now have a little more detail for how action potentials cross synapses. **The spike causes voltage-gated calcium channels to open in the presynaptic side, in-flow of calcium triggers a pathway leading to the release of vesicles of (excitatory) neurotransmitters from presynaptic membrane**, neurotransmitter diffuses across synaptic cleft and binds to receptors on postsynaptic membrane triggering opening of ligand-gated sodium channels on postsynaptic side
- **Spike rate adaptation is achieved through calcium-gated potassium channels.** Opening potassium channels (which is known to decrease the potential) when calcium flows in has the effect of making it harder to spike based on how many times a neuron has spiked previously

## 7.2 Short-term synaptic plasticity

- Short term plasticity allows the brain to respond to relative changes in sensory inputs instead of their absolute magnitudes
- Facilitation (connection strengthening): Accumulation of calcium on presynaptic side increases probability of (neurotransmitter) vesicle release, synaptic currents increase with repeated firing
- Depression (connection weakening): Supply of vesicles on presynaptic side becomes depleted, synaptic currents decrease with repeated firing
- *It's not a bug, it's a feature!*
- If we model vesicle releases as identical and independent with constant probability within each spike which varies between spikes based on time and spike history, then the number of vesicles released in a given spike follows a Binomial distribution. Moreover, for large enough  $n$ , Binomial becomes approximately either Gaussian ( $p$  not small) or Poisson ( $p$  small)

## 7.3 Long-term synaptic plasticity

- The binding of glutamate to **NMDA** receptors on the post-synaptic neuron triggers the **slow opening of associated calcium channels**, but (in line with Hebb's fire together) **only if the post-synaptic neuron is currently spiking** (as the net negative charge inside a neuron at rest attracts  $Mg^{(2)+}$  ions which jam up the calcium channel as they are too large to pass through)
- The number of AMPA receptors on the cell membrane is regulated by the concentration of calcium in the cell. The binding of glutamate to an AMPA receptor triggers the fast opening of associated sodium channels

- **Potentiation (connection strengthening):** If glutamate binds to NMDA receptors and neurons on both sides of synapse fire together, then calcium channels open. **Accumulation of calcium in post-synaptic cell triggers a pathway leading to the more AMPA receptors on the post-synaptic cell membrane.** Binding of glutamate to AMPA receptors causes sodium channels to open causing larger synaptic currents
- **Depression (connection weakening):** If neurons on endpoints of synapse do not fire in sync, then calcium cannot flow through NMDA-gated channels. Decrease in calcium concentration in post-synaptic cell triggers a pathway leading to removal of AMPA receptors from the post-synaptic cell membrane. Fewer AMPA-gated sodium channels mean smaller synaptic currents

## 7.4 Learning

- Hebbian learning: Co-activation of a group of cells (cell assembly) causes connections between them to strengthen. Subsequent activation of a subset of those cells will thus reactivate the whole group, allowing memories to be retrieved in this associative fashion
- Hebbian learning for a single linear neuron:  $y = w \cdot x$  where  $y$  is a scalar and  $w, x$  are vectors,  $\Delta w = \epsilon y x = \epsilon(w \cdot x)x$  — this is not biologically realistic as here neurons can change sign but biologically (Dale's law) neurons cannot change between being inhibitory and excitatory
- Proposition: If we present  $M$  patterns once each and using Hebb's rule update the weights only at the end (as the average update), then  $\Delta w = \frac{\epsilon}{M} Q w$  where  $Q_{ij} = \sum_{\mu} x_i^{(\mu)} x_j^{(\mu)}$

Proof: For each  $i$ ,  $\Delta w_i = \frac{1}{M} \sum_{\mu} \epsilon y x_i^{(\mu)} = \frac{\epsilon}{M} \sum_{\mu} \sum_j w_j x_j^{(\mu)} x_i^{(\mu)} = \frac{\epsilon}{M} \sum_j w_j \sum_{\mu} x_j^{(\mu)} x_i^{(\mu)} = \frac{\epsilon}{M} \sum_j w_j Q_{ij} = \frac{\epsilon}{M} w \cdot (Q_i)^T$ . Thus,  $\Delta w = \frac{\epsilon}{M} Q w$  as required

Corollary: Hebbian learning corresponds to the linear dynamical system  $\tau \frac{dw}{dt} = Q w$ . Thus, as  $Q$  is a sum of outer products and so must have positive eigenvalues, Hebbian learning has unstable dynamics ( $w$  grows exponentially over time)

- The covariance rule extends Hebb's rule to use as  $Q$  the covariance matrix instead of the correlation matrix, this allows for long-term depression as well as the long-term potentiation of Hebb's

- For both Hebb's rule and the covariance rule, the direction of the weights converges to the eigenvector of  $Q$  with the largest eigenvalue (which for the covariance rule is always the largest principal component of the data, and for Hebb's rule is the largest principal component of the data if the data has mean of zero) however the magnitude of the weights diverges to infinity
- Oja's rule extends Hebb's rule with a normalisation penalty term to stabilize the weights:  $\Delta w_i = \epsilon(x_i y - w_i y^2)$
- Oja's rule has steady state equation  $Qw = (wQw)w$  where  $Q$  is the same as in Hebb's rule. Thus, the steady state  $w$  is (the largest eigenvalue) eigenvector of  $Q$  (as  $wQw$  is a scalar (the eigenvalue)), and so Oja's rule picks out the first principal component the same as the other learning rules (but without the divergence of the other learning rules)

- Inhibitory interactions between neurons in a population can be added so that Oja's rule will lead to each neuron's vector of weights being a different principal component
- If non-linear neurons are used, then we could do different unsupervised learning tasks to PCA