# An Investigation of the Hopfield Network and the Hodgkin-Huxley Model of the Axon

**Candidate 5586A**

Part IIB: 4G3 Computational Neuroscience

## Abstract

In this report we investigate the *Hopfield Network* and the *Hodgkin-Huxley* model of the axon from a mathematical and computational perspective. We begin with a discussion on the content-addressable memory properties of the binary Hopfield Network, providing a proof of the stability of encoded memories and the conditions under which memory recall might be compromised. The discussion focuses on comparing the simulated performance of the network to theoretical predictions. We conclude the section on the Hopfield Network by discussing its practical applications in engineering and its biological faithfulness, proposing possible extensions to improve the latter. The second section proceeds with an explanation of the biological relevance of the design and dynamics of the Hodgkin-Huxley model. We present the equations for the behaviour of a single Hodgkin-Huxley axon and implement them in a simulation. We then conclude the report by analysing the behaviour of the simulated Hodgkin-Huxley neuron when receiving inputs from (i) a non-periodic constant step current and (ii) a periodic train of current pulses of variable period, pulse width and amplitude.

## 1 Hopfield Network

### Introduction

In this section we take a physicist's approach to Neuroscience and describe the properties that arise from the collective behaviour of a brain-like model system composed of a large number of simple processing devices which mimic biological neurons. We focus in particular on the *content-addressable memory* capabilities of this system. By content-addressable memory capabilities we refer to a system capable of storing and retrieving entire memory items on the basis of sufficient partial information. In the specific model system presented in this report, the memories to be stored consist of binary strings which are encoded by the binary states of the processing devices of the system. The storage and retrieval capabilities arise from the existence of stable fixed points in the state space, whose location can be selected through an appropriate choice of model parameters. These fixed points act as point attractors and drive the flow through state space of differently-initialised systems towards them, eventually resulting in the settling of the system around these points. The steady-state values of the processing devices of the system hence define the stored memories, and can be reached from partial information of the memory due to the attraction force of the stable fixed points in the state space.

These systems are not uncommon in physics and, in fact, the behaviour presented by a system with content-addressable memory can be conveniently exemplified by a simple system comprising a particle with frictional damping moving in a potential well with several minima, which correspond to the memories. In the case of the particle, the state space would be continuous as opposed to the discrete state space of our model system. It is actually the case that any physical system whose dynamics in state space are dominated by a substantial number of locally stable states to which they are attracted can be regarded as a general content-addressable memory. If, in addition, the physical system can accommodate any prescribed set of states as the stable states of the system, it becomes a useful means for storing and retrieving memories. In the following sections we will discuss the characteristics of one such system: the *Hopfield Network*.

### 1.1 The Model System

Our model system consists of a binary *Hopfield Network* (Hopfield, 1982) consisting of a collection of $N$ processing devices called *neurons*. As in McCulloch and Pitts (1943), at any particular time, each neuron $i$ has one of two states,

$$r_i(t) \in \{0, 1\} \quad \forall i \in [1, N] \tag{1.1.1}$$

where 0 represents the 'not firing' state and 1 represents the 'firing at maximum rate' state. The state of each neuron changes asynchronously over time according to the following algorithm,

$$r_i(t + \Delta t) = F\left(\sum_{j=1}^{N} W_{ij} r_j(t)\right) \tag{1.1.2}$$

where $F$ is the *Heaviside Step Function* defined as follows[1],

$$F(I) = \begin{cases} 1 & \text{if } I \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.1.3}$$

and $W$ is a matrix of *synaptic strengths* whose elements depend on the memories to be stored. Defining the state of neuron $i$ in memory $m$ to be $r_i^{(m)}$, we select memories such that the binary patterns are *uncorrelated* and *balanced* by imposing that,

$$r_i^{(m)} \perp\!\!\!\perp r_{j\neq i}^{(m)} \perp\!\!\!\perp r_j^{(m'\neq m)} \tag{1.1.4}$$

and,

$$P\left(r_i^{(m)} = 0\right) = P\left(r_i^{(m)} = 1\right) = \frac{1}{2}\,. \tag{1.1.5}$$

We then define $W$ using a *covariance rule* as follows,

$$W_{ij} = \sum_{m=1}^{M} \left(r_i^{(m)} - \frac{1}{2}\right)\left(r_j^{(m)} - \frac{1}{2}\right) \tag{1.1.6}$$

$$W_{ii} = 0 \tag{1.1.7}$$

where $W_{ij}$ represents the strength of the connection between neuron $i$ and neuron $j$ and $W_{ii} = 0$ imposes that there be no *autapses* (i.e. self-connections). With this choice of $W$, the network of binary neurons can be used as a form of general (and error-correcting) content-addressable memory to store $M$ binary strings through the states of the individual neurons in each memory $m$. We note that the synaptic strengths in $W$ are kept fixed and hence this system does not model plasticity.[2]

To prove that the network dynamics of this system yield stable fixed points, we show that we can define a *Lyupanov* or *energy function* over the state space of the network,

$$E(\mathbf{r}) = -\frac{1}{2}\sum_i \sum_{j\neq i} W_{ij} r_i r_j \tag{1.1.8}$$

where $\mathbf{r}$ is a vector containing the states of all the neurons in the network. This energy function is inspired by the Hamiltonian function in *Ising models* (Ising, 1925). Energy functions must be both (i) *non-increasing* and (ii) *lower bounded*, properties which enable the existence of stable fixed points in the state space. It is evident that (ii) is satisfied by the function in Equation 1.1.8 due to its finite domain (i.e. $\mathbf{r}$ can only take values between the all-0 and all-1 state). We must, however, prove that the function is non-increasing, which can be done as follows. We first rewrite Equation 1.1.8 in a more convenient form,

$$\begin{aligned} E(\mathbf{r}(t)) &= -\frac{1}{2}\sum_i \sum_{j\neq i} W_{ij} r_i(t) r_j(t) \\ &= -r_k(t)\sum_{j\neq k} W_{kj} r_j(t) - \frac{1}{2}\sum_{i\neq k}\sum_{j\notin\{i,k\}} W_{ij} r_i(t) r_j(t) \end{aligned} \tag{1.1.9}$$

[1] In some implementations of the Hopfield Network the threshold of each particular neuron $U_i$ (i.e. the value of $I$ above which $F$ is 1) can be selected differently for each neuron. In this report, however, we take $U_i = 0$ for all $i$ for simplicity.

[2] See (Hopfield, 1982) for the properties of a version of the Hopfield Network incorporating Hebbian plasticity.

where we have simply taken one element out of the outer sum and one (identical) element out of the inner sum so that we can better show the effect of the asynchronous update step on the energy function,

$$E(\mathbf{r}(t+\Delta t)) = -r_k(t+\Delta t)\sum_{j\neq k} W_{kj} r_j(t) - \frac{1}{2}\sum_{i\neq k}\sum_{j\notin\{i,k\}} W_{ij} r_i(t) r_j(t). \tag{1.1.10}$$

If we now look at the difference between the value of the energy function before and after updating, we obtain,

$$\Delta E = E(\mathbf{r}(t+\Delta t)) - E(\mathbf{r}(t)) = -[r_k(t+\Delta t) - r_k(t)]\sum_{j\neq k} W_{kj} r_j(t) \tag{1.1.11}$$

and we note that the summation term in the product is the same as the argument of the Heaviside step function in Equation 1.1.2, which we refer to as the *local field* or the *input signal* to a particular neuron and denote as,

$$H_k(t) = \sum_{j\neq k} W_{kj} r_j(t) \tag{1.1.12}$$

or also, since $W_{kk} = 0$,

$$H_k(t) = \sum_{j\neq k} W_{kj} r_j(t) = \sum_{j=1}^{N} W_{kj} r_j(t). \tag{1.1.13}$$

As per the update algorithm in Equation 1.1.2, if the value of $H_k(t)$ exceeds (is below) 0, the state of neuron $k$ at time $t+\Delta t$ will be 1 (0). This property comes in useful to prove that $E$ is non-increasing. To this end, we look at the different values that the expression in Equation 1.1.11 can adopt,

$$\begin{aligned}
\Delta E &= -[r_k(t+\Delta t) - r_k(t)]\sum_{j\neq k} W_{kj} r_j(t) \\
&= -[r_k(t+\Delta t) - r_k(t)] H_k(t) \\
&= \begin{cases} -(1-0)H_k(t) \leq 0 & \text{if } r_k(t) = 0 \text{ and } H_k(t) \geq 0 \\ -(1-1)H_k(t) = 0 & \text{if } r_k(t) = 1 \text{ and } H_k(t) \geq 0 \\ -(0-0)H_k(t) = 0 & \text{if } r_k(t) = 0 \text{ and } H_k(t) < 0 \\ -(0-1)H_k(t) < 0 & \text{if } r_k(t) = 1 \text{ and } H_k(t) < 0 \end{cases}
\end{aligned} \tag{1.1.14}$$

which allows us to say that,

$$\Delta E = E(\mathbf{r}(t+\Delta t)) - E(\mathbf{r}(t)) \leq 0 \tag{1.1.15}$$

which concludes our proof of monotonicity.

After having demonstrated the existence of stable fixed points in the state space of the system, we must now demonstrate how the particular choice of $W$ in Equations 1.1.6 and 1.1.7 leads to the emergence of stable fixed points at the locations defined by the collection of $M$ memories we wish to store. We begin by replacing the definition of $W$ into the expression for the local field of an arbitrary neuron $k$,

$$\begin{aligned}
H_k(t) &= \sum_{j\neq k} W_{kj} r_j(t) = \sum_{j\neq k} r_j(t) \sum_m \left(r_k^{(m)} - \frac{1}{2}\right)\left(r_j^{(m)} - \frac{1}{2}\right) \\
&= \sum_m \left(r_k^{(m)} - \frac{1}{2}\right) \sum_{j\neq k} r_j(t) \left(r_j^{(m)} - \frac{1}{2}\right).
\end{aligned} \tag{1.1.16}$$

Note that we have used the definition of $H_k(t)$ from Equation 1.1.12 so that we can conveniently plug in the definition of $W$ from Equation 1.1.6 without worrying about the autapse terms. We now assume that we are in memory state $\mu$ with $r_j(t) = r_j^{(\mu)}$ for all $j$ to calculate the statistics of the input at this location of the state space. The expression for $H_k(t)$ hence becomes,

$$H_k(t)^{(\mu)} = \overbrace{\left(r_k^{(\mu)} - \frac{1}{2}\right)\sum_{j\neq k} r_j^{(\mu)}\left(r_j^{(\mu)} - \frac{1}{2}\right)}^{\text{signal}} + \underbrace{\sum_{m\neq\mu}\left(r_k^{(m)} - \frac{1}{2}\right)\sum_{j\neq k} r_j^{(\mu)}\left(r_j^{(m)} - \frac{1}{2}\right)}_{\text{noise}} \tag{1.1.17}$$

where we have taken out the element of the outer summation in Equation 1.1.16 corresponding to memory $\mu$ so that we can distinguish between what we call the *signal*, that is, the contribution to the input coming only from the rate

pattern in memory $\mu$, and the *noise*, which is the contribution deriving from all the other memories. We can think of this distinction between signal and noise as the signal being the part of the input that strives to maintain the state of neuron $k$ so that it agrees with that encoded in memory $\mu$, and the noise constituting the part of the input corresponding to the same 'force of attraction' from the other memories.

We now prove that the average input to a particular neuron $k$ in memory $\mu$, averaged over the different memory collections (i.e. definitions of $W$) containing memory $\mu$, is such that the location in state space corresponding to memory $\mu$ is, on average, a stable fixed point,

$$\langle H_k(t)^{(\mu)} \rangle_{\mathbf{r}^{(m)}} = \left( r_k^{(\mu)} - \frac{1}{2} \right) \overbrace{\sum_{j \neq k} r_j^{(\mu)} \left( r_j^{(\mu)} - \frac{1}{2} \right)}^{K_+ \geq 0} + \\ + \sum_{j \neq k} r_j^{(\mu)} (M-1) \underbrace{\left\langle \left( r_k^{(m)} - \frac{1}{2} \right) \right\rangle}_{=0} \underbrace{\left\langle \left( r_j^{(m)} - \frac{1}{2} \right) \right\rangle}_{=0} \quad (1.1.18)$$

where the sum denoted by $K_+$ is always greater than or equal to 0 because the elements of the sum are either 0 or $1/2$ and the average of the noise term is 0 due to the *balance* constraint on the collections of memories. We are hence left with,

$$\langle H_k(t)^{(\mu)} \rangle_{\mathbf{r}^{(m)}} = \left( r_k^{(\mu)} - \frac{1}{2} \right) K_+ \quad (1.1.19)$$

which shows that stable behaviour emerges around stored memories since,

$$\begin{aligned} r_k(t) = r_k^{(\mu)} = 1 \rightarrow \langle H_k(t)^{(\mu)} \rangle_{\mathbf{r}^{(m)}} \geq 0 \rightarrow r_k(t + \Delta t) \approx 1 = r_k^{(\mu)} \\ r_k(t) = r_k^{(\mu)} = 0 \rightarrow \langle H_k(t)^{(\mu)} \rangle_{\mathbf{r}^{(m)}} \leq 0 \rightarrow r_k(t + \Delta t) \approx 0 = r_k^{(\mu)} \end{aligned} \quad (1.1.20)$$

indicating that the firing patterns at a memory location, on average, remain constant over time[3]. This expression hence concludes our proof that the binary patterns stored through our definition of $W$ indeed correspond to stable states of the network. Combining the non-increasing property of the energy function with this result further proves that memory states are local minima of the energy function, as required for stability.

### 1.2 Analytical Error Probability

The proof in Section 1.1 of the stability of encoded memories has only shown stability *in expectation*. We proved that the average local field over different collections of memories of an arbitrary neuron in a particular memory is such that the neuron maintains its state. This, however, does not rule out the existence of collections of memories for which some neurons in encoded memories do not have a local field which maintains their state. To better understand the likelihood of this occurrence we must address the following questions:

(i) What form does the distribution of the local field for a particular neuron adopt?

(ii) What is the probability that an encoded memory is erroneously stored or not stored at all?

The first question can be answered by looking at the expression for the local field of a particular neuron $k$ in an arbitrary memory $\mu$,

$$H_k(t)^{(\mu)} = \overbrace{\left( r_k^{(\mu)} - \frac{1}{2} \right) \sum_{j \neq k} r_j^{(\mu)} \left( r_j^{(\mu)} - \frac{1}{2} \right)}^{\text{signal}} + \underbrace{\sum_{m \neq \mu} \left( r_k^{(m)} - \frac{1}{2} \right) \sum_{j \neq k} r_j^{(\mu)} \left( r_j^{(m)} - \frac{1}{2} \right)}_{\text{noise}}. \quad (1.2.1)$$

We observe how the expression consists of large sums of terms which are independent due to the constraint that stored patterns be *uncorrelated* as per Equation 1.1.4. We can hence use the *Central Limit Theorem* (Fischer, 2010) to approximate the form of the distribution for the local field as a Gaussian, with the approximation improving with the size of the network and the number of encoded memories.

[3]Note that this is not the case if we define the all-0 state as a memory. This memory will be deterministically unstable since all of its neurons receive an input of 0 which, due to our definition of $F$, makes the neurons update to become a 1.

To address the second question, we must find the probability that we choose a collection of memories which does not maintain the states of the neurons in an arbitrary memory $\mu$. To do this we look at the statistics of the local field from two perspectives. The first assumes that the system is in a fixed memory state $\mu$ and investigates the distribution of the input to a particular neuron over the different memory collections that could be stored apart from $\mu$. This is the approach taken in Section 1.1 to prove the stability of the network. We denote this distribution by,

$$P\left(H_k(t)^{(\mu)}|\,\mu',\, r_k^{(\mu)}=0\right)=\mathcal{N}(m_0',\sigma_0'^{\,2}) \tag{1.2.2}$$

$$P\left(H_k(t)^{(\mu)}|\,\mu',\, r_k^{(\mu)}=1\right)=\mathcal{N}(m_1',\sigma_1'^{\,2}) \tag{1.2.3}$$

where $\mu'$ represents the states of all neurons in memory $\mu$ except for neuron $k$. We proved in Section 1.1 that,

$$m_0'=\langle H_k(t)^{(\mu)}|r_k^{(\mu)}=0\rangle_{\mathbf{r}^{(m)}}=-\frac{1}{2}K_+, \tag{1.2.4}$$

$$m_1'=\langle H_k(t)^{(\mu)}|r_k^{(\mu)}=1\rangle_{\mathbf{r}^{(m)}}=\frac{1}{2}K_+ \tag{1.2.5}$$

with the definition of $K_+$ remaining unchanged,

$$K_+=\sum_{j\neq k} r_j^{(\mu)}\left(r_j^{(\mu)}-\frac{1}{2}\right). \tag{1.2.6}$$

We can also find the variance of this distribution as follows,

$$\begin{aligned}
\sigma_0'^{\,2} &= \mathbf{Var}\left(H_k(t)^{(\mu)}|\mu', r_k^{(\mu)}=0\right)\\
&= \mathbf{Var}\left(H_k(t)^{(\mu)}|\,r_k^{(\mu)}=0\right)_{\mathbf{r}^{(m)}}\\
&= \mathbf{Var}\left(\underbrace{-\frac{1}{2}\sum_{j\neq k} r_j^{(\mu)}\left(r_j^{(\mu)}-\frac{1}{2}\right)}_{\text{signal}}+\underbrace{\sum_{m\neq\mu}\left(r_k^{(m)}-\frac{1}{2}\right)\sum_{j\neq k} r_j^{(\mu)}\left(r_j^{(m)}-\frac{1}{2}\right)}_{\text{noise}}\right)_{\mathbf{r}^{(m)}}\\
&= \underbrace{\mathbf{Var}\,(\text{signal})_{\mathbf{r}^{(m)}}}_{=0}+\mathbf{Var}\,(\text{noise})_{\mathbf{r}^{(m)}}\\
&= \mathbf{Var}\,(\text{noise})_{\mathbf{r}^{(m)}}
\end{aligned} \tag{1.2.7}$$

where we used the *uncorrelatedness* of the encoded memories to say that the 'signal' is independent from the 'noise' in the fourth line and used the fact that $\mu'$ is known to say that the variance of the 'signal' is 0. We proceed by finding an expression for the variance of the 'noise' term,

$$\begin{aligned}
\sigma_0'^{\,2} &= \mathbf{Var}\,(\text{noise})_{\mathbf{r}^{(m)}}\\
&= \mathbf{Var}\left(\sum_{m\neq\mu}\left(r_k^{(m)}-\frac{1}{2}\right)\sum_{j\neq k} r_j^{(\mu)}\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}}\\
&= \sum_{j\neq k}\left(r_j^{(\mu)}\right)^2\mathbf{Var}\left(\sum_{m\neq\mu}\left(r_k^{(m)}-\frac{1}{2}\right)\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}}\\
&= (M-1)\sum_{j\neq k}\left(r_j^{(\mu)}\right)^2\underbrace{\mathbf{Var}\left(\left(r_k^{(m)}-\frac{1}{2}\right)\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}}}_{=1/16}\\
&= \frac{M-1}{16}\sum_{j\neq k}\left(r_j^{(\mu)}\right)^2
\end{aligned} \tag{1.2.8}$$

where we found the variance of the product in the penultimate line by noting that the two terms in the product are independent and hence,

$$\mathbf{Var}\left(\left(r_k^{(m)}-\frac{1}{2}\right)\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}}=\mathbf{Var}\left(\left(r_k^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}}\mathbf{Var}\left(\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}} \tag{1.2.9}$$

where,

$$\begin{aligned}\mathbf{Var}\left(\left(r_k^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}} &= \mathbf{Var}\left(\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}} \\ &= \left\langle\left(r_j^{(m)}-\frac{1}{2}\right)^2\right\rangle_{\mathbf{r}^{(m)}} - \left\langle\left(r_j^{(m)}-\frac{1}{2}\right)\right\rangle^2_{\mathbf{r}^{(m)}} \\ &= \frac{1}{4}\cdot\frac{1}{2}+\frac{1}{4}\cdot\frac{1}{2}-\left(-\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\right)^2 \\ &= \frac{1}{4} \end{aligned} \tag{1.2.10}$$

and plugging this result into Equation 1.2.9 gives us,

$$\mathbf{Var}\left(\left(r_k^{(m)}-\frac{1}{2}\right)\left(r_j^{(m)}-\frac{1}{2}\right)\right)_{\mathbf{r}^{(m)}} = \frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}. \tag{1.2.11}$$

We can similarly derive the variance for the distribution with $r_k^{(\mu)}=1$ to also be,

$$\sigma_1'^{\,2} = \frac{M-1}{16}\sum_{j\neq k}\left(r_j^{(\mu)}\right)^2. \tag{1.2.12}$$

We are, however, interested in the distribution of the local field for arbitrary *complete* collections of memories, that is, without fixing one of these at $\mu$. In this way, we can obtain an expression for the probability of choosing a collection of memories for which the network does not present stable behaviour around the encoded memories. This constitutes the second perspective through which we look at the input distribution. This distribution can also be approximated using a Gaussian due to the above argument of the local field being a large sum of independent terms. Using this approximation, we define the conditional distribution for the local field of a neuron $k$ in an *arbitrary* memory $\mu$ as follows,

$$P\left(H_k(t)^{(\mu)}|r_k^{(\mu)}=0\right)=\mathcal{N}(m_0,\sigma_0^2), \tag{1.2.13}$$

$$P\left(H_k(t)^{(\mu)}|r_k^{(\mu)}=1\right)=\mathcal{N}(m_1,\sigma_1^2). \tag{1.2.14}$$

To find $m_0$ and $m_1$, we use the expressions that we derived in Equations 1.2.4 and 1.2.5 and further average over the states of all neurons in memory $\mu$ except $k$ (which is fixed by conditioning). This gives the following expressions,

$$m_0 = \langle\langle H_k(t)^{(\mu)}|r_k^{(\mu)}=0\rangle_{\mathbf{r}^{(m)}}\rangle_{r_j^{(\mu)}} = -\frac{1}{2}\langle K_+\rangle_{r_j^{(\mu)}}, \tag{1.2.15}$$

$$m_1 = \langle\langle H_k(t)^{(\mu)}|r_k^{(\mu)}=1\rangle_{\mathbf{r}^{(m)}}\rangle_{r_j^{(\mu)}} = \frac{1}{2}\langle K_+\rangle_{r_j^{(\mu)}} \tag{1.2.16}$$

where due to the linearity of the expectation operator and the balanced distribution of $r_j^{(\mu)}$ we can write,

$$\langle K_+\rangle_{r_j^{(\mu)}} = (N-1)\left\langle r_j^{(\mu)}\left(r_j^{(\mu)}-\frac{1}{2}\right)\right\rangle_{r_j^{(\mu)}} = \frac{N-1}{4} \tag{1.2.17}$$

giving the following expressions for the means,

$$m_0 = -\frac{1}{2}\langle K_+\rangle_{r_j^{(\mu)}} = -\frac{N-1}{8}, \tag{1.2.18}$$

$$m_1 = \frac{1}{2}\langle K_+\rangle_{r_j^{(\mu)}} = \frac{N-1}{8}. \tag{1.2.19}$$

Remember that we are no longer looking at the mean input to a particular neuron in a *fixed* memory in state space. Instead, $m_0$ and $m_1$ represent the mean input of a particular neuron over all the possible states of the neurons in a memory abiding by the constraints of *balance* and *uncorrelatedness*. These expressions hence represent the means of the distribution of the inputs to neurons in random stored patterns.

We now proceed with the natural step of finding an expression for the variance of these distributions. We define the variances in a similar way to $m_0$ and $m_1$ as the average over the states of the neurons in a particular memory $\mu$ of the expressions for the variance in Equations 1.2.8 and 1.2.12, giving,

$$\begin{aligned}\sigma_0^2 = \mathbf{Var}\left(H_k(t)^{(\mu)}|r_k^{(\mu)}=0\right) &= \left\langle \mathbf{Var}\left(H_k(t)^{(\mu)}|\, r_k^{(\mu)}=0\right)_{\mathbf{r}^{(m)}} \right\rangle_{r_j^{(\mu)}} \\ &= \left\langle \frac{M-1}{16}\sum_{j\neq k}\left(r_j^{(\mu)}\right)^2 \right\rangle_{r_j^{(\mu)}} \\ &= \frac{M-1}{16}\sum_{j\neq k}\underbrace{\left\langle \left(r_j^{(\mu)}\right)^2 \right\rangle_{r_j^{(\mu)}}}_{=1/2} \\ &= \frac{(M-1)(N-1)}{32} \end{aligned} \tag{1.2.20}$$

and similarly,

$$\sigma_1^2 = \frac{(M-1)(N-1)}{32} \tag{1.2.21}$$

Summarising, we have found the following distributions,

$$P\left(H_k(t)^{(\mu)}|\, r_k^{(\mu)}=0\right) = \mathcal{N}(m_0,\sigma_0^2) = \mathcal{N}\left(-\frac{N-1}{8}, \frac{(M-1)(N-1)}{32}\right) \tag{1.2.22}$$

$$P\left(H_k(t)^{(\mu)}|\, r_k^{(\mu)}=1\right) = \mathcal{N}(m_1,\sigma_1^2) = \mathcal{N}\left(\frac{N-1}{8}, \frac{(M-1)(N-1)}{32}\right) \tag{1.2.23}$$

which represent the distribution of the input to a neuron $k$ in a particular memory $\mu$ of a Hopfield Network storing a random collection of balanced and uncorrelated memories.

After a long detour, we now come back to addressing the second of the questions that we presented at the beginning of the section: what is the probability that an encoded memory is erroneously stored or not stored at all? These errors could arise for three different reasons:

(i) The first and most obvious reason could be that the stored patterns are unstable fixed points,

(ii) the second is if the stored patterns are not fixed points in the first place,

(iii) and the last is due to the presence of other spurious point attractors.

Using the expressions above, we can now find the likelihood of selecting a collection of memories such that a particular encoded pattern is not stored by the system. We define a useful measure to represent this quantity and call it the 'error probability', which we denote by,

$$p_e = P\left(r_k^{(\mu)}=0\right) P\left(H_k(t)^{(\mu)} > 0\,|\, r_k^{(\mu)}=0\right) + P\left(r_k^{(\mu)}=1\right) P\left(H_k(t)^{(\mu)} < 0\,|\, r_k^{(\mu)}=1\right) \tag{1.2.24}$$

where this is the probability that a particular bit in a stored memory is not 'stable', that is, the probability that the input to a neuron in a particular memory is not such that its state is maintained. Equation 1.2.24 can also be written as,

$$p_e = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma_0^2}}\int_0^\infty \exp\left(-\frac{(x-m_0)^2}{2\sigma_0^2}\right)dx + \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma_1^2}}\int_{-\infty}^0 \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2}\right)dx \tag{1.2.25}$$

or using the transformation to a standard normal distribution and the cumulative distribution function $\phi$, we can write,

$$\begin{aligned} p_e &= \frac{1}{2}\left(1-\phi\left(\frac{0-m_0}{\sigma_0}\right)\right) + \frac{1}{2}\phi\left(\frac{0-m_1}{\sigma_1}\right) \\ &= \frac{1}{2}\left(1-\Phi\left(\sqrt{\frac{N-1}{2(M-1)}}\right)\right) + \frac{1}{2}\Phi\left(-\sqrt{\frac{N-1}{2(M-1)}}\right) \\ &= \Phi\left(-\sqrt{\frac{N-1}{2(M-1)}}\right). \end{aligned} \tag{1.2.26}$$

We have hence found an expression for the probability of storing a bit incorrectly as a function of the number of neurons $N$ and the number of memories encoded in the network $M$.

To illustrate the dependence of $p_e$ on the number of encoded memories $M$, we set $N = 100$ and plot $p_e$ as a function of $M$ below. We also plot the same graph for a network with $N = 1000$ neurons to see the effect of varying the number of neurons in the network for the same number of memories,

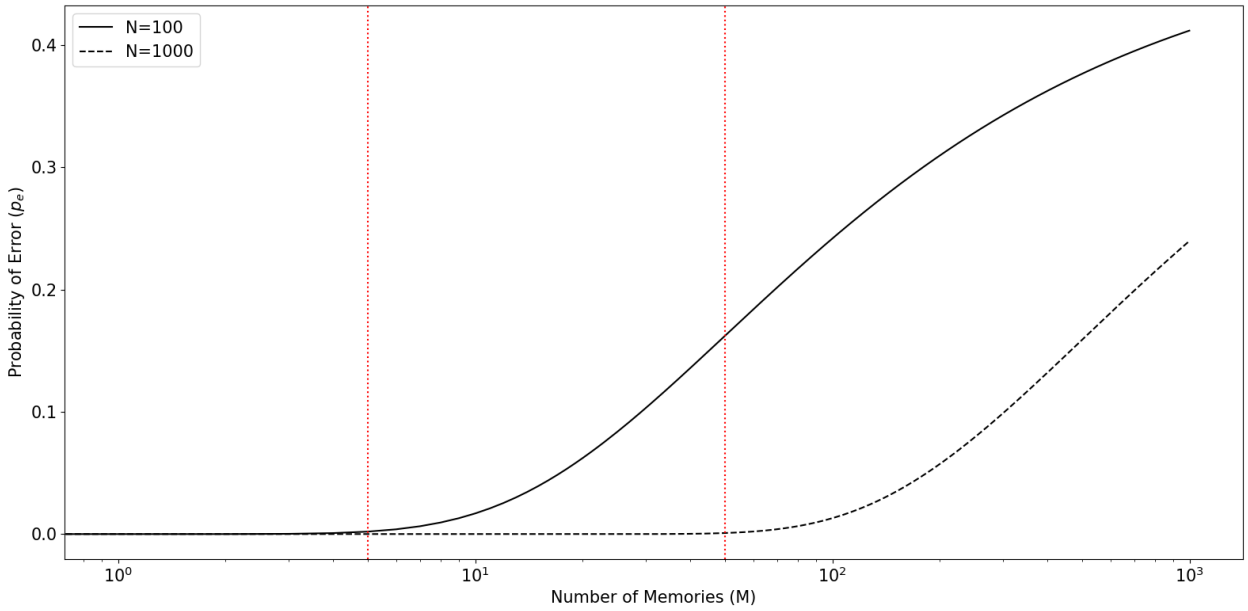
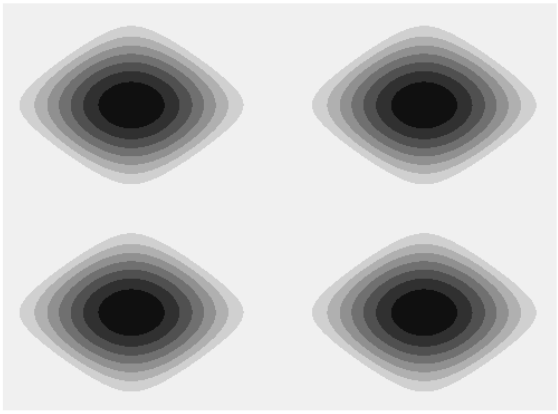


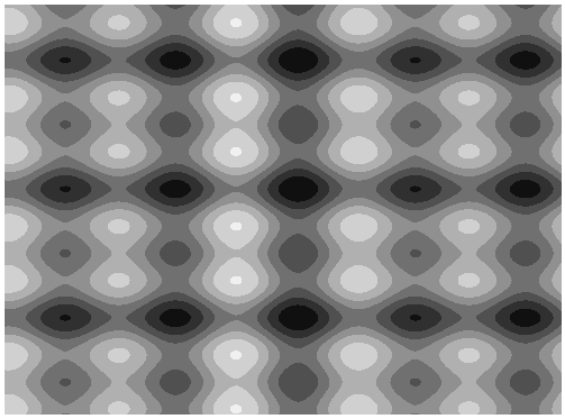

Figure 1: Plots of $p_e$ versus $M$ for (i) $N = 100$ and (ii) $N = 1000$ neurons. We see how the probability of error increases approximately at an inverse exponential rate (linear in the graph due to the logarithmic scale) with the number of encoded memories. We also notice how the number of memories which can be reliably stored with low probability of error are approximately 1/20 of the number of neurons in the network (red lines). Note also how the probability of error for the larger network is everywhere lower than that of the smaller network, with the difference between the two starting off small, then increasing for intermediate values of $M$ and finally decreasing as $M$ becomes large.

Figure 1 shows us that the error probability is strongly dependent on the combination of $N$ and $M$ chosen for the system. In particular, it is important to note that $p_e$ increases with increasing $M$ and that the probability starts off very low and suddenly shoots up at around $M = N/20$ memories. It is also interesting to note that $p_e$ for both networks grows with $M$ until it starts to converges asymptotically to 0.5. This is the error probability for large $M$, which can be found by taking limits in Equation 1.2.26. An error probability of 0.5 occurs when the system cannot recall any memories, since $p_e$ (which can be also interpreted as the probability of a bit in an encoded memory being flipped) is the same as the probability of the individual bit states at the encoding stage. Hence the network ceases to give us any information about the encoded bit pattern and consequently loses its content-addressable memory properties.

The reason for the increase in $p_e$ with $M$ can be seen from the expression for the variance of the local field in Equations 1.2.20 and 1.2.21. Despite the input noise being zero-mean, the variance of the noise term begins to dominate over the signal as $M$ grows, and causes the states of neurons in memories to become unstable. This happens because the influence of other memory point attractors becomes more important as new memories are added to a system with a finite state space whose size is determined by the fixed parameter $N$. This also explains why, if we increase the number of neurons in the network, the probability of error decreases for a fixed number of memories, since the interference between point attractors weakens as the state space becomes larger and the memories become more spread out. A visual representation of this phenomenon is provided in Figure 2 below.



(a) Contour plot of the energy function over a continuous state space for a system storing $M = 4$ memories.



(b) Contour plot of the energy function over a continuous state space for a system storing $M = 15$ memories.

Figure 2: Contour plot of the energy function over the same continuous state space for a system storing (a) $M = 4$ and (b) $M = 15$ memories. Lighter colours represent higher values. We note how the energy function in (a) presents four minima (i.e. memories) which are spread out and do not interfere with each other. On the other hand, the memories in (b) are more closely packed and can be seen to partly fuse together. We also notice the presence of 'spurious attractors' in (b), which manifest themselves as shallower minima between memories. If we picture the state space as a potential field, where each memory is a potential well, and imagine a particle moving through this field, we see that the particle will move until it 'falls' into a potential well (or memory), where it will remain. If, however, we add some random noise to the movement of the particle, it is clear that the likelihood of the particle escaping a potential well and entering a neighbouring one is higher in (b) than in (a). This analogy explains why the error probability increases with increasing $M$ for fixed $N$ in our Hopfield Network.

A useful measure which encompasses the trade-off between the input signal and the noise is the *Signal to Noise Ratio (SNR)* which we define as,

$$\text{SNR} = \frac{\langle \langle H_k(t)^{(\mu)} \rangle_{\mathbf{r}^{(m)}} \rangle^2_{r_j^{(\mu)}}}{\left\langle \mathbf{Var}\left( H_k(t)^{(\mu)} \right)_{\mathbf{r}^{(m)}} \right\rangle_{r_j^{(\mu)}}} = \frac{1}{2} \frac{N-1}{M-1}. \tag{1.2.27}$$

This expression is useful to see when the noise starts to dominate over the signal as we increase $M$. We plot the SNR in decibels, for the same values of $N$ and $M$ as before, to obtain,

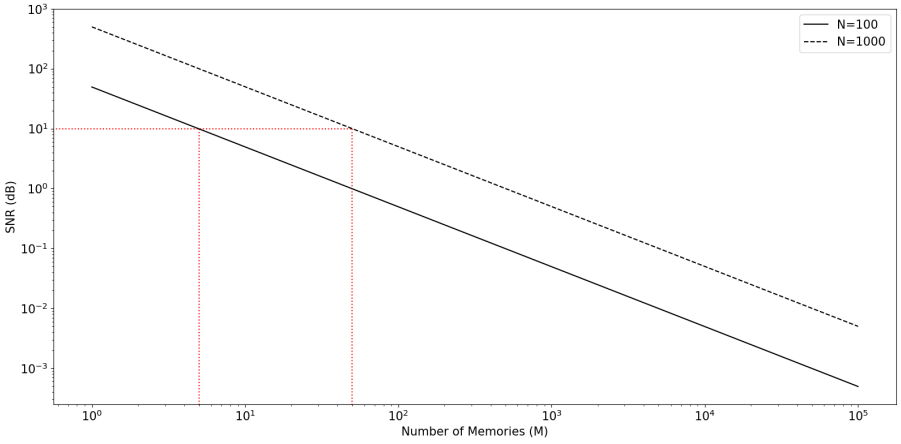



Figure 3: Plots of the SNR versus $M$ for a network with (i) $N = 100$ and (ii) $N = 1000$ neurons. As expected, the SNR decreases linearly (on a logarithmic scale) with increasing $M$ as the noise begins to dominate. We also note how the SNR of the larger network is always higher than that of the smaller network for the same values of $M$. Minimally acceptable values of SNR for reliable information transmission are typically in the range from 10-20dB, which coincides with the values of $M$ close to $N/20$ from before (red lines).

The analysis we have presented in this section may seem rather coarse, abstract and difficult to relate to biology and real neural architectures. However, there are some phenomena, such as the stability of memories and the increasing probability of error with memory count, which can actually be observed in human psychology. It is interesting to realise that the sudden increase in error probability in Figure 1 can be easily related to the 'memory overload' that many of us experience when we are presented with a lot of information at once. The Hopfield Network does not provide us with a precise anatomical and physiological explanation for this occurrence but it shows how, in principle, the behaviour can be recreated with a very simple model of the human brain.

In the next section, we verify that our analytical expression for the error probability coincides with experimental values obtained through an implementation of the Hopfield Network.

### 1.3 Simulated Error Probability

In this section we simulate the Hopfield Network described in Section 1.1 and measure the error probability for different values of $M$. We implement a network with $N = 100$ neurons and generate $M$ random binary memories for storage, along with their corresponding weight matrix $W$. The random character of the memories is faithful to the nature of efficiently encoded real information (e.g. DNA), which also appears random after pre-processing for efficient storage. The scale of $N$, however, is not representative of real neural architectures which tend to have on the order of 100-100,000 cells (Hopfield, 1982). We chose to keep the network small for computational reasons, and because we can still make our claim on the dependence of the error probability on $N$ and $M$ without loss of generality.

For the same values of $M$ as in Section 1.2, the binary memories were generated randomly using the properties of *balance* and *uncorrelatedness* from Equations 1.1.5 and 1.1.4 (i.e. sampling $N$ times from a uniform Bernoulli distribution for each memory). We then computed the corresponding weight matrix $W$ using the definition in Equations 1.1.6 and 1.1.7, and implemented the asynchronous update step in Equation 1.1.2. The error probability was computed by initialising the system randomly at one of the encoded memory states and by averaging, over different memories and different collections of memories, the proportion of times a certain bit in the memory (e.g. the first) changed after one update step. The resulting plot of the error probability versus the number of stored memories is shown in Figure 4 below,

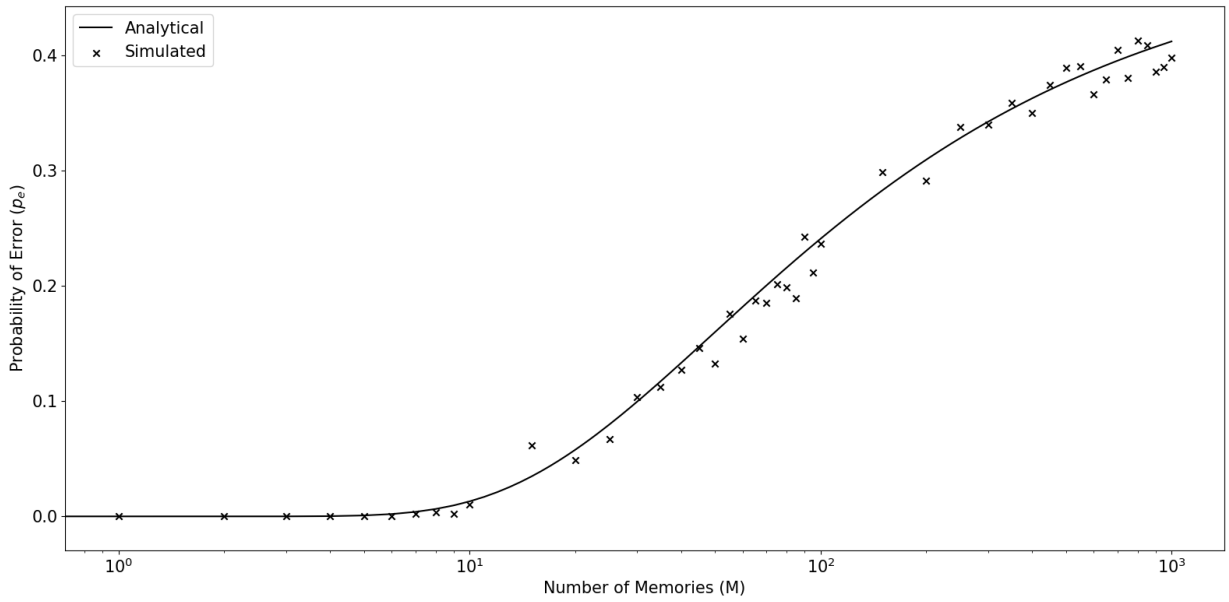



Figure 4: Comparison of the analytical and simulated values of $p_e$ versus $M$ in a network with $N = 100$. The simulated values are the average number of times the first bit in a memory changed after one update step, averaged over 50 stored patterns chosen at random with replacement from each of 50 random collections of memories. We see a good agreement of the theory with the simulated values, confirming the validity of our calculations in Section 1.2.

### 1.4 Recalling Memories from Partial Information

In the previous sections we developed our analysis of the error probability by assuming that the network begins at one of the stored memories. The practically more relevant case is when the network is initialised from a noisy or partial version of one of the originally stored memories and its trajectory through state space is studied. Here lies the true usefulness of content-addressable memory, where we can 'recall' entire memories from their corrupted versions. Relating this behaviour to a biological scenario, we can compare it to the process of *inference* and *pattern completion* that our brain performs when we only see part of an image, for example, or the process of memory retrieval when we are presented with an incomplete memory.

In this section we analyse this scenario through a simulation of the Hopfield Network with $N = 100$ neurons and the same values of $M$ as in the previous sections. Once again we generate $M$ random binary memories from a uniform Bernoulli distribution and compute the corresponding weight matrix as per Equations 1.1.6 and 1.1.7. We also maintain the same implementation of the update algorithm. The difference comes at the time of network initialisation; instead of initialising the network at one of the encoded memories, we initialise it at a corrupted version of a stored pattern. The corruption is performed by flipping each bit in the selected memory randomly with a certain probability, which is denoted as the 'input noise level' $p_{noise}$.

We start by studying the relationship between the error probability and the number of stored memories, as in the previous sections. Here the error probability is the probability that a bit is incorrectly 'recalled', or in other words, the probability of a certain neuron's state being different from that in the original stored memory one update step after initialisation. We plot the error probability obtained through simulations as a function of the number of stored patterns for different input noise levels in Figure 5 below,

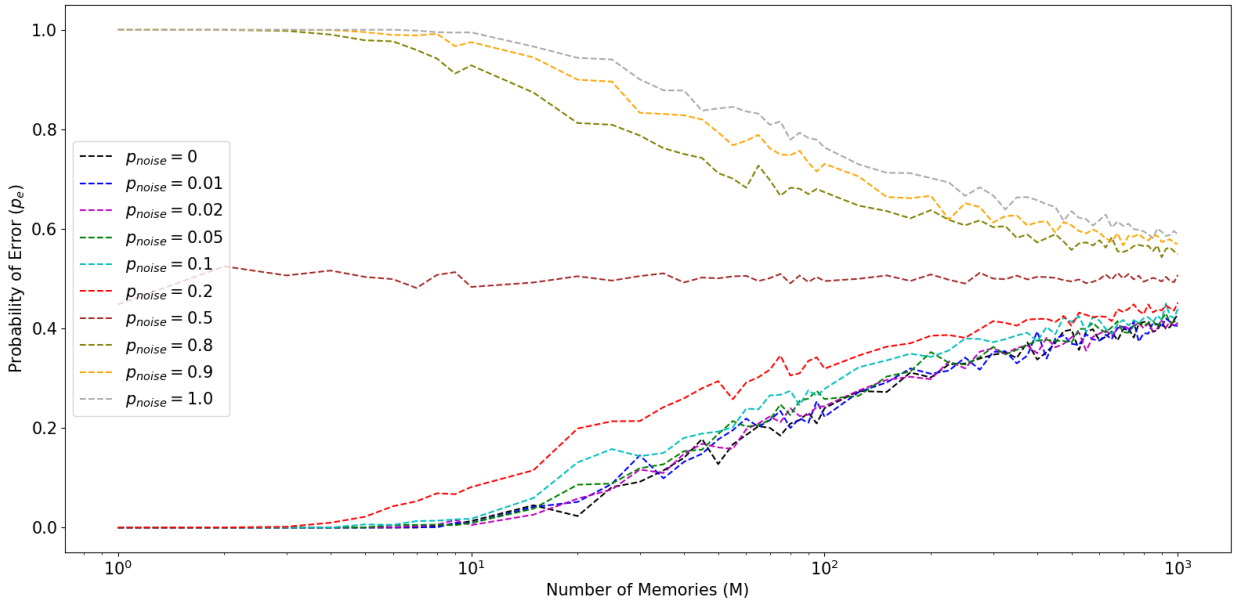



Figure 5: Plots of the error probability versus the number of stored memories for different input noise levels for a network with $N = 100$ neurons initialised at a corrupted memory state. The values of the error probability were obtained by computing the proportion of times that the first neuron in a corrupted memory did not update to its value in the original memory state. The proportion was computed over 50 different memories in a collection, over 50 different collections and 50 different memory corruptions for each memory. The plots for low values of the input noise ($p_{noise} < 0.5$) are similar to the plots in Figure 4, since the probability of starting far away from the original memory is very low. Nevertheless, we observe a slight increase in the error probabilities with input noise. For $p_{noise} = 0.5$ we observe an interesting behaviour in which the error probability remains constant at $p_e = 0.5$ with increasing $M$. Lastly, for $p_{noise} > 0.5$ we observe the opposite behaviour to the plots for low noise. Here the error probability starts off close to 1 and then decreases with increasing $M$. We, however, continue to observe the pattern of higher error probabilities for higher values of $p_{noise}$. All plots tend towards $p_e = 0.5$ for large $M$.

The behaviour in Figure 5 can be explained by deriving an analytical expression for the error probability, much as in Section 1.2. We begin by deriving an expression for the input distribution, which we will then use to derive the expression for the error probability. We wish to find an expression for the distribution over both collections of memories and different 'corrupt' network initialisations of the local field of a particular neuron when the network is in its initial 'corrupted' memory state. Once again, we use the Gaussian approximation from the previous sections since the collections of memories are still balanced and uncorrelated and the individual bits of the corrupted memory are also independent (since the bit flipping is performed independently for each bit with probability $p_{noise}$), making the local field a sum of independent terms as before. We hence define the conditional distribution for the local field of a neuron $k$ in an arbitrary corrupted memory $\widetilde{\mu}$ as follows,

$$P\left(H_k(t)^{(\widetilde{\mu})}|r_k^{(\mu)}=0\right)=\mathcal{N}(\widetilde{m}_0,\widetilde{\sigma}_0^2), \tag{1.4.1}$$

$$P\left(H_k(t)^{(\widetilde{\mu})}|r_k^{(\mu)}=1\right)=\mathcal{N}(\widetilde{m}_1,\widetilde{\sigma}_1^2). \tag{1.4.2}$$

We begin by deriving an expression for the means $\widetilde{m}_0$ and $\widetilde{m}_1$ using the definition of the local field given in Equation 1.1.16,

$$H_k(t)=\sum_m\left(r_k^{(m)}-\frac{1}{2}\right)\sum_{j\neq k}r_j(t)\left(r_j^{(m)}-\frac{1}{2}\right) \tag{1.4.3}$$

where we now assume that, instead of being in one of the encoded memory states as before, we are in one of the initial 'corrupted' memory states $\widetilde{\mu}$ with $r_j(t)=r_j^{(\widetilde{\mu})}$, which gives the following expression for the input,

$$H_k(t)^{(\widetilde{\mu})}=\overbrace{\left(r_k^{(\mu)}-\frac{1}{2}\right)\sum_{j\neq k}r_j^{(\widetilde{\mu})}\left(r_j^{(\mu)}-\frac{1}{2}\right)}^{\text{signal}}+\underbrace{\sum_{m\neq\mu}\left(r_k^{(m)}-\frac{1}{2}\right)\sum_{j\neq k}r_j^{(\widetilde{\mu})}\left(r_j^{(m)}-\frac{1}{2}\right)}_{\text{noise}} \tag{1.4.4}$$

and if we average over all the memories except $\mu$, as before, we obtain,

$$\langle H_k(t)^{(\widetilde{\mu})}\rangle_{\mathbf{r}^{(m)}}=\left(r_k^{(\mu)}-\frac{1}{2}\right)\overbrace{\sum_{j\neq k}r_j^{(\widetilde{\mu})}\left(r_j^{(\mu)}-\frac{1}{2}\right)}^{\widetilde{K}}+\sum_{j\neq k}r_j^{(\widetilde{\mu})}(M-1)\underbrace{\left\langle\left(r_k^{(m)}-\frac{1}{2}\right)\right\rangle}_{=0}\underbrace{\left\langle\left(r_j^{(m)}-\frac{1}{2}\right)\right\rangle}_{=0} \tag{1.4.5}$$

which simplifies to,

$$\langle H_k(t)^{(\widetilde{\mu})}\rangle_{\mathbf{r}^{(m)}}=\left(r_k^{(\mu)}-\frac{1}{2}\right)\widetilde{K}. \tag{1.4.6}$$

We now further average over different memory patterns $r_j^{(\mu)}$ and over different 'corrupted' initialisations $r_j^{(\widetilde{\mu})}$ (both not including neuron $k$) to obtain,

$$\left\langle\langle H_k(t)^{(\widetilde{\mu})}\rangle_{\mathbf{r}^{(m)}}\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}}=\left\langle\left(r_k^{(\mu)}-\frac{1}{2}\right)\widetilde{K}\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}} \tag{1.4.7}$$

which can be developed as follows,

$$\begin{aligned}
&=(N-1)\Big\langle\left(r_k^{(\mu)}-\frac{1}{2}\right)r_j^{(\widetilde{\mu})}\left(r_j^{(\mu)}-\frac{1}{2}\right)\Big\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}}\\
&=(N-1)\Big\langle(1-p_{noise})\left(r_k^{(\mu)}-\frac{1}{2}\right)r_j^{(\mu)}\left(r_j^{(\mu)}-\frac{1}{2}\right)+\\
&\qquad+p_{noise}\left(r_k^{(\mu)}-\frac{1}{2}\right)\overline{r}_j^{(\mu)}\left(r_j^{(\mu)}-\frac{1}{2}\right)\Big\rangle_{r_j^{(\mu)}}
\end{aligned}$$

where $\overline{r}_j^{(\mu)}$ denotes the flipped version of $r_j^{(\mu)}$. Now using the statistics of memory $\mu$ gives us,

$$\begin{aligned}
=(N-1)\Bigg[&\frac{1}{2}\left((1-p_{noise})\left(r_k^{(\mu)}-\frac{1}{2}\right)\left(\frac{1}{2}\right)+p_{noise}\left(r_k^{(\mu)}-\frac{1}{2}\right)\cdot 0\right)+\\
&+\frac{1}{2}\left((1-p_{noise})\left(r_k^{(\mu)}-\frac{1}{2}\right)\cdot 0+p_{noise}\left(r_k^{(\mu)}-\frac{1}{2}\right)\left(-\frac{1}{2}\right)\right)\Bigg]
\end{aligned}$$

and simplifying,

$$= (N-1)\left[\frac{1}{4}(1-p_{noise})\left(r_k^{(\mu)}-\frac{1}{2}\right)-\frac{1}{4}p_{noise}\left(r_k^{(\mu)}-\frac{1}{2}\right)\right]$$
$$= \frac{N-1}{4}(1-2p_{noise})\left(r_k^{(\mu)}-\frac{1}{2}\right). \tag{1.4.8}$$

Hence we have found that,

$$\widetilde{m}_0 = \left\langle \langle H_k(t)^{(\widetilde{\mu})} | r_k^{(\mu)}=0\rangle_{\mathbf{r}^{(m)}}\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}} = -\frac{N-1}{8}(1-2p_{noise}), \tag{1.4.9}$$

$$\widetilde{m}_1 = \left\langle \langle H_k(t)^{(\widetilde{\mu})} | r_k^{(\mu)}=1\rangle_{\mathbf{r}^{(m)}}\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}} = \frac{N-1}{8}(1-2p_{noise}). \tag{1.4.10}$$

The variance can be found as in Section 1.2 by taking the average over $r_j^{(\mu)}$ and $r_j^{(\widetilde{\mu})}$ of the variance of the local field with respect to $\mathbf{r}^{(m)}$,

$$\begin{aligned}
\widetilde{\sigma}_0^2 &= \mathbf{Var}\left(H_k(t)^{(\widetilde{\mu})}|r_k^{(\mu)}=0\right)\\
&= \left\langle \mathbf{Var}\left(H_k(t)^{(\widetilde{\mu})}|\, r_k^{(\mu)}=0\right)_{\mathbf{r}^{(m)}}\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}}\\
&= \left\langle \frac{M-1}{16}\sum_{j\neq k}\left(r_j^{(\widetilde{\mu})}\right)^2\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}}\\
&= \frac{M-1}{16}\sum_{j\neq k}\left\langle \left(r_j^{(\widetilde{\mu})}\right)^2\right\rangle_{r_j^{(\mu)},r_j^{(\widetilde{\mu})}}\\
&= \frac{M-1}{16}\sum_{j\neq k}\underbrace{\left\langle (1-p_{noise})\left(r_j^{(\mu)}\right)^2+p_{noise}\left(\overline{r}_j^{(\mu)}\right)^2\right\rangle_{r_j^{(\mu)}}}_{=1/2}\\
&= \frac{(M-1)(N-1)}{32}
\end{aligned} \tag{1.4.11}$$

and similarly,

$$\widetilde{\sigma}_1^2 = \frac{(M-1)(N-1)}{32}. \tag{1.4.12}$$

Summarising, we have found the following distributions,

$$P\left(H_k(t)^{(\widetilde{\mu})}|\, r_k^{(\mu)}=0\right) = \mathcal{N}\left(-\frac{N-1}{8}(1-2p_{noise}), \frac{(M-1)(N-1)}{32}\right), \tag{1.4.13}$$

$$P\left(H_k(t)^{(\widetilde{\mu})}|\, r_k^{(\mu)}=1\right) = \mathcal{N}\left(\frac{N-1}{8}(1-2p_{noise}), \frac{(M-1)(N-1)}{32}\right). \tag{1.4.14}$$

Using these distributions we can now define the error probability as,

$$p_e = P\left(r_k^{(\mu)}=0\right)P\left(H_k(t)^{(\widetilde{\mu})}>0\,|\, r_k^{(\mu)}=0\right)+P\left(r_k^{(\mu)}=1\right)P\left(H_k(t)^{(\widetilde{\mu})}<0\,|\, r_k^{(\mu)}=1\right) \tag{1.4.15}$$

which is the probability that a particular neuron in the corrupt initial memory state receives an input which does not drive it towards the original memory in the next update step. As in Section 1.2, this can also be expressed as,

$$p_e = \frac{1}{2}\frac{1}{\sqrt{2\pi\widetilde{\sigma}_0^2}}\int_0^\infty \exp\left(-\frac{(x-\widetilde{m}_0)^2}{2\widetilde{\sigma}_0^2}\right)dx+\frac{1}{2}\frac{1}{\sqrt{2\pi\widetilde{\sigma}_1^2}}\int_{-\infty}^0 \exp\left(-\frac{(x-\widetilde{m}_1)^2}{2\widetilde{\sigma}_1^2}\right)dx \tag{1.4.16}$$

and in terms of the cumulative distribution function for Gaussians $\phi$,

$$\begin{aligned}
p_e &= \frac{1}{2}\left(1-\phi\left(\frac{0-\widetilde{m}_0}{\widetilde{\sigma}_0}\right)\right)+\frac{1}{2}\phi\left(\frac{0-\widetilde{m}_1}{\widetilde{\sigma}_1}\right)\\
&= \frac{1}{2}\left(1-\phi\left((1-2p_{noise})\sqrt{\frac{N-1}{2(M-1)}}\right)\right)+\frac{1}{2}\phi\left((2p_{noise}-1)\sqrt{\frac{N-1}{2(M-1)}}\right)\\
&= \phi\left((2p_{noise}-1)\sqrt{\frac{N-1}{2(M-1)}}\right).
\end{aligned} \tag{1.4.17}$$

To verify the correctness of our calculations, we overlay the plots of $p_e$ versus $M$ in Figure 5 on those computed using Equation 1.4.17 to obtain,

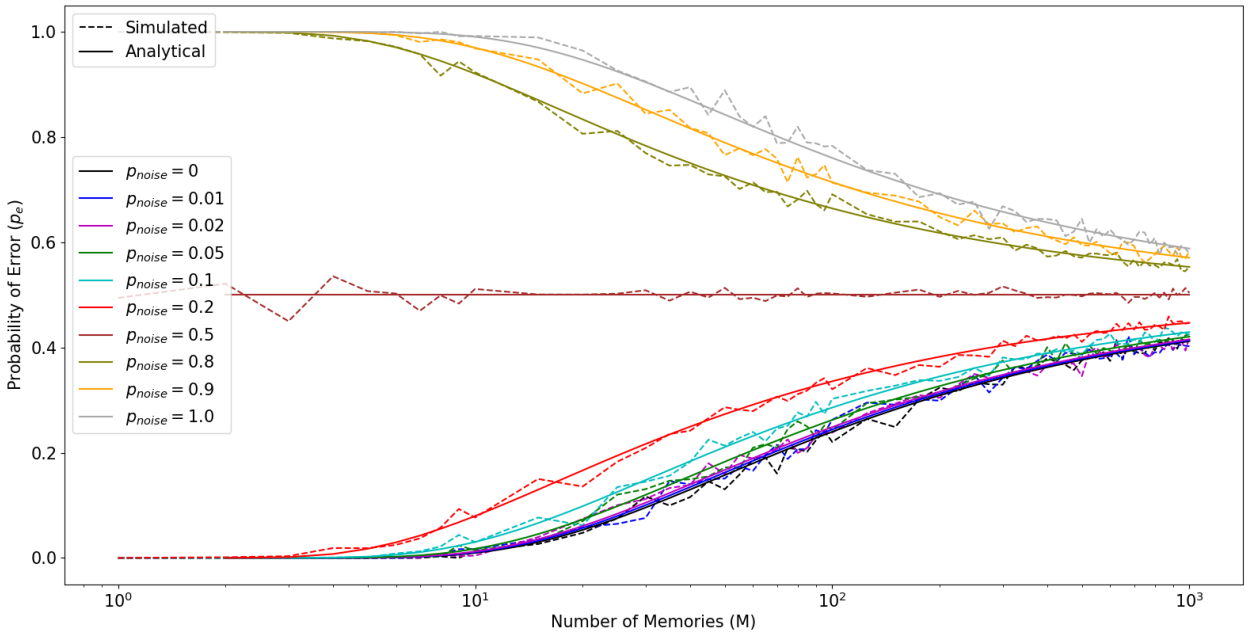



Figure 6: Comparison of the analytical and simulated values of the error probabilities for different numbers of stored memories and noise values for a network with $N = 100$ neurons. The simulated values were obtained in the same way as for Figure 5. We observe a good agreement between the simulated and theoretical values, hence confirming the accuracy of our calculations.

Both Figures 5 and 6 have shown us that the plots of error probability versus $M$ for noisy network initialisations differ considerably from the results of the previous sections, where the network was initialised at one of the encoded memory states. Mathematically, the difference can be explained by comparing the expressions for $p_e$ in Equations 1.2.26 and 1.4.17 which differ by a factor of $(1 - 2p_{noise})$ in the argument of the cumulative distribution function. This factor affects the behaviour of the error probability differently for $p_{noise} < 0.5$, $p_{noise} > 0.5$ and $p_{noise} = 0.5$.

For $p_{noise} < 0.5$, higher values of the input noise give rise to higher error probabilities; this was expected, since the further the initial state of the network is from the original memory (in terms of *Hamming distance*[4]), the less likely it is for a particular neuron in the network to return to its state in the original memory. This is due to two main reasons: firstly, the random bit flipping at the corruption stage can make the network approach another stable point in the state space, where this could be either another encoded memory or a spurious attractor. This would cause the network to tend towards this closer stable state instead of towards the original one. Secondly, this happens because network intialisations which are further away from the original memory feel less attraction from that particular memory. This decreases the proportion of times that a neuron in the network updates towards the original memory, hence increasing the error probability. Looking at this behaviour from a mathematical perspective, it can be seen from Equations 1.4.9 and 1.4.10 that higher values of $p_{noise}$ shift the input mean towards 0 for both $r_k^{(\mu)} = 0$ and $r_k^{(\mu)} = 1$. This means that as the initial state becomes more 'corrupt', the input mean moves further away from values which favour stability towards the original memory, making it more likely for the input to adopt a value which drives the neuron *away* from its state in the original memory. This explains the decrease in 'attraction' from an analytical perspective.

[4]The *Hamming distance* between two strings of equal length is the number of positions at which the corresponding symbols are different.

For $p_{noise} > 0.5$, we observed that the error probability starts at values close to 1 and then decreases with increasing $M$. This might seem puzzling at first, but it can be easily explained by noting a subtle peculiarity of the Hopfield Network. In the proof of the stability of an encoded memory that we presented in Section 1.1, we found that the average input to a neuron $k$ for a network in memory state $\mu$ is such that the state of neuron $k$ remains unchanged, hence favouring stability around the memory, in expectation. It turns out that, due to the symmetry in the definition of the network, the same is also true for the *flipped* memory state $\overline{\mu}$. If the network finds itself in a state identical to that of one of the encoded memories, but with all the states of its neurons flipped, it will present stable behaviour around that point in state space in the same way as it does around the original memory state.

The above property can be proven analytically by analysing the expression for the input to a neuron $k$ in a network at a *flipped* memory state with $r_j(t) = r_j^{(\overline{\mu})}$ as follows,

$$H_k(t)^{(\overline{\mu})} = \left(r_k^{(\mu)} - \frac{1}{2}\right) \sum_{j \neq k} r_j^{(\overline{\mu})} \left(r_j^{(\mu)} - \frac{1}{2}\right) + \sum_{m \neq \mu} \left(r_k^{(m)} - \frac{1}{2}\right) \sum_{j \neq k} r_j^{(\overline{\mu})} \left(r_j^{(m)} - \frac{1}{2}\right) \tag{1.4.18}$$

which can be rewritten using the definition $r_j^{(\overline{\mu})} = 1 - r_j^{(\mu)}$ as,

$$H_k(t)^{(\overline{\mu})} = \left(r_k^{(\mu)} - \frac{1}{2}\right) \sum_{j \neq k} \left(1 - r_j^{(\mu)}\right) \left(r_j^{(\mu)} - \frac{1}{2}\right) + \sum_{m \neq \mu} \left(r_k^{(m)} - \frac{1}{2}\right) \sum_{j \neq k} \left(1 - r_j^{(\mu)}\right) \left(r_j^{(m)} - \frac{1}{2}\right)$$

and taking averages of this expression with respect to the other memories $\mathbf{r}^{(m)}$ gives us,

$$\langle H_k(t)^{(\overline{\mu})} \rangle_{\mathbf{r}^{(m)}} = \left(r_k^{(\mu)} - \frac{1}{2}\right) \overbrace{\sum_{j \neq k} \left(1 - r_j^{(\mu)}\right) \left(r_j^{(\mu)} - \frac{1}{2}\right)}^{K_- \leq 0} + \\ + \sum_{j \neq k} \left(1 - r_j^{(\mu)}\right) (M-1) \underbrace{\left\langle \left(r_k^{(m)} - \frac{1}{2}\right) \right\rangle}_{=0} \underbrace{\left\langle \left(r_j^{(m)} - \frac{1}{2}\right) \right\rangle}_{=0} \tag{1.4.19}$$

hence giving,

$$\langle H_k(t)^{(\overline{\mu})} | r_k^{(\mu)} = 0 \rangle_{\mathbf{r}^{(m)}} = -\frac{1}{2} K_- \tag{1.4.20}$$

$$\langle H_k(t)^{(\overline{\mu})} | r_k^{(\mu)} = 1 \rangle_{\mathbf{r}^{(m)}} = \frac{1}{2} K_- \tag{1.4.21}$$

which is equivalent to,

$$\langle H_k(t)^{(\overline{\mu})} | r_k^{(\overline{\mu})} = 1 \rangle_{\mathbf{r}^{(m)}} = -\frac{1}{2} K_- \geq 0, \tag{1.4.22}$$

$$\langle H_k(t)^{(\overline{\mu})} | r_k^{(\overline{\mu})} = 0 \rangle_{\mathbf{r}^{(m)}} = \frac{1}{2} K_- \leq 0. \tag{1.4.23}$$

This proves that the means of the input to an arbitrary neuron $k$ in a flipped version of an original memory $\overline{\mu}$ are such that the state of the neuron remains unchanged, hence preserving the flipped bit pattern[5]. This proof extends to the flipped versions of all the encoded memories and shows that, in expectation, they will also be stable points in state space.

The above result helps us explain why we observe a decrease in the error probability with increasing number of memories for $p_{noise} > 0.5$ in Figures 5 and 6. High values of $p_{noise}$ cause the corrupted version of the memory to be closer to the *flipped* memory state than to the original one, and since these flipped memory states also act as point attractors, they drive the network towards them. For low values of $M$, the flipped versions of the memories are easier to recall due to there being less interference from neighbouring memories. This causes the error probability to be close to 1 since nearly all the neurons update to become 'flipped' compared to the original memory. However, as the number of stored memories increases, it becomes more difficult for the flipped memories to be recalled and hence the error probability falls as it becomes more likely for the neurons not to update towards the flipped memories - just like it becomes more difficult for corrupted versions of the original memories to return to their original state with increasing $M$. This explains the symmetrical plots in the figures, as we are essentially observing the same behaviour, only that in one case we are dealing with the original memories and in the other with their flipped counterparts. This symmetry can also be seen analytically by replacing $p_{noise}$ with $1 - p_{noise}$ in the expression for $p_e$ in Equation 1.4.17.

[5]Note that this is not true if the original memory is the all-1 state. If this is the case, the flipped version of the memory (i.e. the all-0 state) will be deterministically unstable as any of its neurons will always receive 0 as their input, which causes them to update to a 1. Interestingly, even though the all-0 state is unstable as an original memory, its flipped counterpart will be stable if we choose to store it regardless.

For $p_{noise} = 0.5$ we noticed that the error probability remains constant with increasing $M$. This input noise corresponds to, in expectation, flipping half of the bits and leaving the rest as in the original memory. This means that the network will be initialised around a state which is halfway between an original memory and its flipped counterpart, which are both stable points in the state space. It is hence easy to see that half of the time the network will be updated towards the original memory and in the other half it will be updated towards the flipped version. This is because the system responds to an ambiguous starting state by a statistical choice between the memory states it most resembles. This behaviour is what causes the error probability to remain constant at 0.5 regardless of $M$.

Lastly, we discuss why all the error probabilities tend towards 0.5 for large $M$. The reason is very similar to why the plots in Figure 4 also tends towards 0.5. It happens due to the fact that, at large values of $M$, the memory states become very closely packed and begin to overlap in the finite state space. This makes it very difficult for the network to settle at a single memory state and causes it to continuously update towards different memory states. Since the memory states obey the constraint of balance from Equation 1.1.5, the network will update towards memories with a 1 in the arbitrary position $k$ in half of the updates and towards memories with a 0 in the other half, causing the error probability to tend towards 0.5.

### 1.5 Summary and Extensions

In this section we have provided an in-depth analysis of the content-addressable properties of the binary Hopfield Network. We started by proving the existence of stable fixed points in the state space of the network through the definition of a lower-bounded and non-increasing energy function. We continued with a proof that we can encode memories into the network through the definition of an appropriate weight matrix and that these memories correspond to stable fixed points in the network state space (in expectation). We then concluded with a discussion of the probability with which we select a collection of memories for which (i) we update a neuron *away* from its state at an encoded memory at which we initialise the network and (ii) we update a neuron *away* from an encoded memory if we initialise the network at a corrupt version of the memory.

The binary Hopfield Network is clearly a very simple and rudimentary model of the human brain and may seem to have little biological relevance. However, the emergence of content-addressable memory properties in the collective behaviour of its constituent neurons is indeed impressive as it shows that we can recreate one of the most important features of the human brain through very simple concepts. The underlying mechanism is so simple that it has been proposed to employ it in the design of memory chips for better error correction and reliability of information storage. There have been several attempts to improve the biological relatability of the Hopfield Network, one designed by John Hopfield himself in his follow-up paper (Hopfield, 1984) where he uses a more realistic smooth activation function. This extension does not involve too much additional complexity and proves that the content-addressable memory properties extend to this more realistic model as well. There is also the effect of changing the firing threshold for each neuron which has not been discussed in this report and could be an interesting topic for further work. Lastly, I also encourage the investigation of the effect of introducing Hebbian Learning (Hebb, 1974) to the definition of the weight matrix to model the effect of plasticity in the brain and analyse the resulting properties. Another important aspect that we have overlooked in this report is what happens if we vary the sparsity of the connections in the network, or in other words, the number of synapses per cell. It turns out that in our calculations we have always been assuming a fully-connected network (see when we take the average of the sums, we do not omit any terms) and that the error probability (and consequently the SNR) actually depends on the combination of the *number of encoded memories* and *synapses per cell* rather than the total number of cells in the network.

The content-addressable memory properties of the brain have been found to be located in the hippocampus, with several studies having shown that if this area of the brain is impaired or removed, the capabilities of both animals and humans at recalling concepts and locations are significantly affected (Nakazawa, 2002). Neuroscientists have also found that groups of 'place cells', which are involved in storing information about places and have been found to present location-elicited firing, behave similarly to the neurons in the Hopfield Network, as they map out entire zones and develop 'stable states' or 'memories' (where firing is higher) in specific locations of interest (Wills, 2005).

## 2 Hodgkin-Huxley Model of the Action Potential

### Introduction

The mathematical model of the action potential developed by Hodgkin and Huxley in their seminal articles (Hodgkin and Huxley, 1952) accurately describes the properties of the channels that are essential for generating and propagating an action potential. More than a half-century later, the Hodgkin-Huxley model stands as the most successful quantitative model in neural science if not in all of biology.

Action potentials are the signals by which the nervous system receives, analyses and conveys information over long distances. They are sharp electrical impulses that are regenerated at regular intervals along an axon. These signals are highly stereotyped throughout the nervous system, although they are initiated by a wide variety of events in our environment. They arise when the membrane potential of an axon reaches a certain threshold, which causes a rapid influx of $Na^+$ ions followed by a slightly slower efflux of $K^+$ ions, giving rise to the characteristic depolarisation and subsequent repolarisation of the membrane shown in Figure 7.

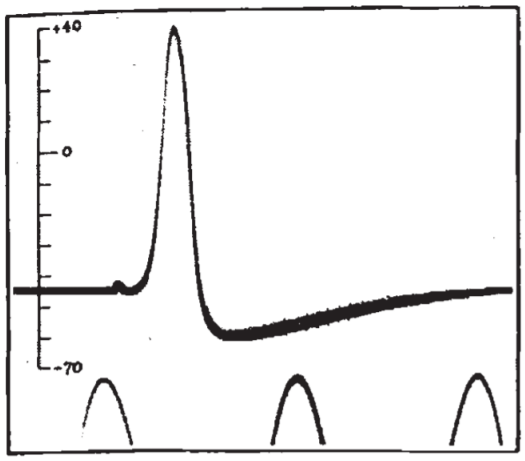



Figure 7: The first published intracellular recording of an action potential. It was recorded in 1939 by Hodgkin and Huxley from a squid giant axon. The vertical scale indicates the potential of the internal electrode in millivolts, with the external sea water the axon was placed in being taken as zero potential. Note the characteristic shape of the action potential, whereby the membrane potential is depolarised from its resting potential of approximately -0.65 mV to close to 40 mV, after which the membrane is repolarised back to its resting potential. Reproduced from (Hodgkin and Huxley, 1939).

Action potentials have four properties important for neuronal signalling. First, they have a threshold for initiation. This is in the form of a threshold membrane potential which is typically at around -50 mV (with respect to the extracellular potential). Second, the action potential is an all-or-none event. The size and shape of an action potential initiated by a large depolarising current is the same as that of an action potential evoked by a current that just surpasses the threshold.[6] Third, the action potential is conducted without decrement thanks to a self-regenerative feature that keeps the amplitude constant even over great distances. Fourth, the action potential is followed by a refractory period during which the ability to fire a second action potential is suppressed. The Hodgkin-Huxley model is capable of predicting three of these features faithfully, namely the *threshold*, the *all-or-none* behaviour and the *refractory period* .

These four properties are highly unusual for biological processes, which tend to respond in a graded fashion to changes in the environment. These properties, in fact, puzzled many biologists for almost a century after the action potential was first measured in the mid 1800s. It was only after the late 1940s and early 1950s studies of the squid giant axon by Alan Hodgkin, Andrew Huxley and Bernard Katz that we first acquired a quantitative understanding of the mechanisms underlying the action potential. Several years of experimentation were condensed into a deterministic set of equations describing the behaviour of the membrane potential of a nerve cell: the Hodgkin-Huxley model.

[6]The *all-or-none* property applies to an action potential that is generated under a certain set of conditions. The size and shape of an action potential *can* be affected by changes in membrane properties, ion concentrations, temperature and other variables.

According to the Hodgkin-Huxley model, an action potential involves the following sequence of events

1. An initial depolarisation of the membrane causes $Na^+$ channels to open rapidly, resulting in an inward $Na^+$ current.
2. This current discharges the membrane capacitance and causes further depolarisation, thereby opening more $Na^+$ channels, resulting in a further increase in inward current.
3. This regenerative process drives the membrane potential up very quickly until it starts to approach the reversal potential for $Na^+$, that is, the potential at which the positive membrane voltage opposes the influx of $Na^+$ down to its chemical concentration gradient. If the membrane potential exceeds the reversal potential, the electrical driving force pushing $Na^+$ out is now greater than the chemical driving force pulling $Na^+$ in, leading to an efflux of $Na^+$ and hence a decrease in the membrane potential.
4. The depolarisation limits the duration of the action potential in two ways: (i) it gradually inactivates the voltage-gated $Na^+$ channels and (ii) it opens, with some delay, the voltage-gated $K^+$ channels, which lead to an outward $K^+$ current that repolarises the membrane. These effects, together with the limit on the growth of the amplitude of the action potential given by the reversal potential of $Na^+$, cause the membrane potential to depolarise back towards its resting potential.
5. The action potential is then followed by a hyperpolarising *after-potential*, a transient shift of the membrane potential to values more negative than the resting potential. This occurs because the $K^+$ channels that opened during the repolarisation of the membrane remain open for some time after the membrane potential has returned to its resting value, causing it to decrease past the resting potential for a brief period of time until the $K^+$ channels close again.
6. The combined effect of this transient increase in $K^+$ conductance and the residual inactivation of the $Na^+$ channels underlies the *absolute refractory period*, that is, the brief period of time following an action potential during which it is impossible to elicit another action potential.
7. As some $K^+$ channels begin to close and some $Na^+$ channels recover from inactivation, the membrane enters the *relative refractory period*, during which it is possible to trigger an action potential, but only by applying stimuli that are stronger than those normally required to reach threshold.
8. After this, the membrane potential returns to its resting state, ready to receive an input large enough to trigger another action potential.

This detailed 'routine' leads to simulations of neuronal dynamics of formidable biological accuracy. In this section we discuss the intricacies of the model, along with their supporting biological explanations. We present an in-depth discussion on the inclusion of the individual terms of the mathematical expressions and outline the effect of each on the overall behaviour of the model axon.

With this preamble, we will be in a good position to proceed to investigate the dynamics of the Hodgkin-Huxley model and its response to different inputs in more detail. We implement the model and analyse the response of a single axon to external current inputs with different shapes and amplitudes. We discuss the biological relevance of our findings and provide qualitative and quantitative explanations for the resulting behaviour. Lastly, we compare the performance of this physiologically realistic model to that of simplified models such as the Leaky Integrate-and-Fire model, rate models or the McCulloch and Pitts neuron and outline the main limitations of these simplified models.

## 2.1 Model

The Hodgkin-Huxley model we use in this report can be depicted using the circuit diagram below,

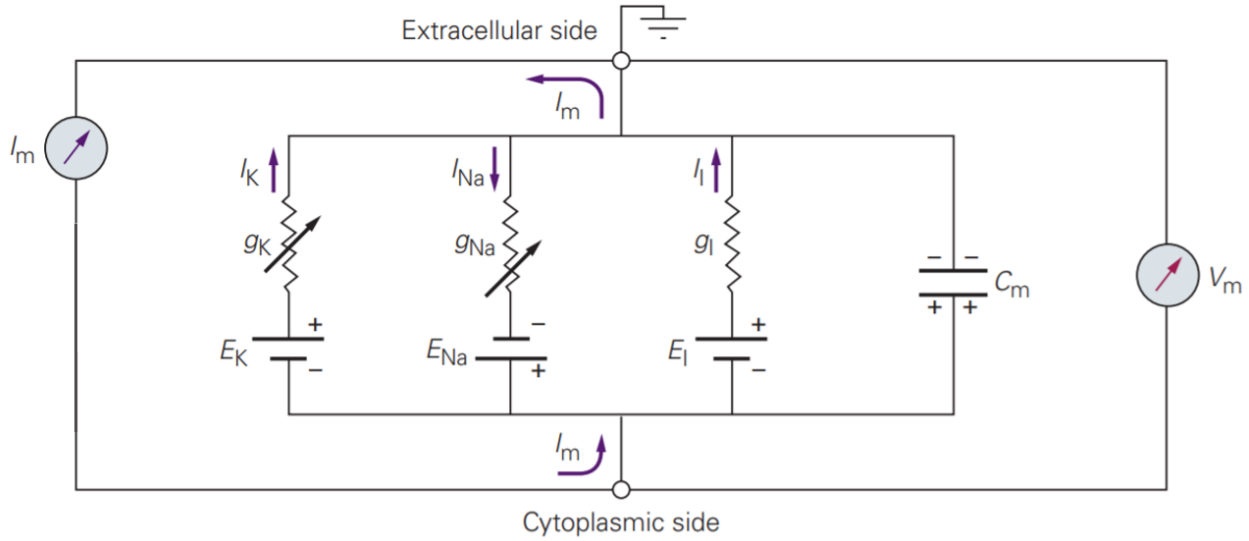



Figure 8: Circuit diagram illustrating the structure of a Hodgkin-Huxley model of an axon using voltage/current sources, resistors and capacitors. $V_\mathrm{m}$ stands for the *membrane potential*, $I_\mathrm{m}$ for the *total membrane current* and $C_\mathrm{m}$ for the *membrane capacitance* (i.e. the capacitance of the lipid bilayer). We also see three different channels through which current can flow from one side of the membrane to the other. The potassium channel has a *reversal potential* of $E_\mathrm{K}$ and a *conductance* of $g_\mathrm{K}$, with the current that flows through it being denoted by $I_\mathrm{K}$. The same definitions apply to the sodium (Na) and leakage (l) channels, where the leakage channel accounts for the natural permeability to ions of the membrane. The differential equations describing the dynamics of this model are presented below. Adapted from (Kandel et al., 2012).

The diagram in Figure 8 defines the following ordinary differential equations which describe the classic Hodgkin-Huxley model of the action potential,

$$\dot{V}_\mathrm{m} = -\overbrace{\overline{g}_\mathrm{Na}\, m^3\, h\, (V_\mathrm{m} - E_\mathrm{Na})}^{I_\mathrm{Na}} - \overbrace{\overline{g}_\mathrm{K}\, n^4\, (V_\mathrm{m} - E_\mathrm{K})}^{I_\mathrm{K}} - \overbrace{\overline{g}_\mathrm{l}\, (V_\mathrm{m} - E_\mathrm{l})}^{I_\mathrm{l}} + I_\mathrm{ext} \tag{2.1.1}$$

$$\dot{m} = \alpha_\mathrm{m}(V_\mathrm{m})(1 - m) - \beta_\mathrm{m}(V_\mathrm{m})\, m \tag{2.1.2}$$

$$\dot{h} = \alpha_\mathrm{h}(V_\mathrm{m})\, (1 - h) - \beta_\mathrm{h}(V_\mathrm{m})\, h \tag{2.1.3}$$

$$\dot{n} = \alpha_\mathrm{n}(V_\mathrm{m})\, (1 - n) - \beta_\mathrm{n}(V_\mathrm{m})\, n \tag{2.1.4}$$

where $n$, $m$ and $h$ are state variables for the membrane currents, $\alpha$ and $\beta$ are functions describing the rates of ion channel gating as a function of voltage, and the constants in the equations have the following meanings and typical values,

- $\overline{g}_\mathrm{Na} = 120\,\mu$S/nF is the maximal *conductance density* (normalised to membrane capacitance $C_\mathrm{m}$) for the voltage-gated $\mathrm{Na}^+$ channels. It is a constant which, when multiplied by the gating variables $m$ and $h$, describes the varying conductance of the $\mathrm{Na}^+$ channels over the course of an action potential as a function of time and membrane potential.
- $\overline{g}_\mathrm{K} = 36\,\mu$S/nF is the maximal *conductance density* (normalised to membrane capacitance $C_\mathrm{m}$) for the voltage-gated $\mathrm{K}^+$ channels. Similar to $g_\mathrm{Na}$, it is a constant which, when multiplied by the gating variable $n$, describes the varying conductance of the $\mathrm{K}^+$ channels over the course of an action potential as a function of time and membrane potential.
- $\overline{g}_\mathrm{l} = 0.3\,\mu$S/nF is the fixed *conductance density* (normalised to membrane capacitance $C_\mathrm{m}$) of the leak channel. Leak channels account for the natural permeability of the membrane to ions and their dynamics can be expressed by an equation like those for voltage-gated channels, where the conductance $g_\mathrm{l}$ is now a constant.

- $E_{\text{Na}} = 50$ mV is the *reversal potential* for the *sodium current* $I_{\text{Na}}$. This is the value of the membrane potential at which $I_{\text{Na}} = 0$, since the electrical driving force (voltage source in Figure 8) ejecting $Na^+$ ions is the same as the chemical driving force pulling $Na^+$ in (determined by channel conductance in Figure 8).
- $E_{\text{K}} = -77$ mV is the *reversal potential* for the *potassium current* $I_{\text{K}}$. This is the value of the membrane potential at which $I_{\text{K}} = 0$, since the electrical driving force ejecting $K^+$ ions is the same as the chemical driving force pulling $K^+$ in.
- $E_{\text{l}} = -54.4$ mV is the *reversal potential* for the *leakage current* $I_{\text{l}}$. This is the value of the membrane potential at which $I_{\text{l}} = 0$, since the membrane potential is such that, on average, the electrical driving potentials and chemical concentration gradients moving ions across the cell membrane cancel out.
- $I_{\text{ext}}$ is the externally applied current (in mA/nF) to the axon, which will be defined differently throughout our experiments.

The state variables $n$, $m$ and $h$ serve a very useful purpose in the Hodgkin-Huxley model presented above. They are used to model the *voltage-dependent conductances* of the sodium (Na) and potassium (K) channels, which are a key feature in the dynamics of the channel currents. Most of the interesting electrical properties of neurons, including their ability to fire and propagate action potentials, in fact arise from nonlinearities associated with active membrane conductances. Individual channels fluctuate rapidly between open and closed states in a stochastic manner. Models of membrane conductances must hence describe how the probability that a channel is open at any given time depends on the membrane potential.

In our formulation of the Hodgkin-Huxley model, we assume that each of the macroscopic Na and K channels, with graded conductances $g_{\text{Na}}$ and $g_{\text{K}}$ in Figure 8, is composed of a large number of microscopic binary subchannels (i.e. which can either be open or closed) of a given type. This assumption has a biological justification, as there are indeed many such Na and K channels along an axonal membrane (see Figure 9). We further assume that the subchannels fluctuate between open and closed states independently from each other (which they do, to a good approximation). Hence, due to the law of large numbers, the fraction of channels open at any given instant is approximately equal to the probability that any one channel is open $P_i$. This observation allows us to move between single-channel probabilistic formulations and macroscopic deterministic descriptions of membrane conductances.

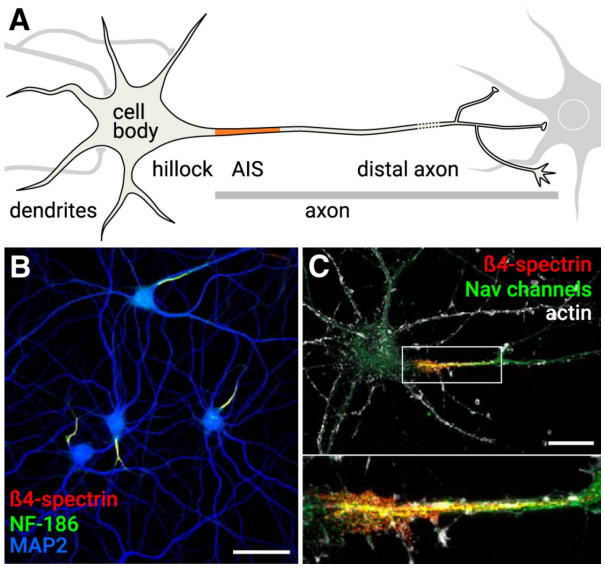



Figure 9: The Axon Initial Segment (AIS). This figure highlights the presence of multiple sodium channels along the membrane of a neuron in green. (A) A typical neuron receives input on the cell body and dendrites (left). The hillock leads to the axon, which contains the AIS (orange). The distal axon contacts downstream neurons (right). (B) Hippocampal neurons after 22 d in culture labeled for the AIS components NF-186 (green) and $\beta$4-spectrin (red). The somatodendritic compartment is labeled using an anti-MAP2 antibody (blue). Scale bar, 50 $\mu$m. C, Hippocampal neuron after 14 d in culture labeled for actin (gray), $\beta$4-spectrin (red), and Nav channels (green). Bottom, The zoomed image represents the AIS. Scale bar, 20 $\mu$m. Reproduced from (Leterrier, 2018).

Having denoted the conductance due to a set of ion channels of type x by $g_{\text{x}}$, the value of $g_{\text{x}}$ at any given time is determined by multiplying the conductance of an open channel $\overline{g}_{\text{x}}$ (or the maximal conductance) by the fraction of channels which are open at that time. Since this fraction is equivalent to the probability of finding any given channel in the open state $P_{\text{x}}$, we can write,

$$g_{\text{x}} = \overline{g}_{\text{x}} P_{\text{x}}. \tag{2.1.5}$$

The dependence of the conductance of a channel $g_{\text{x}}$ on voltage, time and other factors is hence determined by the dynamics of the open probability $P_{\text{x}}$. In our model of the action potential, we simulate the dynamics of the open probabilities (and hence conductances) of the voltage-dependent Na and K channels through the state variables $n$, $m$ and $h$ as follows,

$$P_{\text{Na}} = m^3 h \tag{2.1.6}$$

$$P_{\text{K}} = n^4 \tag{2.1.7}$$

as can be seen in Equation 2.1.1. The leak conductance is assumed to remain fixed at its nominal value of $\overline{g}_l = 0.3\,\mu$S/nF, as its change is negligible. We then multiply these conductances by the difference between the membrane potential and the corresponding reversal potential to get the respective channel currents,

$$I_{\text{Na}} = g_{\text{Na}}(V_{\text{m}} - E_{\text{Na}}) = \overline{g}_{Na} P_{Na}(V_{\text{m}} - E_{\text{Na}}) = \overline{g}_{Na}\, m^3\, h\, (V_{\text{m}} - E_{\text{Na}}) \tag{2.1.8}$$

$$I_{\text{K}} = g_{\text{K}}(V_{\text{m}} - E_{\text{K}}) = \overline{g}_{K} P_{K}(V_{\text{m}} - E_{\text{K}}) = \overline{g}_{K}\, n^4\, (V_{\text{m}} - E_{\text{K}}) \tag{2.1.9}$$

$$I_{\text{l}} = \overline{g}_{l}(V_{\text{m}} - E_{\text{l}}). \tag{2.1.10}$$

Upon looking at the expressions for the open probabilities in Equations 2.1.6 and 2.1.7 one naturally questions why the Na and K channels are treated differently. This is due to a fundamental difference in the biological nature of the two types of channels. There are two different mechanisms by which voltage-dependent channels open and close as a function of membrane potential, and these are modelled through two different types of channel conductances: persistent and transient. Channels for these two types of conductance are depicted in Figure 10 below,

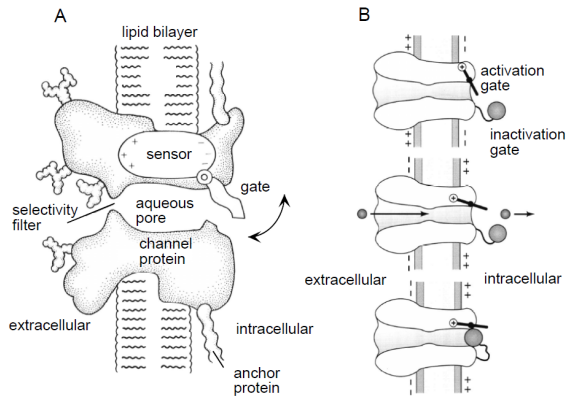



Figure 10: Gating of membrane channels. In (A) we can see an illustration of the gating of a persistent conductance. A gate is opened and closed by a sensor that responds to the membrane potential. The channel also has a selectivity filter which only allows certain kinds of ions to pass through the channel, for example, $Na^+$ ions for a sodium channel. (B) shows a depiction of the gating of a transient conductance. The activation gate is controlled by a voltage sensor (denoted by $\oplus$ in the figure) like in (A). We, however, observe a further gating mechanism (denoted by a ball) which can block the channel once it is open. The top figure shows the channel in a deactivated and deinactivated state, the middle shows an activated channel and the bottom an inactivated channel. (A) from (Hille, 1992) and (B) from (Kandel et al., 1991). Combined figure from (Dayan and Abbott, 2005).

The potassium channel is a *persistent conductance* channel and the sodium channel is a *transient conductance* channel. Persistent conductance channels act as if they had a single type of gate (although we model them as a number of identical subgates) and produce a *noninactivating conductance*. Opening of the gate is called *activation* of the conductance and the closing of the gate is called *deactivation*. For this type of channel, the probability that the gate is open, $P_K$, increases as the axon is depolarised and decreases when it is hyperpolarised. The conductance of the potassium channel (also known as the delayed-rectifier $K^+$ conductance) is hence responsible for repolarising the axon after an action potential, as it favours an outward current of $K^+$ ions when the membrane is depolarised. The delayed-rectifier $K^+$ conductance is constructed from 4 identical subgates that must all be open at the same time for the macroscopic channel to open. Denoting the probability that one of these subgates opens by $n$, and treating their behaviour as independent, allows us to express the probability that the potassium channel is open at any particular point in time as,

$$P_K = n^k \tag{2.1.11}$$

where this $n$ is the same as that in Equations 2.1.7 and 2.1.1. Here, $n$ varies between 0 and 1 and is called a *gating* or *activation* variable, and a description of its dependence on time and membrane potential amounts to a description of the dependence of the overall channel conductance, $g_{\mathrm{K}}$,on these variables. We model this dependence using a simple kinetic scheme in which the transition from closed to open of the subunits occurs at a voltage-dependent rate $\alpha_{\mathrm{n}}(V_{\mathrm{m}})$ and the reverse transition from open to closed occurs with rate $\beta_{\mathrm{n}}(V_{\mathrm{m}})$. The probability that a subgate opens over a short period of time is proportional to the probability of finding the subgate closed $(1-n)$ multiplied by the opening rate $\alpha_{\mathrm{n}}(V_{\mathrm{m}})$ and likewise, the probability that a subgate closes is proportional to $n\beta_{\mathrm{n}}(V_{\mathrm{m}})$. Hence the rate at which the open probability for a subgate changes over time is given by the difference of these two terms,

$$\dot{n} = \alpha_{\mathrm{n}}(V_{\mathrm{m}})\,(1-n) - \beta_{\mathrm{n}}(V_{\mathrm{m}})\,n \tag{2.1.12}$$

as in Equation 2.1.4. Where this can also be expressed in the more convenient form by dividing both sides by $\alpha_{\mathrm{n}}(V_{\mathrm{m}}) + \beta_{\mathrm{n}}(V_{\mathrm{m}})$,

$$\tau_n(V_{\mathrm{m}})\dot{n} = n_\infty(V_{\mathrm{m}}) - n, \tag{2.1.13}$$

where,

$$\tau_n(V_{\mathrm{m}}) = \frac{1}{\alpha_{\mathrm{n}}(V_{\mathrm{m}}) + \beta_{\mathrm{n}}(V_{\mathrm{m}})}, \tag{2.1.14}$$

$$n_\infty(V_{\mathrm{m}}) = \frac{\alpha_{\mathrm{n}}(V_{\mathrm{m}})}{\alpha_{\mathrm{n}}(V_{\mathrm{m}}) + \beta_{\mathrm{n}}(V_{\mathrm{m}})}. \tag{2.1.15}$$

The opening and closing rate functions $\alpha_{\mathrm{n}}(V_{\mathrm{m}})$ and $\beta_{\mathrm{n}}(V_{\mathrm{m}})$ are obtained by fitting experimental data to a functional form derived from thermodynamic arguments. In our model we use the following expressions,

$$\alpha_{\mathrm{n}}(V_{\mathrm{m}}) = \frac{0.01(V_{\mathrm{m}}+55)}{1-e^{-(0.1V_{\mathrm{m}}+5.5)}}, \tag{2.1.16}$$

$$\beta_{\mathrm{n}}(V_{\mathrm{m}}) = 0.125\,e^{-0.0125(V_{\mathrm{m}}+65)} \tag{2.1.17}$$

which are the expressions fitted by Hodgkin and Huxley in units of 1/ms with voltage in mV.

We now move on to the *transient conductance* sodium channels, as the name suggests these open only transiently when the membrane potential is depolarised because they are gated by 2 processes with opposite voltage dependencies (instead of one as in persistent conductance channels). The channel structure that gives rise to this behaviour is shown pictorially in Figure 10B as a channel controlled by two gates, an activation gate and an inactivation gate. The activation gate behaves in exactly the same manner as the gate in potassium channels, having a probability of being open of $m^k$ where $m$ is the probability of a subgate being open and $k$ is an integer representing the number of subgates that make up the overall channel. Hodgkin and Huxley found that a value of $k=3$ for the number of activation gates in a sodium channel gave biologically accurate results.

The inactivation gate, however, behaves slightly differently to the activation gates. We denote the probability that the channel is *not* blocked by an inactivation gate by $h$, which we call our *inactivation variable*. This variable has precisely the opposite voltage dependency to an activation variable: depolarisation causes $h$ to decrease and hyperpolarisation causes it to increase. Hodgkin and Huxley found that an adequate number of inactivation gates per sodium channel was 1, suggesting that a typical sodium channel is composed of three activation gates (which are typically fast at reacting to changes in membrane potential) and one (slower) inactivation gate. Since both the activation and inactivation gate have to be open for the channel to conduct, assuming both channels to act independently, we obtain the following open probability for sodium channels,

$$P_{\mathrm{Na}} = m^3 h. \tag{2.1.18}$$

Once again, since both the activation and inactivation variables represent a probability, they vary between 0 and 1. The voltage-dependent dynamics of these variables are determined by the opening and closing rate functions $\alpha_{\mathrm{m}}(V_{\mathrm{m}})$, $\beta_{\mathrm{m}}(V_{\mathrm{m}})$ and $\alpha_{\mathrm{h}}(V_{\mathrm{m}})$, $\beta_{\mathrm{h}}(V_{\mathrm{m}})$ in exactly the same way as in Equations 2.1.12 - 2.1.15, with the appropriate change in subscript and with the functional forms inverted for the inactivation variable $h$. In our model we continue to use the rate functions fitted by Hodgkin and Huxley as below,

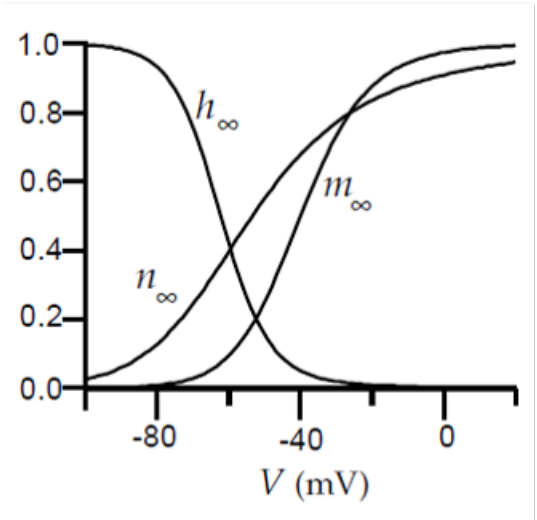
$$\alpha_{\mathrm{m}}(V_{\mathrm{m}}) = \frac{0.1(V_{\mathrm{m}}+40)}{1-e^{-(0.1V_{\mathrm{m}}+4)}}, \tag{2.1.19}$$

$$\beta_{\mathrm{m}}(V_{\mathrm{m}}) = 4\,e^{-0.0556(V_{\mathrm{m}}+65)}, \tag{2.1.20}$$

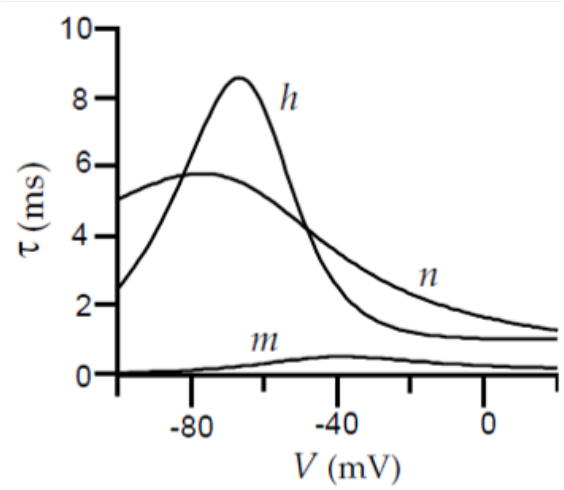
$$\alpha_{\mathrm{h}}(V_{\mathrm{m}}) = 0.07\,e^{-0.05(V_{\mathrm{m}}+65)}, \tag{2.1.21}$$

$$\beta_{\mathrm{h}}(V_{\mathrm{m}}) = \frac{1}{1+e^{-(0.1V_{\mathrm{m}}+3.5)}}. \tag{2.1.22}$$

Using the above expressions, we plot the steady-state values of the gating variables (as in Equation 2.1.15) and their corresponding time constants (as in Equation 2.1.14) in Figure 11 for all three state variables $n$, $m$ and $h$,




(a) The steady-state level of the activation ($m_\infty$) and inactivation ($h_\infty$) variables of the $Na^+$ conductance, and the activation variable ($n_\infty$) of the $K^+$ conductance.




(b) Voltage-dependent time constants that control the rate at which the three state variables $n$, $m$ and $h$ reach their steady-state values.

Figure 11: The voltage dependent functions of the Hodgkin-Huxley model. (a) shows how the steady-state values of the three state variables $n$, $m$ and $h$ change as a function of membrane potential. These are the values that the variables tend to over time. We note that the inactivation variable $h_\infty$ is flipped relative to the activation variables $m_\infty$ and $n_\infty$ and approaches 0 as the membrane becomes depolarised, as opposed to the activation variables which approach 1. (b) shows the voltage-dependence of the time constants of these three state variables, which define the speed with which the variables approach their steady-state values. We note how the time constant for the activation variable of Na channels ($\tau_m$) is always very small compared to the other two, indicating that the activation gates of these channels (or more precisely, their subgates) react very quickly to changes in membrane potential. In general, the slowest gating variable is $h$, which also displays a peak in its time constant at a value close to the resting potential (-65 mV), indicating that the Na inactivation gates take a long time to close at the onset of an action potential. This is what allows the positive feedback loop between the depolarisation and the opening of the activation gates to rapidly increase the voltage and create the characteristic spike shape of an action potential. Both (a) and (b) are adapted from (Dayan and Abbott, 2005).

The presence of two competing factors in Equation 2.1.18 gives the transient conductance some interesting properties. For example, if we want the transient conductance to reach its maximum value, it may first be necessary to hyperpolarise the membrane below its resting potential and then to depolarise it. This is because hyperpolarisation increases the value of $h$, which due to its higher time constant remains high as the membrane potential then depolarises and increases the value of $m$ much more quickly. This results in high values of both $h$ and $m$, which lead to a rapid escalation in the channel conductance. This competing effect also models the *refractory period* after an action potential, since the value of $h$ remains low for some time after the peak due to its slower dynamics, and the value of $m$ will also be low due to the low membrane potential and its faster dynamics. This causes the open probability of the sodium channel to be close to zero shortly after an action potential, making it initially impossible, and then very difficult to elicit another action potential until the inactivation variable $h$ starts to grow again.

Having explained the reasoning and biological effects behind the Hodgkin-Huxley model, here we conclude our description of the model and proceed in the next section with a simulation and investigation of its behaviour when subject to different external currents $I_{\text{ext}}$.

### 2.2 Simulation: Response to Fixed-Length Pulses of Applied Current

The Hodgkin-Huxley model defined by Equations 2.1.1 - 2.1.4 was implemented in Python by numerically integrating the differential equations using forward Euler integration (Euler, 1768) with a fixed timestep of 0.001ms. In all our simulations the model was initialised with the neuron membrane at its resting potential of $V_{\text{m}} = -0.65$ mV. The initial values of the state variables were taken to be their steady-state values at the resting potential, since we assume that we begin our simulation after the neuron has spent sufficient time at rest. These initial values were calculated using Equation 2.1.15 evaluated at $V_{\text{m}} = -0.65$ mV and the results are tabulated below,

Table 1: Initial Values of the State Variables

| $m_{\text{init}}$ | $h_{\text{init}}$ | $n_{\text{init}}$ |
|---|---|---|
| 0.05 | 0.6 | 0.32 |

We began our investigation of the Hodgkin-Huxley model by simulating its response to 200ms-long pulses of applied current $I_{\text{ext}}$. The amplitude of the pulses was initially varied in the range 0.1-5 mA/nF as shown in the figure below,

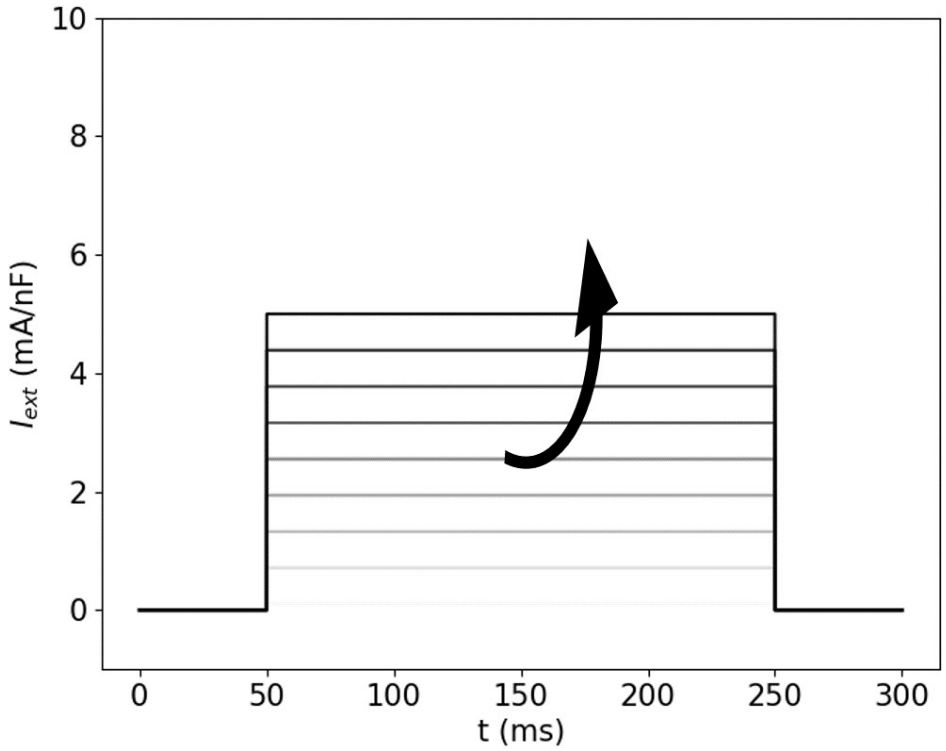



Figure 12: The amplitude of the 200ms-long pulses was initially increased from 0.1 to 5 mA/nF sequentially over 9 steps: 0.1, 0.713, 1.325, 1.938, 2.55, 3.163, 3.775, 4.388 and 5 mA/nF. Current injection begins at 50ms and ends at 250ms.

The resulting responses of the membrane potential for the different current amplitudes were recorded and plotted in the figure below,

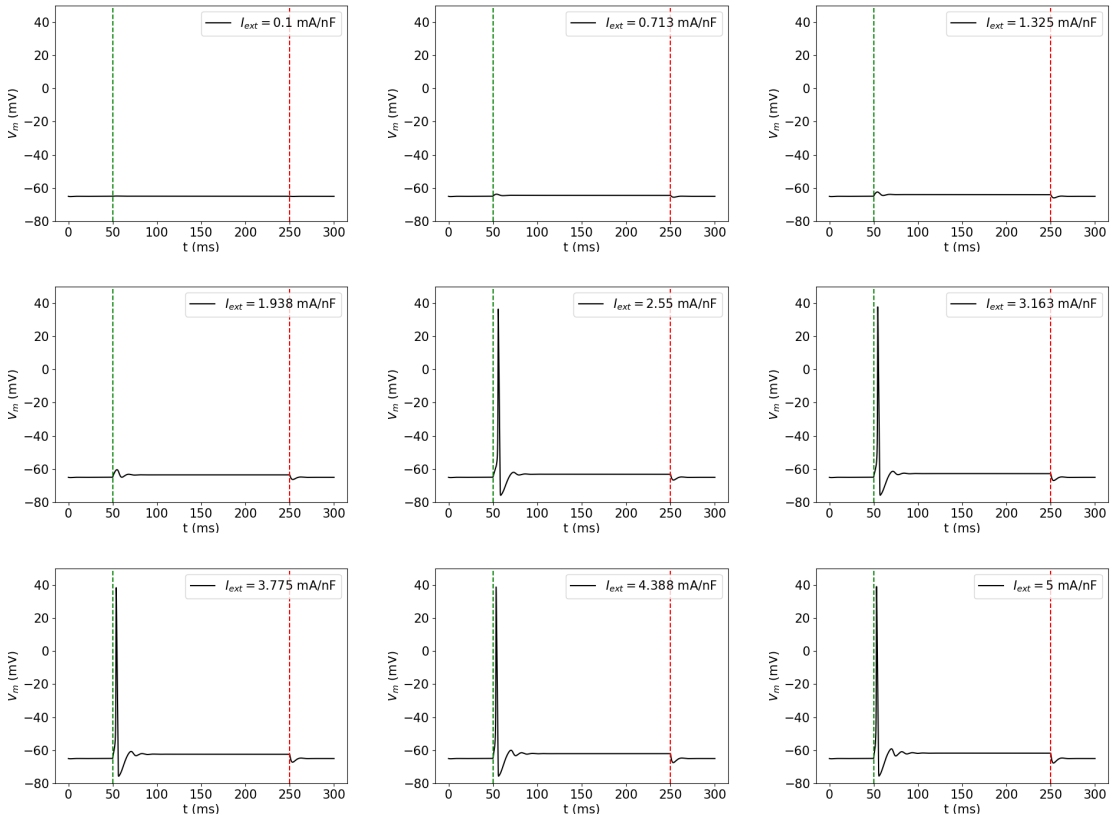



Figure 13: Plots of the membrane potential of a Hodgkin-Huxley axon when injected with 200ms-long current pulses of amplitude varying in the range 0.1-5 mA/nF. The green line represents the onset of current injection (50ms) and the red line represents the end (250ms). We notice that, initially, for small current amplitudes, little change is seen in the membrane potential. We see how the membrane depolarises slightly to higher voltage values throughout the duration of the pulse to accommodate the extra charge being injected, but no spiking behaviour is observed. At the end of the current pulse, the membrane potential returns to its rest value of -65 mV. As the input current amplitude is increased, we reach a point after which we suddenly start to observe action potentials; they manifest as a damped oscillation with a maximum about 20ms after the onset of the current step. This sudden appearance of spiking behaviour indicates the presence of a *threshold effect*, that is, the existence of a threshold input after which we begin to observe spiking behaviour in the membrane potential. This effect is faithful to the known behaviour of biological neurons, and hence a significant advantage of the Hodgkin-Huxley model. Another important feature to notice is that the amplitude and duration of the action potentials that we observe for input amplitudes at and above 2.55 mA/nF remain the same regardless of the input amplitude. The action potentials all have a maximum value of approximately 40 mV and their effect on the membrane potential lasts for approximately 30ms. This shows us that the Hodgkin-Huxley model also simulates the *all-or-none* property of action potentials that we described in Section 2.1.

The plots in Figure 13 reveal a *threshold effect* whereby currents smaller than or equal to $I_{\text{ext}} = 1.938$ mA/nF do not elicit action potentials, and those larger than or equal to $I_{\text{ext}} = 2.55$ mA/nF do. The complex dynamics of the Hodgkin-Huxley model make it very difficult to obtain an analytical expression for this input threshold without making simplifications (one can see (FitzHugh, 1961) and (Nagumo et al., 1962) for examples of simplified versions of the model where this is possible). We hence investigate this region through our simulation and look for the value of the threshold amplitude empirically. We note that it is not the amplitude of the current itself that elicits the action potential, but rather the total input charge which causes the membrane potential to cross the threshold voltage and hence trigger an action potential. Plots of the membrane potential in the range from $I_{\text{ext}} = 2.1828$ mA/nF to $I_{\text{ext}} = 2.3052$ mA/nF are plotted in Figure 14 below,

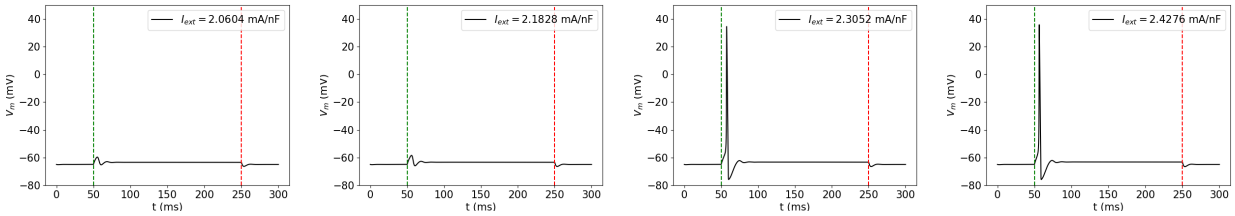



Figure 14: Plots of the membrane potential for 200ms-long input current pulse amplitudes close to the input threshold for the initiation of action potentials. We observe that the threshold lies between amplitude values of $I_{\text{ext}} = 1.938$ mA/nF to $I_{\text{ext}} = 2.55$ mA/nF. Further investigation revealed that the threshold input current amplitude was approximately 2.241 mA/nF, to three significant figures.

We ventured into higher values of the external current pulse amplitude and found that there is a range of values for which regular repetitive firing occurs for the duration of the current pulse, or in other words, the system exhibits *limit cycles*. The range of values was found empirically to be those above 6.25 mA/nF. Plots of the behaviour of the membrane potential for values of the current amplitude around this threshold showing the transition from single action potentials to sustained firing are shown in Figure 15 below,

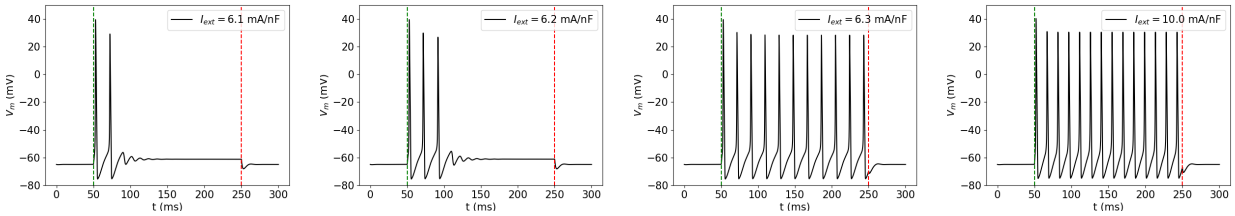



Figure 15: Plots of the membrane potential responding to 200ms-long input current pulses with amplitudes just before and after the input threshold for repetitive firing of approximately 6.25 mA/nF. For values just before the threshold, we can already see multiple action potentials being elicited by the input current. These action potentials, however, do not repeat for the entire duration of the input pulse and instead decay over time until the membrane potential returns to its resting value. For values of the input current amplitude past the threshold we observe repetitive firing over the entire duration of the input pulse. Most neurons will respond to suprathreshold current steps with a spike train where intervals between spikes increase successively until a steady state of periodic firing is reached. In the particular figures shown above, however, this period of *adaptation* is so quick that it is barely visible, hence we say that the neuron is exhibiting *fast firing*. We notice that the frequency of the firing of action potentials increases with increasing amplitude past the threshold. We also observed that as the input current amplitudes were increased even further, they led to action potentials with smaller amplitudes, this decrease is however difficult to see from the above plots.

From the above plots for infrathreshold current amplitudes, we can tell that the membrane response exhibits an increase in the firing threshold after the first action potential from the ringing following the second action potential. This increase in threshold is caused by residual $K^+$ conductance and the decrease in the $Na^+$ conductance caused by the inactivation gates not having had time to return to equilibrium. This, together with the small rise in $V_{\text{m}}$ that it causes, forces more $Na^+$ inactivation gates to remain closed and less activation gates to open, hence leading to a decreased amplitude of the subsequent action potentials or oscillations if the threshold is not reached. For $I_{\text{ext}} = 6.1$ mA/nF, the model produces two action potentials spaced about 20 ms apart, by which we can also tell that the neuron's minimum repetitive firing rate will be greater than 50 Hz (see Figure 15), as otherwise the above effects would lead to a decay in the amplitude of the action potentials and eventually in the emergence of decaying oscillations.

The dependence of the action potential frequency on the input current amplitude observed in Figure 15 was studied in more detail. We plotted this dependence in a firing rate versus input current amplitude plot presented in Figure 16. This plot allows us to see more clearly the threshold at which repetitive firing begins. At the onset of firing, the firing rate jumps from no repetitive firing to a relatively high rate abruptly for a small increase in $I_{\text{ext}}$. In a whole neuron the dependence of the firing rate on the input stimulus is determined predominantly by the neuron's cell body and dendrites rather than by its axon, since the axon's main task is that of transmitting action potentials. From this point of view it does not seem odd that the axon of a cell would present a threshold effect in its firing response.

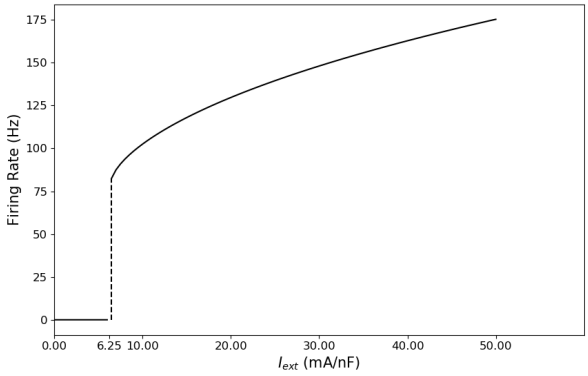



Figure 16: Firing rate of the Hodgkin-Huxley model of an axon receiving input from a 200ms-long current pulse of varying amplitude. As we saw in Figure 15, the threshold for the onset of sustained repetitive is around $I_{\text{ext}}$ = 6.25 mA/nF where we observe a discontinuity in the plot as the firing rate increases from 0 to approximately 80Hz (a value above 50Hz as predicted). The firing rate then increases gradually with increasing external current. A firing rate curve like this one, where we find a discontinuity, is called a *type II* model in the theoretical literature. There are some species of neurons where this discontinuity is not observed. Rather, the onset of firing rates rises from zero without a large jump from no-firing to high frequency firing. Neurons of this type are called *type I* model neurons (Wells, 2010).

The limit cycles of the Hodgkin-Huxley model have been studied in great detail (see (Keener and Sneyd, 2008), for example). One of the ways in which we can illustrate the behaviour of the membrane potential's periodic orbits as $I_{\text{ext}}$ is increased is in the form of a bifurcation diagram, as seen in Figure 17. For each value of $I_{\text{ext}}$ we plot the value of $V_{\text{m}}$ at the resting state, and the maximum and minimum values of $V_{\text{m}}$ over the associated periodic orbit (i.e. the peaks and troughs of the plots in Figure 15). This diagram gives a precise outline of the limit cycle behaviour of the membrane potential. A rigorous investigation of bifurcations would require long derivations and is beyond the scope of this report.

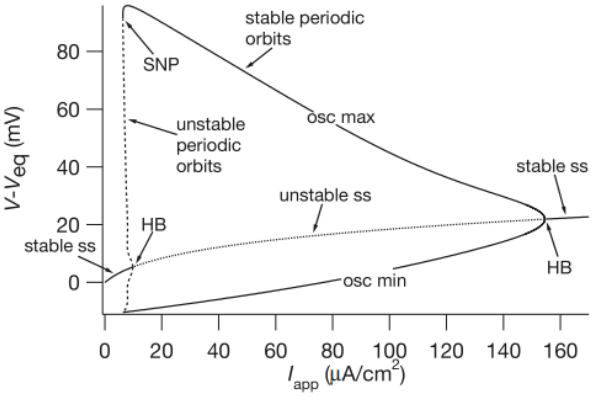



Figure 17: Bifurcation diagram of the Hodgkin–Huxley equations, with the applied current, Iapp ($I_{\text{ext}}$ in this report) as the bifurcation parameter. HB denotes a Hopf bifurcation, SNP denotes a saddle-node of periodic bifurcation, osc max and osc min denote, respectively, the maximum and minimum of an oscillation, and ss denotes a steady state. Solid lines denote stable branches, dashed or dotted lines denote unstable branches. Adapted from (Keener and Sneyd, 2008).

Interestingly, this sustained repetitive firing is actually a discrepancy between Hodgkin and Huxley's theory and experiments, since the squid giant axon that the model was originally intended for does not present this behaviour and instead fires only once at the beginning of the pulse. This is called *type 3 excitability*, as opposed to the *type 2 excitability* of sustained firing and *type 1 excitability* of infrathreshold depolarising current pulses. John Clay in Clay et al. (2008) provides are more in depth discussion of *type 3 excitability* and how the Hodgkin-Huxley model can be amended to include this behaviour.

### 2.3 Simulation: Response to Periodic Square Pulse Current Inputs

In this section we perform a similar investigation to that in Section 2.2 where we now add two more degrees of freedom to the definition of the input current. We specify our input current as a periodic square pulse input with variable period $T$, variable pulse width $p$ and variable amplitude $I$ as in the figure below,

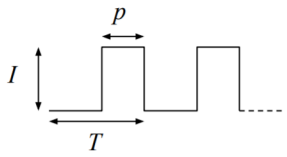



Figure 18: Current Input Waveform. Adapted from the coursework handout.

and systematically explore the effect of driving the model with such an input. The aim of the section is to illustrate the *property of resonance* which is well modelled by the Hodgkin-Huxley model. We begin by exploring the following input parameters,

$$T = 10, 11, 12 \ldots 20 \text{ ms with } p = 5 \text{ ms, and } I = 2.3\text{mA/nF}. \tag{2.3.1}$$

The resulting plots of the response of the membrane potential are presented in Figure 19 below,

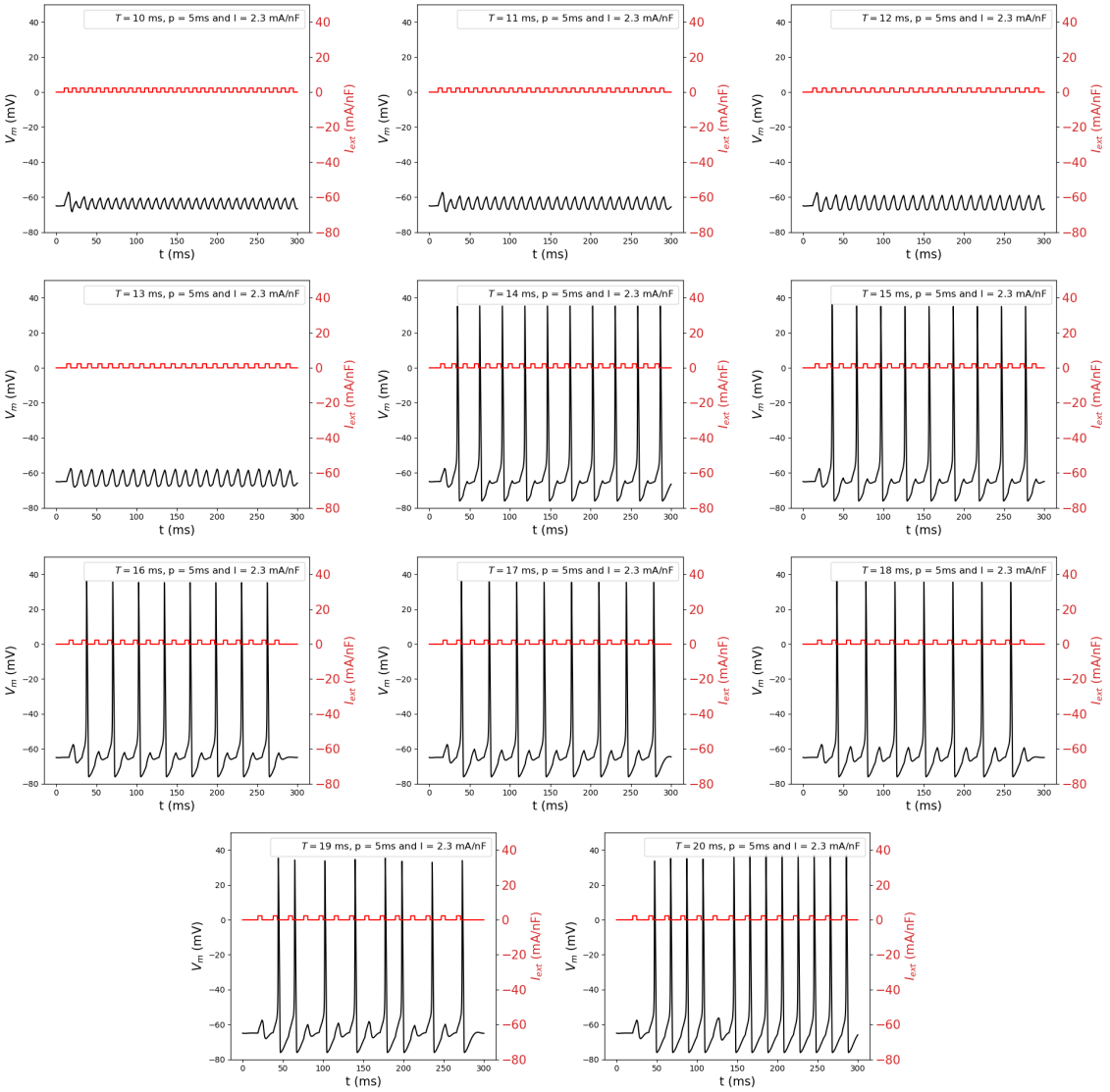



Figure 19: Plots of the membrane potential response to periodic pulses with a pulse width of $p = 5$ ms, an amplitude of $I = 2.3$ mA/nF and 11 different periods in the range $T = [10, 20]$ ms. The trace for the membrane potential is shown in black and the trace for the external current input is in red. We notice that, although the amplitude and duration of the pulses, and hence their charge, are fixed, the membrane potential behaves in significantly different ways depending on the period of the input waveform. The membrane potential does not present any action potentials until the period reaches $T = 14$ ms, after which we observe regular firing with spikes separated by approximately $2T$ ms. This behaviour persists until $T = 19$ ms and $T = 20$ ms where we begin to observe a more irregular firing pattern, which we call *stuttering*. These results might seem to contradict the claim that we made in the previous section about there being a threshold input charge at which action potentials begin to be elicited, since the figures seem to show that action potentials can be generated by subthreshold inputs. The behaviour in these plots can however be explained through the phenomenon of *resonance*, an effect that we will describe in more detail in the following paragraphs.

The phenomenon of *resonance* is one of the main strengths of the Hodgkin-Huxley model; it distinguishes this model from simpler models of the neuron such as the McCulloch and Pitts, Integrate and Fire, rate models or Leaky-Integrate and Fire (LIF) neuron models. We define resonance as the formation of regular firing patterns due to a coherence that emerges between the fluctuations of the membrane potential and the periodicity of the input signal. As seen in Figure 19, the phenomenon only arises at certain input frequencies for a given set of other conditions. We can describe this behaviour from a biological perspective by comparing the dynamics of the gating variables for periodicities at which firing does not occur, such as $T = 11$ ms, and during the regular firing pattern seen at, for example, $T = 16$ ms,

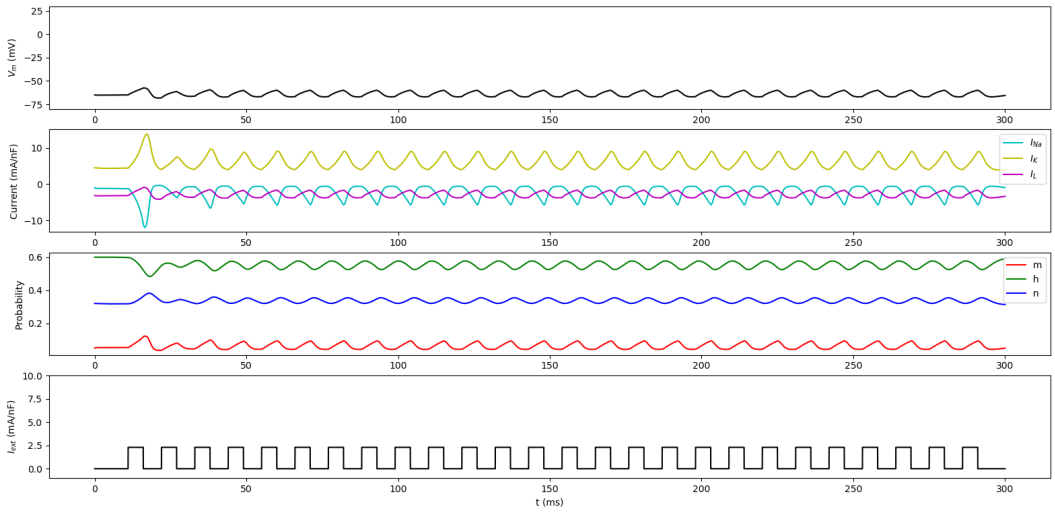



Figure 20: Starting from the top, plots of the membrane potential, ionic currents, gating variables and input current as a function of time for a neuron *not* displaying resonant behaviour at $T = 11$ms. The current pulses occur at too high a frequency for the gating variables to enter the action potential 'routine'. Consequently, the ionic currents do not display any peaks in their dynamics, a sign that neither the positive nor negative feedback circuits is entering its full potential.

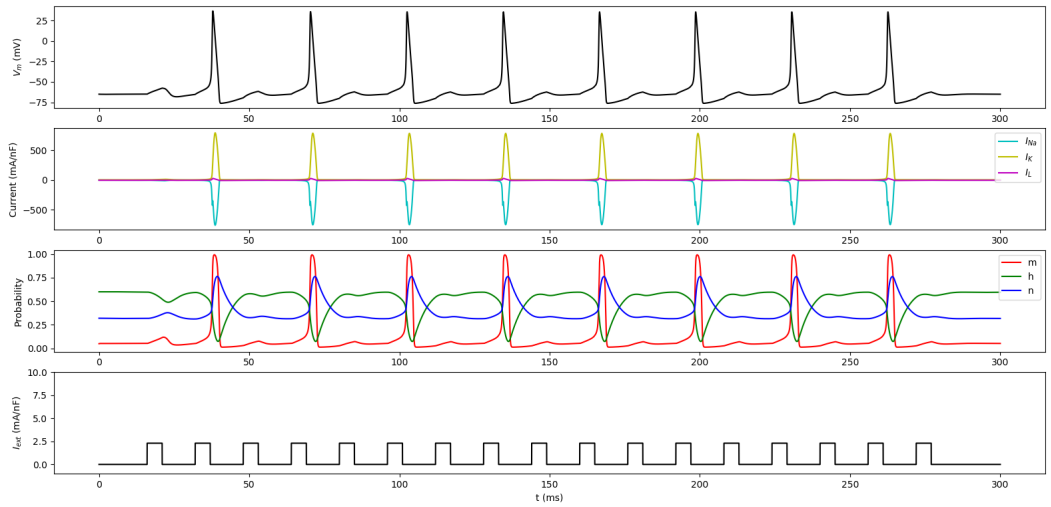



Figure 21: Starting from the top, plots of the membrane potential, ionic currents, gating variables and input current as a function of time for a neuron displaying regular resonant behaviour. The current pulses occur at the right time for the gating variables to enter the action potential 'routine' and elicit regular spikes in the membrane voltage plot. As a consequence, we observe clear and regular spikes in the dynamics of the ionic currents, which is a sign that the sodium channels have entered (and exited, at the end of the action potential) their positive feedback circuits which lead to the rapid escalation in membrane potential. The potassium channel similarly enters its negative feedback circuit which repolarises the membrane after the action potential.

The above plots clearly show the differences in the dynamics of the different variables involved in the Hodgkin-Huxley equations of Section 2.1. From Figure 21 we see how the input pulse frequency is perfectly timed so that it reinforces the behaviour of the gating variables and leads to the formation of action potentials. Take the case of the inactivation variable $h$. This variable decreases whenever there is an increase in the membrane potential and has the slowest dynamics out of the three state variables. In Figure 20, we observe how the initial input pulse leads to a small dip in the value of $h$ as the membrane voltage increases. This pulse has a subthreshold charge and is hence insufficient to elicit an action potential by itself. The next pulse arrives as the variable is returning back up to its rest value, causing it to decrease again, but not by much, as the variable has still not recovered its rest value and is hence more 'inactivating' than at rest, it also carries some upwards 'momentum', which partly cancels the effect of the new input. On the other hand, in Figure 21 the new input arrives after $h$ has already recovered its rest value. In Figure 21 it is even the case that it is starting to travel down towards *lower* values as the new pulse arrives and hence carries a downwards 'momentum', which is reinforced by the new input pulse and triggers a deep decrease in the value of $h$. The fact that the $h$ variable has already recovered its rest value when the new input arrives, and that the state variables are still in an oscillating state after the first pulse, allows the membrane voltage to increase and enter the action potential 'routine' whereby the $m$ activation variable grows very large very quickly and the $n$ variable follows more slowly, allowing the Na current to spike just before the K current and thereby elicit a rapid depolarisation of the membrane. This is quickly followed by the repolarisation of the membrane enabled by the K current. This is, in essence, the reason why resonance occurs; it is all a matter of the dynamics of the gating variables, which in turn model the dynamics of the Na and K channels. These all have their particular time constants and dynamics which, if adequately triggered (as in Figure 21) can lead to very interesting behaviour in the Hodgkin-Huxley model and in real life. We note that in this particular illustrative example we have only discussed why input pulses with higher frequencies than the resonance frequencies do not elicit firing. It is however easy to argue that a similar issue will arise if we wait too long after an input pulse before receiving another, as the effect of the first pulse will have been damped and we will be in the initial situation where a subthreshold input is not sufficient to elicit firing in a neuron at rest.

The more irregular behaviour that we observe for $T = 19$ ms and $T = 20$ ms occurs as the coherence between the input current and the gating variables is slightly lost as the pulses become more spread out. This can be seen in Figure 22 which shows, once again, the dynamics of the Hodgkin-Huxley variables,

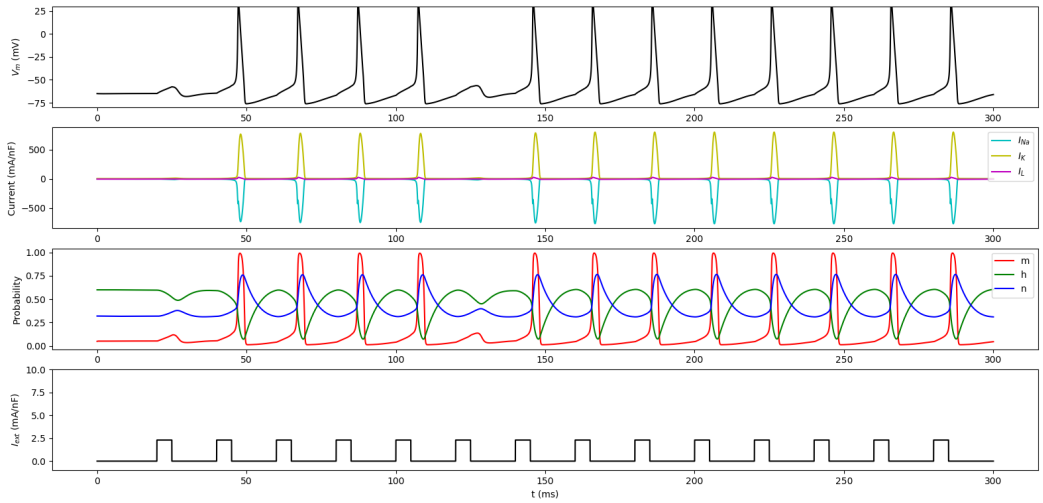



Figure 22: Starting from the top, plots of the membrane potential, ionic currents, gating variables and input current as a function of time for a neuron displaying irregular resonant behaviour. The current pulses arrive a bit too late for the regular firing to occur and instead we observe a fractional (i.e. not continuous) firing behaviour.

Venturing into even higher values of the period (i.e. above $T$ = 20 ms) yielded no more firing. However, for values between $T$ = 19 ms and $T$ = 20 ms we found some interesting behaviour as shown in Figure 23,

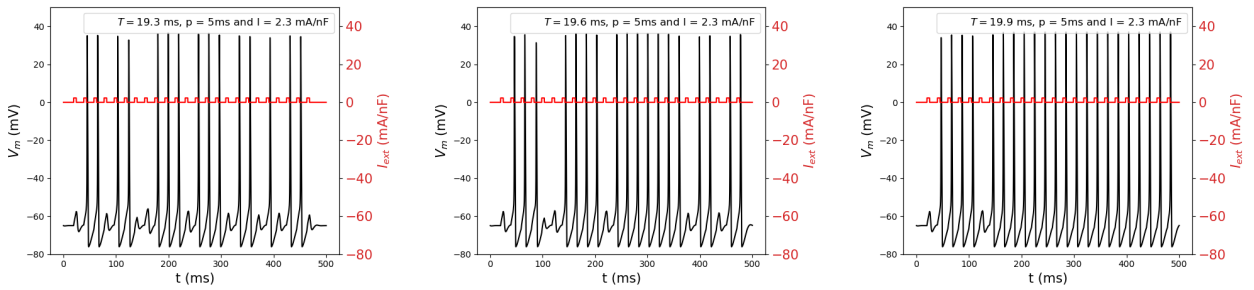



Figure 23: Plots of the membrane potential responding to periodic pulses with a pulse width of $p = 5$ ms, an amplitude of $I = 2.3$ mA/nF and three different periods $T = 19.3,\ 19.6,\ 19.9$ ms from left to right. We notice interesting behaviour as the firing pattern transitions from the approximately 2/3 fractional firing for the leftmost figure to more irrational firing in the rightmost figure.

Several studies have been conducted to find the combinations of current amplitude and frequency (or periodicity) which elicit resonance. In (Parmananda et al., 2002) they present a U-shaped curve which shows the combinations which give rise to repetitive firing,

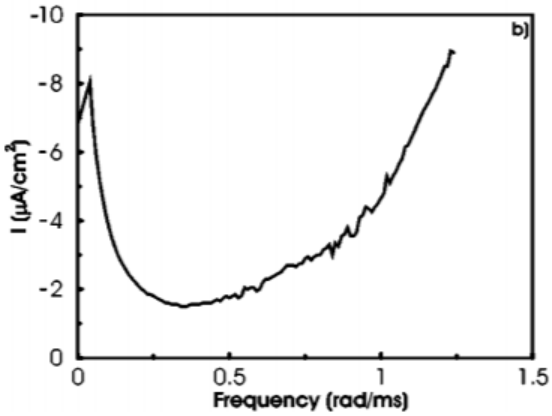



Figure 24: The U-shaped curve encapsulates the region in parameter space (amplitude-frequency domain) where periodic modulations of the input current trigger spike trains in the model system. The curve was constructed empirically using a sinusoidal input current. Despite the difference in form of the input current their results provide insights into our scenario. Adapted from (Parmananda et al., 2002).

The same researchers were also capable of finding a plot which indicates how the firing patterns change as the frequency of the input is varied with all other parameters kept fixed,

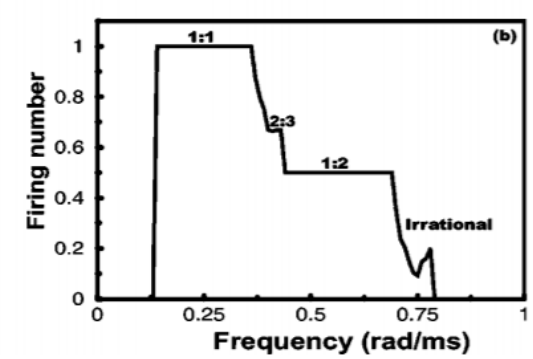



Figure 25: Devil's-staircase-like structure encapsulated by the U-shaped region for $I$ = 3 mA/cm$^2$. It indicates the inception of both rational and irrational firing numbers under the influence of continuous periodic modulations. Adapted from (Parmananda et al., 2002).

We also tried negative values of the amplitude of the input current and investigated the possibility of eliciting action potentials using a negative current pulse. This is indeed possible and is called 'post-inhibitory' rebound. This occurs when we remove an inhibitory current from an axon and leads to an action potential with a slightly different 'routine' from those elicited by depolarising currents. An example of such an action potential is presented in Figure 26 below,

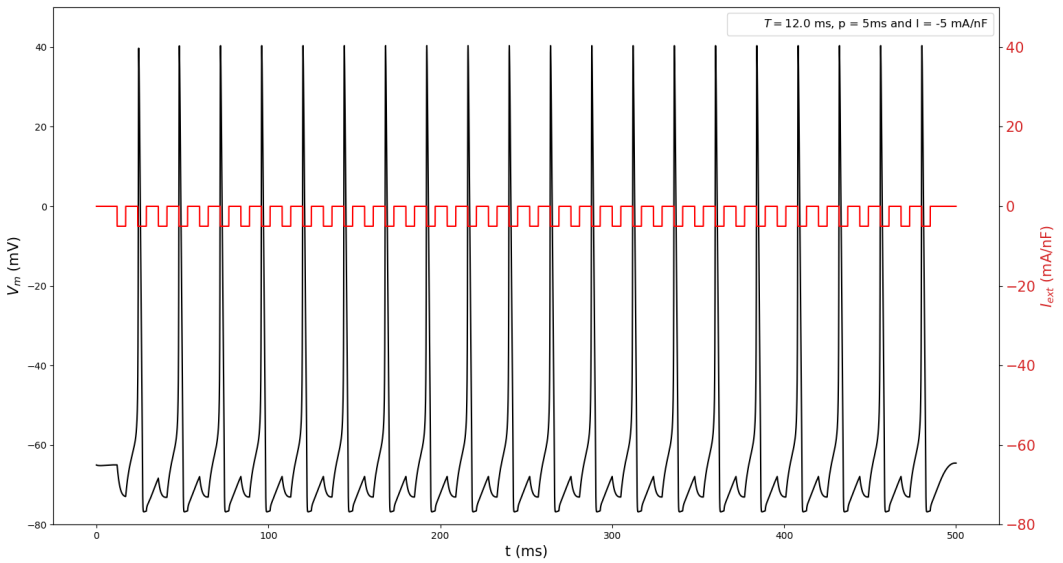



Figure 26: Plot of the membrane potential for an axon responding to a negative periodic current with $I$ = -5mA/nF, $p$ = 5 ms and $T$ = 12 ms. We notice that the membrane potential rises each time the inhibitory current is released, an effect called the 'post-inhibitory rebound'. This is yet another property that distinguishes the more basic LIF or rate models from the Hodgkin-Huxley model. We also notice that here too, the firing action potentials are separated by 2 periods of the input current.

The reason why we can elicit an action potential with a negative current pulse is once again due to the dynamics of the gating variables, though it can also be explained from a physiological perspective. At a low membrane potential, the $h$ variable will be higher than at rest and the $m$ and $n$ variables will be lower than at rest, indicating that most of the Na inactivation gates are open and that most of the Na and K activation gates are closed. However, upon an increase in the membrane potential due to the removal of a negative applied current, the $h$ variable remains high, and the $m$ variable quickly shoots up (i.e. Na activation channels open) allowing a quick influx of $Na^+$ ions with very little 'subchannels' being blocked by the inactivation gate, since the probability of them being blocked is very low due to the high value of $h$. This allows the membrane voltage to increase very quickly and elicit a rapid depolarisation which is then, as before, followed by a repolarisation enabled by both the K channels and the inactivation gates of the Na channels. This results in an action potential. The dynamics of the gating variables described above can be seen in Figure 27 below,

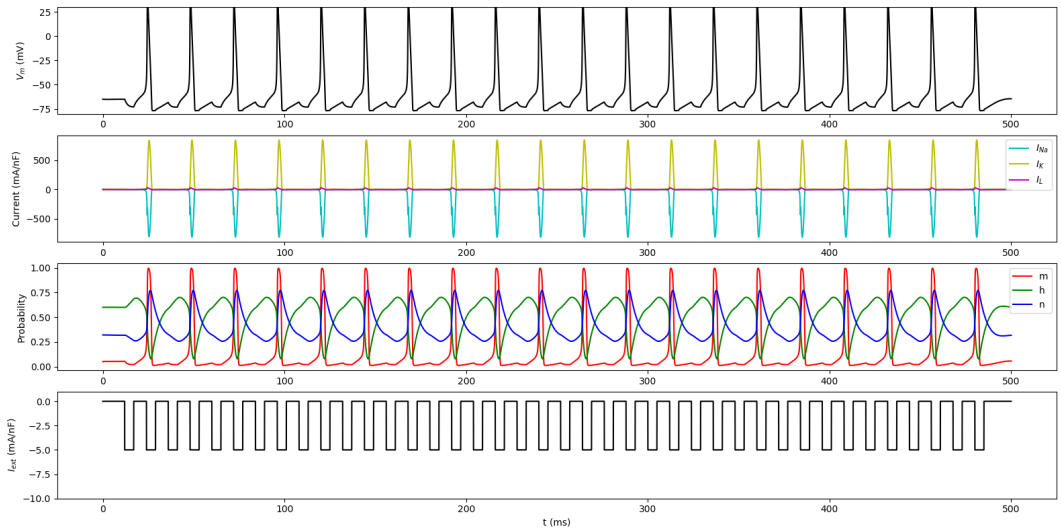



Figure 27: Starting from the top, plots of the membrane potential, ionic currents, gating variables and input current as a function of time for a neuron displaying resonant behaviour elicited by negative current pulses. We observe how the plots look very similar to those of Figures 21 and 22 but with the action potential 'routine' elicited by the sudden elimination of the negative applied current.

### 2.4 Comparison of the Hodgkin-Huxley Model to Simplified Models of the Neuron

We begin our last section by discussing the limitations of the Leaky-Integrate-and-Fire simplified model of the neuron. The dynamics of this model are determined by the following equation,

$$\tau_{\mathrm{m}} \frac{dV_{\mathrm{m}}}{dt} = -(V_{\mathrm{m}}(t) - V_{\mathrm{m}}^{\mathrm{rest}}) + RI_{\mathrm{ext}}(t) \tag{2.4.1}$$

where $R$, $\tau_{\mathrm{m}}$ and $V_{\mathrm{m}}^{\mathrm{rest}}$ are constants representing the leakage resistance, the membrane time constant ($= RC$, where $C$ is the membrane capacitance) and the resting membrane potential respectively. This model has an additional equation which determines the firing behaviour,

$$\text{if } V_{\mathrm{m}}(t) = V_{\mathrm{m}}^{\mathrm{thresh}} \text{ then } \lim_{\delta \to 0; \delta > 0} V_{\mathrm{m}}(t+\delta) = V_{\mathrm{m}}^{\mathrm{rest}} \tag{2.4.2}$$

which essentially means that whenever the membrane potential exceeds the threshold given by $V_{\mathrm{m}}^{\mathrm{thresh}}$ we note the simulation time as the firing timestamp and reset the voltage to its resting value. This indicates that no memory of previous action potentials is kept in the dynamics of the membrane potential.

A limitation of this model can clearly be seen from its defining equations: the external input current, which may arise from presynaptic neurons or from current injection, is integrated linearly, independently of the state of the postsynaptic neuron. This is very different from the non-linear dynamics of real neurons, most of whose properties can be modelled by Hodgkin-Huxley neurons. Another clear limitation is that the effect of the action potentials on the membrane potential is completely forgotten, and hence spike-dynamics-dependent behaviour cannot be modelled. To discuss the rest of the limitations, we present in Figure 28 a collection of the different firing pattern classes that we have observed in Section 2.3 during our investigation of the Hodgkin-Huxley model,

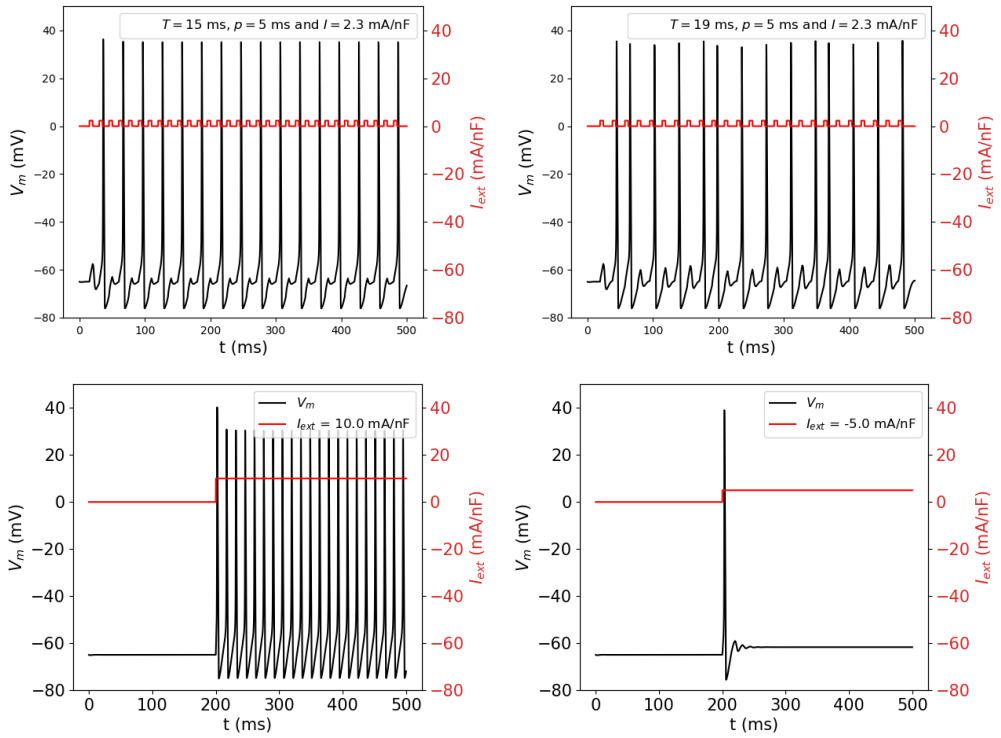



Figure 28: Plots from the investigations carried out in the previous sections outlining the different classes of firing patterns. Top left: *fast-firing* response to a periodic input current. Top right: *stuttering* response to a periodic input current. Bottom left: *adaptation* behaviour of action potentials (albeit very fast) in response to a step input current. Bottom right: *Post-inhibitory rebound* action potential elicited after an inhibitory current is turned off.

Reviewing the different types of firing patterns observed in the plots above, we note the following properties and limitations of the LIF model as compared to the Hodgkin-Huxley model:

1. The fast-firing observed in the top left plot of Figure 28 can indeed be modelled by an LIF neuron as these neurons show no adaptation and can therefore be well approximated by non-adapting integrate-and-fire models. The signals simply consist of action potentials separated by fixed time intervals and hence can be easily recreated through an LIF model with the right time constant.
2. The stuttering behaviour observed in the top right plot of Figure 28 cannot be modelled by LIF neurons as it is a behaviour which arises due to the firing history of the neuron and the dynamics of the membrane potential during and after action potentials. LIF neurons forget their firing history due to the spike-and-reset mechanism and are hence unable to model such behaviour.
3. LIF neurons reset the membrane potential immediately after an action potential is fired, hence the adaptation observed in the bottom left plot of Figure 28 cannot be modelled as it depends on the past firing history. The LIF model may be amended to mimic this process by adding up the contributions to refractoriness of several spikes back in the past. This can be done by using a filter $\eta$ for refractoriness with a time constant much slower than that of the membrane potential, or by combining the differential equation of the leaky integrate-and-fire model with a second differential equation describing the evolution of a slow variable (Gerstner et al., 2014).
4. LIF neurons do not present the 'post-inhibitory-rebound' behaviour seen in the bottom right plot of Figure 28. This is because LIF neurons only fire when the membrane potential rises above the threshold for action potentials $V_{\mathrm{m}}^{\mathrm{thresh}}$. A negative current would drive the membrane potential down in an LIF neuron, and upon inactivation of the applied current the membrane potential would simply return to its resting value.
5. In the plots with the periodic input current in Figure 28 above, spiking behaviour is observed despite the amplitude of the charge injected being subthreshold. This behaviour is clearly impossible with an LIF model as action potentials only arise when the membrane potential exceeds a certain threshold voltage. In fact, if we take the top right plot displaying stuttering behaviour as an example, we see that even if the amplitude of the input current were enough for the membrane potential to exceed the threshold, there would be an inconsistency

in that sometimes we observe an action potential after a single pulse and at other times two are required. This is because the LIF neuron is an 'integrate-and-fire' model as opposed to the 'resonate-and-fire' properties displayed by the Hodgkin-Huxley model.

6. The plots in Figure 28 above, and in the previous sections, have also shown how the shape and amplitude of the action potentials is not always identical across different scenarios, this is also true of biological neurons. This indicates that the nature of the input to a neuron can actually have a slight effect on the characteristics of its spikes, despite them being quite homogeneous overall. This clearly cannot be modelled by an LIF model as it simply records the timestamps of action potentials, completely disregarding their shape, amplitude and duration.

Firing-rate models are more difficult to compare to the Hodgkin-Huxley model as they go up a level of abstraction and do not look at individual spiking patterns, instead defining outputs to be firing rates. Firing-rate models have the advantage that they avoid the short time scale dynamics required to simulate action potentials and thus are much easier to simulate on computers. They are also simple enough that we can derive analytic calculations of some aspects of network dynamics that could not be treated in the case of spiking neurons, and they have fewer free parameters to fine-tune/fit. Another argument in favour of rate models is that they better model the intrinsic stochasticity of neuronal networks, as they do not define what the firing patterns in response to a particular input will look like deterministically, instead limiting the analysis to the more flexible concept of a firing rate. Lastly, firing-rate models have the added benefit that it is not difficult to simplify and reduce the number of components in large neuronal networks by averaging the firing rate of individual elements of the network, an aspect which can be useful to reduce computational requirements. On the other hand, for spiking models like the Hodgkin-Huxley model, it is difficult to define an accurate way of merging network subunits. The way averaged spiking models are typically constructed is by saying that an action potential fired by the larger averaging unit duplicates the effect of all the neurons it represents firing synchronously. Not surprisingly, such models tend to exhibit large-scale synchronization unlike anything seen in a healthy brain.

Firing-rate models however also have important limitations. They cannot account for aspects of spike timing and spike correlations that may be important for understanding nervous system function. Firing-rate models are restricted to cases where the firing of neurons in a network is uncorrelated, with little synchronous firing, and where precise patterns of spike timing are unimportant. Furthermore, they completely disregard the underlying biological mechanism in charge of eliciting spiking behaviour and hence overlook the low-level effects of changes in environment or type of neurons.

## References


John R Clay, David Paydarfar, and Daniel B Forger. A simple modification of the hodgkin and huxley equations explains type 3 excitability in squid giant axons. *Journal of The Royal Society Interface*, 5(29):1421–1428, Dec 2008. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2008.0166.

Peter Dayan and L. F. Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational neuroscience. MIT Press, first paperback ed edition, 2005. ISBN 9780262541855.

L. Euler. *Institutionum calculi integralis*. Number Volume 1 in Institutionum calculi integralis. Academia Imperialis Scientiarum, 1768.

H. Fischer. *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer New York, 2010. ISBN 9780387878577.

Richard FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1 (6):445–466, Jul 1961. ISSN 00063495. doi: 10.1016/S0006-3495(61)86902-6.

Wulfram Gerstner, Werner M. Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: from single neurons to networks and models of cognition*. Cambridge University Press, 2014. ISBN 9781107060838.

Donald O. Hebb. *The organization of behavior: a neuropsychological theory*. Wiley, 11. [print.] edition, 1974. ISBN 9780471367277.

Bertil Hille. *Ionic channels of excitable membranes*. Sinauer Associates, 2nd ed edition, 1992. ISBN 9780878933235.

A. L. Hodgkin and A. F. Huxley. Action potentials recorded from inside a nerve fibre. *Nature*, 144(3651):710–711, Oct 1939. ISSN 0028-0836, 1476-4687. doi: 10.1038/144710a0.

A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, Aug 1952. ISSN 0022-3751, 1469-7793. doi: 10.1113/jphysiol.1952.sp004764.

J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. ISSN 0027-8424. doi: 10.1073/pnas.79.8.2554. URL `https://www.pnas.org/content/79/8/2554`.

J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, May 1984. ISSN 0027-8424, 1091-6490. doi:

10.1073/pnas.81.10.3088.

Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, Feb 1925. ISSN 0044-3328. doi: 10.1007/BF02980577.

Eric Kandel, James Schwartz, Thomas Jessell, Department of Biochemistry Jessell, Molecular Biophysics Thomas, Steven Siegelbaum, and A. J Hudspeth. *Principles of Neural Science, Fifth Edition.* McGraw-Hill Publishing, 2012. ISBN 9780071810012.

Eric R Kandel, editor., Eric R Kandel, 1932-2006 Schwartz, James H. (James Harris), and Thomas M Jessell. *Principles of neural science*. New York : Elsevier, 3rd ed edition, 1991. ISBN 0444015620 (hardcover : alk. paper). Includes bibliographical references and index. Donated by RACO Victorian Branch, Qualification & Education Committee.

James Keener and James Sneyd. *Mathematical physiology. I: Cellular physiology. 2nd ed*, volume 8/1. SpringerLink, 01 2008. doi: 10.1007/978-0-387-75847-3.

Christophe Leterrier. The axon initial segment: An updated viewpoint. *Journal of Neuroscience*, 38(9):2135–2145, 2018. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1922-17.2018. URL `https://www.jneurosci.org/content/38/9/2135`.

Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, Dec 1943. ISSN 0007-4985, 1522-9602. doi: 10.1007/BF02478259.

J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10):2061–2070, 1962.

K. Nakazawa. Requirement for hippocampal ca3 nmda receptors in associative memory recall. *Science*, 297(5579): 211–218, Jul 2002. ISSN 00368075, 10959203. doi: 10.1126/science.1071795.

P. Parmananda, Claudia H. Mena, and Gerold Baier. Resonant forcing of a silent hodgkin-huxley neuron. *Phys. Rev. E*, 66:047202, Oct 2002. doi: 10.1103/PhysRevE.66.047202. URL `https://link.aps.org/doi/10.1103/PhysRevE.66.047202`.

Richard B Wells. *Introduction to Biological Signal Processing and Computational Neuroscience*. University of Idaho, 2010.

T. J. Wills. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876, May 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1108905.