Neural Networks

(The contents of this document are used purely for academic learning purposes)

Muhtasim Ishmum Khan

ID: 2131908642

Department of Electrical and Computer Engineering

1 Introduction

Artificial intelligence (AI) has made machines more capable than ever before and has unlocked great opportunities for humans. However, present AI algorithms are computationally expensive, limiting their use in many situations where it is not possible to have relatively large and powerful hardware. Interestingly, the human brain is able to make large computations while consuming lesser energy using the neural networks and other associated feedback and memory mechanisms. AI algorithms are mathematics based and seldom employ any bio-chemical aspects to boost efficiency and reduce computation costs. Hence, our goal is to explore the mechanisms that are derived from the neural networks of the human brain and tune them to operate in a similar fashion as the brain in terms of efficiency. The objective is to be able to install these mechanisms in minuscule devices (nano-bots could be an example) that are programmable and can, thereby, be used to traverse the human body leading to the possible analysis and cure of numerous diseases such as cancer. In order to do so, however, the human brain has to be understood first and is considered in the following sections in greater detail.

2 The Neuron

The diagram above shows a single neuron. A neuron consists of numerous **dendrites** that branch of to link to other neurons. It is at the dendrites that a neuron receives information in the form of an electric pulse from the input through organs. The end of each dendrite is known as the **synapse** and it is this region which connects neurons, through the receipt and release of neurotransmitters. The synapse that release neurotransmitters are known as the pre-synapse and the synapse that receives them are known as post-synapses. The neurotransmitters cause the dendrites to receive an electric pulse and these can be either excitatory or inhibitory depending on the type of neurotransmitters received. Pulses from each dendrite is, then, sent to the **soma**, or the neuron body, for processing. The accumulation of the pulses after being processed moves down the **axon** to be sent to a following neuron. The axon is insulated by **myelin sheath** with periodic gaps along the length of the axon known as the **nodes of Ranvier**. The purpose of the myelin sheath and nodes of Ranvier are to ensure fast and smooth transmission of electric pulse along the axon. These



Figure 1: An overview of a neuron.

mentioned parts of the neuron and their roles are guided by chemical changes which are discussed in the following section.

2.1 Underlying Chemical Interaction within the Neuron

The environment around the cell or neuron is ionized and has a higher concentration of Na^+ ions than inside the neuron K^+ ions are greater in quantity inside the cell compared to the outside. And when neurotransmitters are received at the post-synapse, ion channels are opened up at the synapse causing an influx of the Na^+ ion, setting up a concentration gradient.

2.1.1 The Resting Potential

The resting potential of the neuron is -70mV and this is maintained by leakage of K^+ ions out of the cell by leakage channels in the cell membrane. However, only K^+ ions are permeable through these channels and so Na^+ ions are pumped inwards through active transport. This maintains an a constant ionic concentration gradient and since there is more positive potential outside the neuron than inside, -70mV is the resting potential across the membrane. The leakage does not significantly influence the passage of the electric pulse when the neuron fires.

2.1.2 The Action Potential

Once the concentration gradient is set up and Na^+ ions enter the neuron at the synapses, they start to diffuse within the neuron since the concentration of Na^+ ions are less in the other regions. In these regions, the potential becomes less negative as the difference in positive charge from outside the cell decreases. Consequently, the potential rises to -55mV and results in voltage-gated Na^+ channels to open up which let allow more Na^+ ions to move into the neuron, diffuse further across the neuron and open more ion channels. Such a diffusion of the ions in known as electrotonic motion. This mechanism is followed throughout the entire neuron to conduct the electric pulse. Along the axon, the ion channels are present at the nodes of Ranvier to let more ions diffuse in the neuron, while the zones covered by the myelin sheath contribute mostly to conduction by preventing the Na+ ions from losing too much energy as they flow. As Na^+ ions increase, the potential across the cell membrane becomes less negative and moves towards positive, eventually reaching a maximum of +30mV. This maximum potential is the action potential of the neuron and is achieved when the neuron fires for any specific input. The increase in potential is known as the process of **depolarisation**.

At +30mV, the ion channels for Na^+ ions close while the ones for K^+ open up. Since there are more K^+ ions within the neuron, as mentioned earlier, another concentration gradient is set up, with K^+ ions diffusing out of the neuron and decreasing the positive potential again until there is more positive charge outside the neuron compared to the inside; a negative potential below -70mV, about -80mV is reached. The decrease of potential across the cell membrane is called **repolarisation**. Then, the leakage and active transport of the mentioned ions in **section 2.1.1** takes place to reestablish the resting potential. The state during which the resting potential is regained is known as **hyperpolarisation**.



Figure 2: The time period for the entire process of (a)depolarisation, (b)repolarisation and (c)hyperpolarisation takes only about 5ms.

3 Deeper Analysis of the Synapse

In order to increase the efficiency of any artificial neural network, the specific interactions that occur at the synapse need to analyzed at a greater depth. The chemical dynamics at this neural junction involves the release, receipt and re-uptake of **neurotransmitters** and the more accurately these processes can be re-created in an artificial network system, the greater the efficiency will be.

3.1 Neurotransmitters

The gap between the pre-synapse and the post-synapse is called the synaptic cleft. At the pre-synaptic terminal, the action potential causes voltage-gated ion channels to open up, but causing an influx of Ca^+ ions into the pre-synaptic terminal. Neurotransmitters are stored in vesicles, which are membranes that isolate the chemicals, preventing any sort of reactions. The Ca^+ ions release synaptotagmin, which basically dissolves the vesicles with the cell membrane causing the neurotransmitters to flow in the synaptic cleft. From the cleft, the neurotransmitters bind to receptors present at the post synaptic completing the transmission of information.



Figure 3: The Synaptic Cleft. Image source: www.simplypsychology.org

All neurotransmitters are synthesized from proteins and each type of neurotransmitter can only cause excitation or inhibition at the post-synapse. For instance, **Glutamate**, causes an excitatory effect and promotes electrotonic spread of Na^+ ions. **GABA**, on the other hand is an inhibitory neurotransmitter, and allows for a very small number of Na^+ ions to enter and does not add to the electrotonic spread of ions that simultaneously takes place from other excited synapses. The different neurotransmitters are responsible for the memory of humans - both long-term and short-term. The implementation of learning depends on the storage and access of past inputs and corresponding outcomes and hence, requires the use of neurotransmitters[2].

3.2 Neurotransmitters and Learning

Based on the neurotransmitters in action, memory is categorized as follows:

- The Immediate Memory
- The Working Memory
- The Verbal Memory
- The Visual Memory

The following sections provide further details about the different memories

3.3 The Immediate Memory

Also known as the short term memory, this is where instant processing data are held. An analogy of Random Access Memory(RAM) of a computer can be used to signify the task of the short term memory. Upon the reception of an input pulse, the output pulse of each neuron in the series of neurons operating for that specific input, is, in a way, stored to determine which neuron to fire next. This is the first part of learning in the sense that the immediate memory helps in predicting the specific path of neurons to pass a certain input to. The release of **Glutamate** (which is excitatory in nature) allows for this functionality.

3.4 The Working Memory

The working memory functions from the secretion of **Dopamine**, which is a type of inhibitory neurotransmitter. When a path for a specific input is stored at the immediate memory, within 30ms it is sent to the working memory. The advantage is that this information is now accessible from the working memory and so can be used by the immediate memory to invoke reactions based on similar but not necessarily the same input. Therefore, a layer of versatility is added due to the inter-dependent operation of the memories.

Moreover, this memory is responsible for longer term processes like muscle memory and so increase the biological neural network's efficiency since there is no need to assess muscle actions every time by sequential neuron fires. Rather it kind of provides a shortcut to certain actions or motor outputs. The neurotransmitter **Acetylecholine** is responsible for this specific role of the working memory.

3.5 The Verbal and Visual Memory

GABA and **Serotonin** are two types of inhibitory neurotransmitters that allow for the recognition of shapes, object and scenarios that are seen, making the visual memory. And the verbal memory, also caused by the same neurotransmitters, enable people to resemble sounds heard, sentences and words said. This memory can be considered a subset of the working memory since the eventual use the visual and audible interactions are sent to the working memory for being processed, recognized and result in the desired reaction.

4 Memory

The different classifications of memory, as described earlier, require further analysis in order to simulate the idea of learning and intelligence in an artificially designed neural model. To do so, memory formation and functioning is looked at greater detail in the following segments.

4.1 Memory Formation

The continuous activation of a specific network of neurons forms memory depending on the strength of the relation between the neurons. The said strength refers to the frequency at which neurons in the specific network is fired - and this phenomenon is called synaptic plasticity. For instance, considering two connected neurons and for a certain input the

former neuron in the network causes the latter one to fire for each time the same input is received. And with each fire, the connection between these two neurons get stronger, becoming, initially, part of the working memory and later (if the strong relation remains) part of the long-term memory.

4.2 Chemical Dynamics that Create Memory

Memory sets the foundation to learning and leads to the presence of intelligence. The interneuronal connection strength that results in memory is caused, like all other neural activities, by chemical interactions - the volume of neurotransmitter being released into the synapse(1) and the number of ions channels that allow for the entrance of Na^+ and Ca^{2+} (Calcium ions) into the post-synaptic terminal through the dendrite branches(2). The following diagram shows this in action:



Figure 4: Synaptic plasticity changes either due to increase in neurotransmitter or inlet channels of the Na^+ ions and Ca^{2+} ions or both. Image source: Queensland Brain Institute

5 Chemical interactions contributing to synaptic dynamics

Post Synapses are not only one dimensional terminals with receptors on their surface, rather these can be modelled according to the Singer-Nicholson model of the fluid mosaic membrane. This suggests that the membrane follows a two-dimensional orientation solution of integral proteins embedded in a viscous phospholipid bilayer. Following this model, there are numerous sub domains of protein complexes at the post-synaptic membrane. This region that acts as a membrane, below the surface is known as the post-synaptic density(PSD)[3].

5.1 Recycling of neurotransmitters and receptors

Proteins from the PSD are responsible for the exocytosis and endocytosis of receptors present at the post-synapse surface. However the rate of degradation and production of the receptors depend on other stimuli[3] that are discussed shortly. For the neurotransmitters that are re-released from the receptors, glial cells present adjacent to the post-synapses - sometimes



Figure 5: The inner layer of post-synaptic membrane houses numerous protein complexes as shown in the diagram. The entire sub-surface membrane is the postsynaptic density (PSD). *Image source:* [3].

almost overlapping them - let out other proteins that act as sort of carriers, taking them through the glial cells called astrocytes and re-inserting them into the vesicles at the presynapse. In addition to contributing for the re-uptake of neurotransmitters, astrocytes also release calcium ions Ca^{2+} that play a crucial role in long-term potentiation[4].



Figure 6: Astrocytes allow for the re-uptake of the neurotransmitters while allowing for electro-chemical equilibrium by influencing the concentration of $Ca^{2+}ions$. The ATP are the carrier proteins that transport the neurotransmitters to the astrocytes. *Image source:* [4].

5.2 AMPA and NMDA receptors

From a general point of view, the activity mentioned at the beginning of **section 5** refers to the fact that for identical inputs, the same sequence of pre-synapses and post-synapses are used and this usage causes the connection of this path to strengthen every time the input is received by the brain. The strengthening occurs in the form of increase in the number of receptor channels at the post-synapse. Therefore, to put it in simpler terms, plasticity increases with increase in the strength of a neural pathway and this occurs through increase in neurotransmitter receptors.

AMPA is the type of receptor that purely allows the enter of Na^+ ions in the synapse. NMDA receptors do not open up solely due to the binding of a neurotransmitter, the channel is blocked by a magnesium ion (Mg^{2+}) , which moves outwards when a potential difference is set up by the influx of Na^+ ions through AMPA receptors. Both Ca^{2+} ions (released from astrocytes mentioned in 5.1) and Na^+ ions enter the post-synapse. The Ca^{2+} ions stimulate the post-synapse and PSD to create more receptors through exocytosis[9] and neurotransmitters have been observed to show greater affinity towards post-synaptic branches that have the highest number of receptors.



Figure 7: Inter-dependency between AMPA and NMDA receptors.

6 Synaptic plasticity and its role in artificial neural networks

The ability of the human brain to learn through experience occurs via the phenomenon of long term potentiation and to achieve this, the synaptic cleft between neurons have to be analyzed. Neuronal communication takes place through chemical interactions at the synaptic cleft, and the behavioural traits of the post-synapse and the pre-synapse changes depending on activity between these two terminals[6]. Therefore, synapses are dynamic in nature and the pattern of their behaviour is discussed as follows: The release of neurotransmitters are due to the firing of a neuron and these neurotransmitters bind to receptors of a post-synapse across the synaptic cleft. However, there are numerous post-synaptic branches that have the receptors for the specific neurotransmitters released. The determination of the specific post-synaptic branch is based on activity of the neuron, that is, the neurotransmitters choose to bind to the post-synaptic branch that had been previously bounded with for the same input. This specific path that is followed for identical inputs creates what is known as synaptic plasticity.

7 Artificial Neural Networks

Now that the functionality and inter-connectivity of biological neurons have been discussed at reasonable depth, the task at hand is to look at how the biological aspects have been implemented to re-create the computational power and intelligence of the human brain.

The first-ever attempt in this regard was by neuroscientist Warren McCulloch and logician Walter Pitts in 1943. They produced a mathematical model that mimics the information (electric pulse) processing and transmission within a neuron. The following diagram shows a McCulloch-Pitts neuron:



Figure 8: Structure of the McCulloch-Pitts neuron.

The inputs to this neuron are binary, i.e. 0 or 1. Inside the neuron body are two functions that are responsible for processing the inputs, the function g causing aggregation or summation of the inputs and the function f is the threshold function that determines whether the aggregated input passes a given threshold. If the threshold is reached or exceeded, the neuron fire is re-created and the output is 1, otherwise, the output is 0. A simplified example is as follows: the neuron will be used to identify the picture of a dog, the inputs represent the following statements:

- x_1 is it an animal
- x_2 does it bite
- x_3 can it be trained

If x_1 is 0 then the output is 0 by default. Such a deterministic input as x_1 represents an inhibitory input. On the other hand, x_2 can be 1 or 0 but it will not determine the outcome, rather the other inputs have to be considered and finally the threshold has to be checked. Such an input is considered excitatory. In the case that all the three inputs are 1, then the output is also 1.

However, this was just a standard mathematical representation of the neuron and in reality there are so much more factors to be considered to generate an evaluation based on the input. So, the Perceptron was designed to take a step closer in developing a more realistic model of the neuron.

7.1 The Perceptron

[10] First proposed by American psychologist Frank Rosenblatt in 1958, the perceptron uses a retinal framework to receive inputs from its surrounding and has functions within its body(based on the McCulloch-Pitts neuron) which allow for an output to be evaluated. The following diagram shows the perceptron model:



Figure 9: The organization of the perceptron. Image source: [10].

The most significant upgrade in the perceptron model is the fact that each input has a weight assigned to it and the input can be refined using layers of these artificial neurons with varying weights in successive processes. These weights allow to distinguish which specific input has the greatest influence on the output to be evaluated. In this framework, the input is also considered as binary and is multiplied to its weight. Therefore, the following function is used for the first step of processing the input information:

$$\sum_{i=0}^{n} w_i * x_i \leq \theta$$

n is the number of inputs, *w* is the weight assigned, *x* is the input and θ is the threshold value that allows for the neuron to fire. The lines linking the retina to the projection area (A_I) ,

in figure 6 are x_i . In A_I are multiple A-units which emulate the presence of the multiple branched dendrites and every single input is received in an A-unit with different weights. The A-units that reach the threshold pulse or value fire and send this signal to A_{II} where there are similar A-units, analogous to the pre-synapse region. The A-units that do not fire at the A_I area do not send further signal down to the A_{II} region. The final aggregation occurs at A_{II} with the received values by each A-unit in this region and the output is either 1 or 0, denoted by R_1 and R_2 in figure 10.



Figure 10: The A-units in the projection area (A_I) drop out the unlikely outputs and send the most likely outputs to the A-units in the Association area (A_{II}) . *Image source:* [10].

7.2 The Hodgkin-Huxley Model and Spiking Neural Network

As mentioned, the spiking neural network is associated with the action potential of a neuron. The dynamics of changes in membrane potential over the course of information transmission that causes a neuron to fire has been discussed in **2.1** and the Hodgkin-Huxley model mathematically incorporates the ionic interchanges that give rise to the flow of current as the neuron fires. The diagrams below shows the representation of the Hodgkin-Huxley model:





The circuit diagram on the right shows the behaviour of the ion channels that allow for the ionic exchange.

Image source: neuronaldynamics.epfl.ch

The role of Na^+ , K^+ ions and leakage channels have been discussed in section 2.1 and extending on that theory, the permeability of the neuron membrane changes with change in potential, thereby regulating the influx and efflux of Na^+ ions and K^+ ions respectively. In the circuit, above, C is the membrane capacitance, R, R_{Na} and R_K denote the varying ability of the leakage ions, Na^+ ions and K^+ ions respectively to flow across the cell membrane. And since, each type of ion channel requires a specific potential to open, E_L , E_{Na} and E_K represent those voltages for the leakage channels, Na^+ channels and K^+ channels respectively (found by experiments conducted by Sir Alan Hodgkin and Sir Andrew Huxley on a squid giant axon). The total current, I, is input to the neuron and regulated using the resistors such that it allows for the peak voltage to hit consecutively for each individual input of I. The following graph shows the result: The behaviour of this graph can be described



Figure 12: The Spikes of action potential are consecutively hit when an artificial neural network is controlled using the Hodgkin-Huxley model.

mathematically using a differential equation derived from the circuit in figure 11. C = q/u, here, the capacitance, C, of the membrane equals to the charge, q, stored for the membrane potential, u, per unit. Considering the current across the cell membrane is I_C and it is known that I = q/t or current equals charge per unit time. Therefore, substituting I_C into the equation of C gives: $C = I_C * t/u$, or $I_C = C * u/t$ And since they flow of current in the neuron is continuous when the neuron becomes activated or is about to fire, the instantaneous current in the cell membrane as a function of time would be:

$$I_C(t) = C \frac{\partial u}{\partial t}$$

The currents for the ion channels and leakage channels are I_L , I_K , I_{Na} and the final form of the equation is:

$$I(t) = I_C(t) + I(t)_L + I(t)_K + I(t)_{Na}$$

$$\Rightarrow I(t) = C\frac{\partial u}{\partial t} + \sum_{k} I_k(t)$$

In the final equation, k is not the K^+ ions rather it is a dummy variable. The use of this equation helps to construct a more efficient artificial neural network as it only utilizes on the firing part of neural activity and therefore, drops out instances or iterations that do not cause the neuron to fire, hence, lowering the time taken to train the model[5].

8 Forming Artificial Networks

Combining the perceptron model with chemical aspects of neuronal activity as discussed above, a type of neural network, known as a Biomolecular Neural Network(BNN)[7] can be formed in the attempt to make the network more efficient in terms of its ability to determine which specific combinations of computations it has to carry out for a specific input. The expected outcome is the decrease in computational time and power and increase in accuracy in the power of recognition for identical inputs due to the use of the concept of plasticity. The following equations have been proposed as a result:



Figure 13: Simplified biological equations to map onto the biomolecular framework.

8.1 Biological Regulators

The proposition provided above is, however, deficient of implementable aspects which should make the system an efficient artificial network. For instance, the production of \mathbf{R} is a noisy process which is likely to affect the performance of the system. Furthermore, \mathbf{T} has no source and no regulation and hence, it is a matter of investigation to identify the most bio-physically relevant method of introducing \mathbf{T} in the system without hindering run time. To solve such issues, regulatory molecules that govern the behaviour of \mathbf{T} or neurotransmitters are analyzed.

8.2 The AMPA-NMDA dependency

Using the concept in section 5.2, there is an up-regulation of the neurotransmitter receptors or \mathbf{R} due to the activation of NMDA receptors triggered from AMPA receptor and neurotransmitter binding. Furthermore, NMDA kinetics are much slower compared to AMPA and other types of receptor kinetics [12] and so these receptors can be considered as a type of regulatory co-agonists controlling the concentration flow of neurotransmitters in and out of the pre-synapse and post-synapse respectively. Drawing inspiration from this knowledge, the model in section 8 can is modified to the following diamond network.





Although the concept of these receptors is used from biology, the network design is customised, targeted to be an optimised ANN. It is designed as a combinational network with two double positive feedback loops between **A** and **O** and **N** and **O**; autoregulation in **A** and **O**; feed-forward connection from **T** to **A** and **N**[1]. The dynamics of this network are explained as follows.

8.3 ANN version 2.0 [explained]

The open channel due to neurotransmitter-receptor binding is treated as the output molecule of the system and as per this formation, the production of \mathbf{O} up-regulates the production

of **A** and **N** while both **A** and **N**, but not necessarily at the same time, leads to the production of **O**. The feedback relation here is not linear; the production of **N** is dependent on a threshold concentration of **O** which is produced from **A**. When the specific threshold concentration is reached, **N** is takes part in the reaction and starts controlling the number of **O** until reaching a steady state concentration.

Nevertheless, this control mechanism set up by the AMPA and NMDA receptors is also dependent on many protein-based co-agonists in the PSD (section 5) and, surrounding glial cells as well as others present in the pre-synapse. Hence, a different biological aspect is considered in the following section.

9 Glycine in the Synaptic Cleft

Glycine is an amino acid that can enter the synaptic cleft from either the glial astrocytes mentioned in **section 5** or diffuse from extra-synaptic regions, which in layman's terms, is the outside region surrounding the synaptic cleft and the synaptic branches. To work as a co-agonist, it also requires that D-serine, another amino acid, be present at the site of binding. Working together, glycine and D-serine act as partial activators of the NMDA receptor, thereby, controlling synaptic plasticity[8].



Figure 15: Gylcine enters the synaptic cleft from the astrocytes and regulates the binding of neurotransmitters with receptors. Image source: [11].

Hence, instead of considering NMDA and AMPA as individual regulators, the role of Glycine and D-serine are taken into account and the following modification to our existing model is made: $\mathbf{T} \Rightarrow \mathbf{N}$; $\mathbf{A}, \mathbf{N} \Rightarrow \mathbf{R}$; $\mathbf{O} \Rightarrow \mathbf{NR}$; the role of Glycine and D-serine are cumulatively accounted for and represented in our model as a custom molecule.

References

- [1] U. Alon. An introduction to systems biology: design principles of biological circuits. CRC press, 2019.
- [2] M. Bear, B. Connors, and M. A. Paradiso. *Neuroscience: exploring the brain, enhanced edition: exploring the brain.* Jones & Bartlett Learning, 2020.
- [3] D. Choquet and A. Triller. The dynamic synapse. Neuron, 80(3):691–703, 2013.
- [4] C. Eroglu and B. A. Barres. Regulation of synaptic connectivity by glia. *Nature*, 468(7321):223–231, 2010.
- [5] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal dynamics: From* single neurons to networks and models of cognition. Cambridge University Press, 2014.
- [6] D. R. Hammond-Weinberger, Y. Wang, A. Glavis-Bloom, and N. C. Spitzer. Mechanism for neurotransmitter-receptor matching. *Proceedings of the National Academy of Sciences*, 117(8):4368–4374, 2020.
- [7] A. Moorman, C. C. Samaniego, C. Maley, and R. Weiss. A dynamical biomolecular neural network. In 2019 IEEE 58th Conference on Decision and Control (CDC), pages 1797–1802, 2019.
- [8] T. Papouin, L. Ladépêche, J. Ruel, S. Sacchi, M. Labasque, M. Hanini, L. Groc, L. Pollegioni, J.-P. Mothet, and S. H. Oliet. Synaptic and extrasynaptic nmda receptors are gated by different endogenous coagonists. *Cell*, 150(3):633–646, 2012.
- [9] L. Pickard, J. Noel, J. M. Henley, G. L. Collingridge, and E. Molnar. Developmental changes in synaptic ampa and nmda receptor distribution and ampa receptor subunit composition in living hippocampal neurons. *Journal of Neuroscience*, 20(21):7922–7931, 2000.
- [10] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [11] M. W. Sherwood, S. H. Oliet, and A. Panatier. Nmdars, coincidence detectors of astrocytic and neuronal activities. *International Journal of Molecular Sciences*, 22(14):7258, 2021.
- [12] S. Zhu, R. A. Stein, C. Yoshioka, C.-H. Lee, A. Goehring, H. S. Mchaourab, and E. Gouaux. Mechanism of nmda receptor inhibition and activation. *Cell*, 165(3):704– 714, 2016.