SAVONIA

University of Applied Sciences

THESIS – BACHELOR'S DEGREE PROGRAMME
TECHNOLOGY, COMMUNICATION AND TRANSPORT

HIDDEN ALGORITHMS IN DNA AND DISCOVERING APPROACHES

AUTHOR/S Ngoc Dung Lai

SAVONIA UNIVERSITY OF APPLIED SCIENCES THESIS
Abstract

Field of Study Technology, Communication and Transport	
Degree Programme Degree Programme in Internet of Things	
Author(s) Ngoc Dung Lai	
Title of Thesis Hidden Algorithms in DNA and Discovering Approaches	
Date 27 May 2023	Pages/Number of appendices 59
Client Organisation /Partners Savonia University of Applied Science.	

Abstract

Many phenomena on floras and faunas, such as their unconditional reflections and complex vision, provide evidence of natural mathematical algorithms embedded in DNA. This thesis aims to discuss possible methods to exploit those mathematical algorithms significantly. By using suitable techniques, unknown natural algorithms can be brought to lighten up.

To explore these algorithms, two approaches are represented. The first approach uses the Boolean network model to analyze the gene expression data. From the analysis, the arrangement between the fundamental functions in the system can be inferred. This approach can explore the information at the cell level, but it requires a large amount of observation and high computing complexity. The second approach is to capture the neural signal from the brain and sensory organs, after that, a mapping function can be explored. This thesis does not cover the scope of this problem, only the first step of this problem, which is to find an implantable device that can observe the neural signal at the position near a sensory organ, will be discussed.

The result shows that the Boolean network model can be used as an effective tool to discover the arrangement and the causal-effect relationship between the components of a biological system. There are still problems making the model not describe the biology process accurately, hence, it needs to be modified based on the real observations. For the second method, research efforts in recent years have facilitated the creation of such devices.

In conclusion, this is an interesting but challenging issue that is difficult for any individual or organization to solve on its own. Hence, cooperation between people from different fields is necessary to discover complete answers.

Keywords

Gene expression, Boolean network, neural signal processing, stentrode, biological intelligence.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The cognition of biological creatures has always been always a big question for centuries, and the answer to this question may be partially found in DNA. Many phenomena in nature have shown evidence about similar behaviors and responses between individuals in the same species, which are spontaneously assigned to them from the born. Interestingly, animals or plants do not have to learn to get these behaviors. The unconditioned reflex is a prominent example of this type of ability. This phenomenon raises up a concern about common controlling algorithms which are embedded in DNA and shared between individuals. The problem will be represented more clearly in the body of the thesis. In this scenario, the reversing technique on genomes may help to examine the DNA, understand its operation, and provide evidence about the natural cognitive mechanism. The expression from genome to cognitive and behaviors is not represented in tangible forms, but it is expressed through an abstract process. Therefore, information stored in genomes must not only include sections encoding for tangible components such as DNAs, RNAs or proteins, but also sections to regulate the intangible process such as the arrangement of events happening in the cells or localization of tangible components in the cell. And if there is any arrangement between functions in a fundamental function set, it can be considered being an algorithm. In the thesis, two approaches which are potential to partially solve the problems will be addressed separately in two parts:

- Boolean network model: Among many aspects of genomes being studied, gene expression analysis is a suitable tool to infer biological algorithms. By capturing the messenger RNA in the cell at different time, the instruction from DNA can be recorded and decoded to tell which function the cell is doing. This information about functions can help to build a Boolean network model, which show the arrangement between these functions. Analyzing result can be the evidence to detect operating mechanism of the cells. The theory and algorithms used to analyze the network will be discussed in chapter 3.
- Signal processing approaches: Doubts about the existence of common mapping functions between individuals, which convert signals from receptors into sensations, gives an impetus to develop a device that could record neural signal implanted near sensory organ. If the record from the device can prove that the mapping functions truly exist, they will be more concrete evidence about the algorithms embedded in the genomes, and these mapping functions then can be used to apply on electronic devices to help them mimic the sensation of natural creatures. This section describes the progress researchers have made in the field of neural signal acquisition and describes a simple experiment for signal acquisition and processing. In the experiment, existing tools from Savonia's laboratory will be used to create a simple circuit which can collect and process signals from skin and muscle.

Both approaches have their own advantages and disadvantages. The Boolean network can help to discover the problem in greater detail, down to the information at the cellular and tissue level. The weakness of this method is that the computational space is very large, and there are currently no effective methods to solve it in an acceptable time. The computational complexity of the signal processing method is lower; however, the implementation of sensor is complex, and it can only discover the algorithm in the nerve system.

2 PROBLEM IDENTIFICATION

2.1 A mathematical issue of operator's arrangement

The coordination rule and arrangement between operators is an important topic, which can affect the results of mathematical problems. Let's examine a following ancient math problem: Adding operators of the set {+, -, ×, ÷} to complete the equation below:

$$9 \quad 9 \quad 9 \quad 9 = 81 \tag{1}$$

There are five arrangements which satisfy the equation, they are divided into 2 sets: {×, +, -}, {+, -, ×} and {×, ×, ÷}, {×, ÷, ×}, {÷, ×, ×}. It can be noticed that the order and coordination rule of operators are already specified (coordination rule: from left to right, coordination rule: multiply and divide are calculated first then plus, and minus are calculated). The coordination rule of the calculation, which is not expressed explicitly in the equation, play a role in determining the result. In fact, people use parathesis to interfere with this rule. For example: 9 x 9 + 9 ÷ 9= 81 but 9 × (9 + 9) ÷ 9 = 18. In the solution if the order is changed: 9 - 9 × 9 + 9 = -63 ≠ 81. Therefore, arrangement of operators has as vital role as numbers and variables. Other sub-issues are the conditions for the system of equations to have solutions, and the number of solutions of the system.

This mathematical problem plays an important role in solving genome reversing problem. Genome includes many parts where each part provides a fundamentals function, they play the same roles as operators in the above example. These functions need to be organized in a reasonable arrangement to make sure the biology system works properly, otherwise it can lead to the chaos. Hence it is necessary to put other information with an accurate arrangement together, to explain the operation of DNAs and its products. However, the analysis about these arrangements have not received sufficient attention. In latter part of this document, a graph tool presenting the gene expression network will show a clearer picture.

Unfortunately, there is no mathematics theory studying this problem at this moment. One simple approach is to present the calculation in the graph form, each vertex is assigned to a number and each edge is assigned to an operator. Then iterations take place all over the searching space to find the right solution. To find better solutions, this mathematical problem should be examined further in the future.

2.2 Evidence about the common algorithm embedded in the genome.

The starting point of the concern about biological algorithms begins with the vision abilities of natural creatures. While it may take a computer several hours to days for analyze the 3D structure of a complex object, some simple insects can solve the problem in real time, outperforming computer in the image processing field. Then there are two problems that be clarified: does it stand for biology algorithms, and if yes what do they come from? To answer the first question, let's dive deeper into the eyes structure and their working mechanism. In humans, three types of photoreceptor cells exist in the retinas, they are L-cones, M-cones, and S-cones with the quantities portion being sequentially 60%, 30% and 10%. They contain different kinds of macro proteins called iodopsin, which are highly conjugated. When a beam of photon is injected to these proteins, physiological processes take place,

converting light energy to chemical potential energy. The signal then is transmitted to the cortex by optic nerves. One noticeable point is among sixty distinct types of neurotransmitters, transmission process in optic nerve at first stage only use L-glutamate (Gegenfurtner,1999). It leads to a concern of how the visual cortex can differentiates between three types of signals red, green, and blue. There are two possible ways to explain this phenomenon. The first possibility is the existence of a communication protocol between brain and retinas. By using this protocol, the brain can base on some special features of the signal to decode the signal, such as signal pattern, signal intensive or interruption between two adjacent cycles of the signal. This concept is similar to the communication protocol in electronic devices such as SPI, I2C or UART. The second explanation is the segmentation of brain cortex, visual cortex may include three different functional sub-regions which are responsible for decoding signal of special wavelengths red, green, or blue. After building the monochromatic images, all of them will be combined to create our vision. However, no matter what mechanisms are used by the optic nerve cells and brain, both above hypotheses require a set of strict rules to regulate and combine all these fundamental functions into a complex function, so the brain can create an accurate cognitive about the surrounding environments. Therefore, it can stand for a sequence of order between these functions, hence an algorithm exists in this scenario. In fact, not only vision cognitive but other human senses also represent these features. For example, the spicy taste of tongue is a pain signal sent by the nerves that transmit touch and temperature sensations. Or when a person is exposed to cold weather, he properly feels the pain as some needles penetrate his skin (organ that senses temperature but also sends out pain signals). That paresthesia suggest that the sense seem to be built by the brain by decoding the patterns of signal.

Figure 1. Experiment set up (Lai, 2023)

Now moving on the second question, It can be noticed that humans share the same method of image processing. An experiment is conducted to inspect the correctness of this hypothesis. An experiment is organized in a quiet room, in the library of Savonia University of Applied Sciences, 2 Microkatu Kuopio. Figure 1 shows the experimental setup. There were seven isolated tables, equipped with private panels, 60 cm in height x 120 cm in width and a narrow slit between two adjacent panels which allowed people to observe the scenery outside. Nine participants were asked to sit in the middle table and slowly turn their head from left to right and describe their vision. The destination view was another studying table which was in the right corner and contained many edges and details. Four of the participants were taken photos from above to compare their head axis and measure head angle. All participants experienced one special point, at that point their vision suddenly switched from the vision of almost left eye to the vision of almost right eye. Their brains also tended to keep as many details as they can when switching the vision happened. After that, four participants were required to stop moving when they experienced the switching point. The taken pictures then were overlapped to

measure mean deviation of head angles when the switching point happened, the result was smaller than 4 degrees. The experiment reveals the similarity of image processing function between individuals. From this evidence, another concern raises up: how two individuals can have the same sensation and method of visual processing. Using the neuron network model to explain this question, by somehow is not reasonable. Assume that each person gets those abilities independently and their cognitive functions are determined only by forming connections network inside their brain. As human brain contains 86 billion neurons, and each person has different input data set, the probability in which two people create two neuron networks, which produce the same way of decoding the pattern of signal, is very low. (Grossi, 2007). The more reasonable explanation is that a part of their cognitive functions is inherited from a common original or common ancestor. And if they come from a common ancestor, they are embedded in DNA.

2.3 Potential applications

Studying DNA algorithms can provide many useful information in biology research and medical application. However, with the eyes of an engineer, one could find potential applications of nature algorithms in his field as well.

- Artificial intelligence: AI technology has gained huge improvements in recent years, however some statistical approaches in machine learning and AI are getting close to their maximum limit potential. To solve complex problems, these algorithms may require both strong computer re-sources as well as sometimes time consuming due to the large number of learnable variables used in the training process (Browne, 2020). Artificial vision algorithms usually lack flexibility, poorly adapt with unexpected changes from the operating environment and input datasets. The inflexibility of current AI model cause by the high dependence on the quality and quantity of the training dataset. For example, assume an AI model is designed to detect dogs and the training dataset only include pictures of dogs taken in sunny days. If the model is asked to recognize the dog from a picture, however some noise is introduced by putting a layer of rain drops or a big obstacle before the target dog, the prediction accuracy drops markedly. The example shows that adding factors not included in the training dataset will confuse the AI model. In another situation, natural organisms appear to be more flexible in handling unexpected changes even though they may not have encountered such situations before (Fukushima,1982). Another challenge AI scientists are facing is to build a mathematics model to mimic biology organisms cognitive. Inspecting DNA algorithms as well as its related "biology hardware" can provide data to solve these challenges.
- Data storage technique: The birth of high throughput sequencing technique has opened the door to store data in DNA. With the old technology, it takes scientists nearly twenty years to decode the whole human genomes. Nowadays, the process takes only more than one day. One prominent advantage of this idea is the high density of information storage and high durability. But there exist several risks which need to be considered. First, the DNAs are attacked by radiation or virus, hence it tends to cause mutations which broke integrity of data. Second, cross contamination can add unwanted noise to data. Finally, the transfor-

mation phenomenon of DNA also gives a risk to inadvertently create harmful strains of bacteria. In nature, organisms have developed sophisticated error-correction mechanisms and strong immune systems to protect their DNAs with high precision. Understanding this mechanism could help to develop new data storage materials in a trustable way. (Yiming, 2020 and Antkowiak, 2021)

- Universal smart robots: A robot which can perform multiple tasks and adapt with the change in operating requirements is always desirable. In many situations, robot's versatility features are required more than productivity. For example, a family needs a plow, seeding machine, watering machine to handle a garden in 3 days, then keep them in storage for the rest of year. Otherwise, it takes a general-purpose robot 2 weeks to finish all the tasks in the garden, but then the users can use it to support other tasks in their house. In this case, robots do not provide the highest productivity for each task of garden work, however it is more efficient overall. Natural organisms often evolve towards multitasking; therefore, properties of natural models can be learnt to implement universal robot idea.
- Quantum problems: In the recent year, quantum mechanism is introduced and used to explain biological phenomena, for examples photosynthesis, enzyme catalysis, avian navigation, etc. Some hypotheses about relationship between cognitive and quantum have been proposed. In this scenario, biology algorithms itself cannot explain the phenomena, but it can help to trace the signal paths and provide data for further examination. If the hypotheses about quantum's usage in brain is correct, then it will be the premise for building quantum computers that work at room temperature. (Fleming, 2011 and Alessandro, 2011)

3 GENE EXPRESSION ANALYSIS APPROACH

3.1 Gene expression and motivation of Boolean network models

Instruction information encoded in DNA is expressed through two processes: transcription and translation. First, DNA is unwound and serves as a template for mRNA synthesis in the cell nucleus, this process is called transcription. The mRNA molecules after that leave the nucleus to enter the cytoplasm, here they are translated to proteins, which conduct the cell's functions. This process is called translation (Goldberg,2017). By measuring the quantity of mRNA at given points of time in the cell, people can know exactly what segment of DNA is expressed and which proteins can be produced. If a gene is expressed through mRNA, it is turned "ON", otherwise it is turn "OFF". The arrangement of "ON" and "OFF" state is important. All the cells in the body share the same genome, but with different combination of gen turned "ON" at different point of time, they can express different behaviors and become different types of cells. The regulation process on the gen is determined by many factors, such as protein interaction in intracellular environment or stimulation from extracellular environment. The figure 2 shows the expression process of genes in eukaryotic.

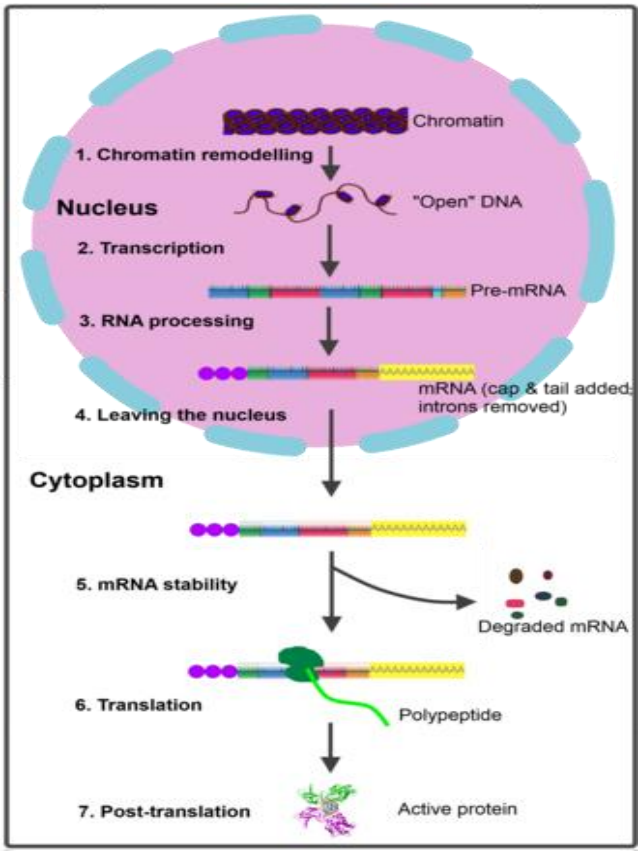


Figure 2: Gene expression process in eukaryotic (CKRobinson, 2015).

Let's take an example to examine how this process can be operated. Beside the coding regions carrying the information for protein synthesis, DNA are also equipped with regulatory regions and ending regions. Regulatory regions are usually located before the start point of transcription site of DNA, and this region carry several regulatory sites. These regulatory sites can be recognized by a regulatory molecular such as a particular protein or an RNA-protein complex. Once regulators molecular bind to a regulatory site, it can activate or inhibit the transcription of a gene. Notice that these regulatory factors are also the products of other genes expression processes. Hence, each gene on the DNA

may have a role to turn “ON” or turn “OFF” other gene. Suppose a bacterium response to a surplus of amino acids environment. It will turn “ON” a set of genes in a tight arrangement to metabolize the amino acid, while turn “OFF” a set of other genes to synthesize the amino acid. To do that, first some genes are induced and self-regulated by the stimulate from environment. Activator proteins are then produced and bind to other regulated sites on DNA near by the starting point of metabolizing gens to switch them on, while repressors protein bind to starting point of synthesizing genes to switch them off.

The mechanism of self-expression by the stimulation from environment, that is mentioned above, is also important. The expression of a gene does not only depend on effects of other genes, but it is also determined by the environment factors. Please follow an example about RNA thermometer to have an overview about this problem. RNA thermometers are a family of RNA molecules which form a clip-liked 3D structure by base pairing of distance nucleotides. When there is a change in temperature, their pairing bonds become unstable and tend to break down, hence their 3D structure and conformation are altered by the fluctuation of temperature. When the mRNA strands are unwound, it allows ribosome to access and start the translation process. This mechanism helps the bacteria response effectively with the change of environment. (Kortmann,2012) This example demonstrates that the information of operation does not store in the controlling code itself but also depend partly on their surroundings. Hence, to successfully apply reversing technique on biology code, it is necessary to put it in its operating context. Please keep it in mind as it the motivation for extension of Boolean network model later. Figure 3 shows an example of mRNA thermometer. In the Left picture, the mRNA thermometer forms the clip-liked structure, blocks the initiation codon (AUG) in the middle of the clip. Ribosome cannot bind to the initiation site and hence the translation cannot start. In the right picture, the mRNA thermometer is unwound by high temperature, expose initiation codon. Ribosome now can bind to this site and start translation and produce proteins carrying the functions that help bacteria response with hot temperature.

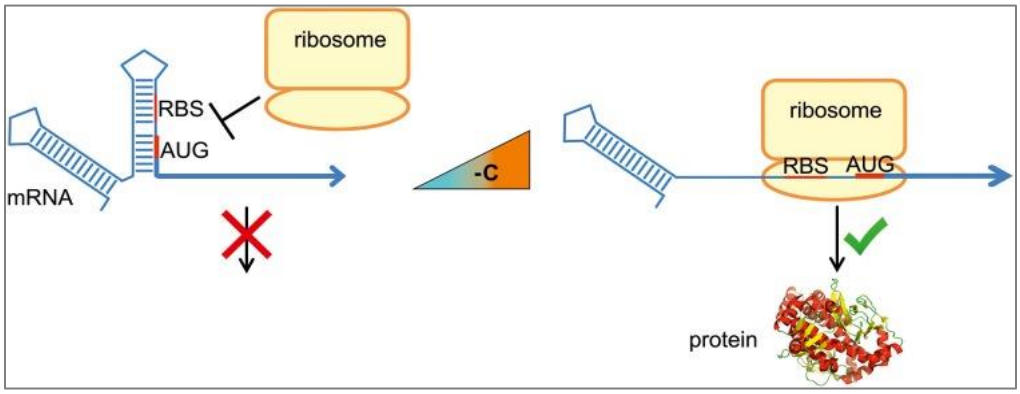


Figure 3: Gene expression regulated by mRNA thermometer. (Righetti and Narberhaus, 2014).

The description above shows the regulation effect of a set of genes on other gene, which act as the switch “ON” or “OFF”. In fact, these relationships establish a complicated network and each component of the network have causality effect on others. Each components provide a fundamental function, then they are combined in a tight arrangement way to create more complex function. Therefore, a directed graph model can be used to represent this type of networks, where each vertex stands for a gene and each edge stands for regulated functions assigned to this pair of vertices.

The existence of algorithms underlying behind this network can be used to explain two phenomena in biology. The first paradox is about C value:" The original C-value paradox was based on the perplexing observation that total genome size was unrelated to organism complexity, which in turn was taken as a plausible indicator of the number of genes in a genome." (Elliott,2015). The observations from the programing engineering field show that the length of an algorithm does not determine the advance of it. Some of the advance algorithm can be shorter than the others, however with clever methods they can solve the problem better in shorter time. The pair of plus and multiply are a good example. The multiply are result of adding a same number for multiple times. This definition can be used to calculate the results, however, human has created a more efficient way to multiply. They store the result of fundamentals multiplies in a table and use them to calculate more complex calculations. By using this method, the length as well as the time for multiplying decrease many times. The same phenomena could happen between the length and the complex of a genome, where more advance creatures own a set of better algorithms.

The second phenomena are intron-exon ratio difference in prokaryotic vs eukaryotic. The intron and exon are both segments of an immature RNA. Introns will be spliced out from RNA to create mature RNA, which includes only exon, then mature RNA is translated to proteins. Therefore, intron is usually called the non-coding sections, sometimes is refer to the junk DNA. In the prokaryotic and simple organisms, the introns sections are almost absent, while in their complex counterparts the intron section can count to 45-50% of the genomes, such as in human. Although the hypothesis of junk DNA cannot be rejected, the existence of algorithms in DNA can add a hypothesis to help explain this phenomenon. Back to the mathematics issue in section 2.1, notice that parentheses and bracket can be used to determine the order of operations. When the number of variables and operators increase, to arrange the order between calculation, number of parentheses and brackets can also increase and sometimes they create a complex combination. Considering a controlling program formed by a set of simple functions, the ratio of the code segments used to arrange the coordination between functions will be increase when the number of functions increases. This phenomenon can be observed clearly in website programing field with MVC model. When a simple static website is built, the ratio for the visual components usually counts for high percentage in the code. In contrast, in complex dynamic websites which interacts with the user and database, the ratio of code for defining visual components decrease while, ratio of the code segment for controlling cooperation between these components increase quickly. The same problem can appear in a complex organism, as they contain more functional proteins than the simple one and need more mechanism to coordinate these proteins. The non-coding sections may have a role for these tasks and therefore they are found more in eukaryotic. In fact, functional benefits of introns in genomes have been addressed such as mRNA alternative splicing, enhance the expression, etc.

3.2 Gene expression data and input of Boolean network models

Boolean network model used gene expression data in time series as input to inspect the dynamic of the biology system. To obtain gene expression data, types, and number of mRNAs in cells are collected and measured. Large amount of cell samples is taken in a sequence of time to record the transcription in time series. After data is normalized and is analysis, each sample yells a single vector of cells state

with size equal to number of gens in the cell. Each element in this vector take value 0 or 1, genes switched "OFF" at time τ are assigned to value 0, while genes switched "ON" are assigned to value 1. After repeat the process for all n samples, scientist can collect a set of cell state vectors which is represent the evolution of cells in time series. These data can be also represented in table form or matrix form. Two main methods used to measure the expression of genes are microarrays and RNA-sequencing.

- Microarray: The microarray foundation bases on the complementary DNA fragment synthesis and fluorescence labelling. Firstly, a microarray chip is printed with thousands tiny pots, each pot is coated with multiple identical probes which are short segments of DNA. The RNA population of samples are harvested from target cells and introduced into a container. In this container, reverse transcriptase is added, and reaction of complementary DNA synthesizing is started using harvested mRNA. Two different labeling fluorescent dyer such as Cy3-dUPT and Cy5-dUPT are used to mark the origin of cDNA, than two samples are mixed in small volume and put on the microarray chip. The cDNA in the mix will pair with the probes if they are complementary, the remnant cDNA remains unbound and move freely in the sample liquid, this process is called hybridization. After that unbound cDNA is washed over the chip and laser scan is applied on each tiny pot which reveals which cDNA is presented at this location. The experimental process could encounter with various errors, which finally add noise to the data. For example, cross hybridization can happen if there is a weak reaction between cDNA and probes, or multiple cDNA can accumulate and bound to a probe. (Govindarajan, 2012)
- RNA sequencing: RNA-seq also uses complementary DNA fragment for operating, but instead of using a chip, the procedure applies short-gun DNA sequencing or next generation sequencing techniques to measure the present of cDNA in the sample. Similarly to microarray, first mRNA is harvested from cells, but it will be cut into smaller section which have the size 100 -1000 bases. Then it is put in an environment with the present of reversing transcriptase, it leads to the synthesize of the first strand of DNA. Now DNA polymerase are added to complete the transcription, results in double strand cDNA fragments. Next step, these fragments are ligated with small special designed DNA sections called adapters, these adapters play important roles to recognize the start and end of cDNA fragments, or to amplify the amount of these cDNA. With a library of cDNA fragments, now once can use variety techniques to decode the cDNA such as sequencing by synthesis. Thousands of cDNA reads are collected, the results are then put in a computer program to compare the overlaps between these cDNA fragments and mapped them to the original RNA sequence. (Elyasigomari, 2017)

Gene expression data usually contains vast number of elements. These elements are usually put together in matrix or table form, where columns stand for the sample (or the context that the sample is taken) and the rows stand for the individual genes in the gene list. Raw data is usually converted to transcripts per million (TPM) and fragments per kilobase per million (FPKM), the purpose of using these units is to make the data become comparable across the genes. Another way to normalize the data is to use expression level of a sample as a reference value, other test sample value will be calculating as log value of relative ratio to this reference value: relative express level = log (test sample/ reference sample). Figure 4 shows the gene expression data in heat map form.

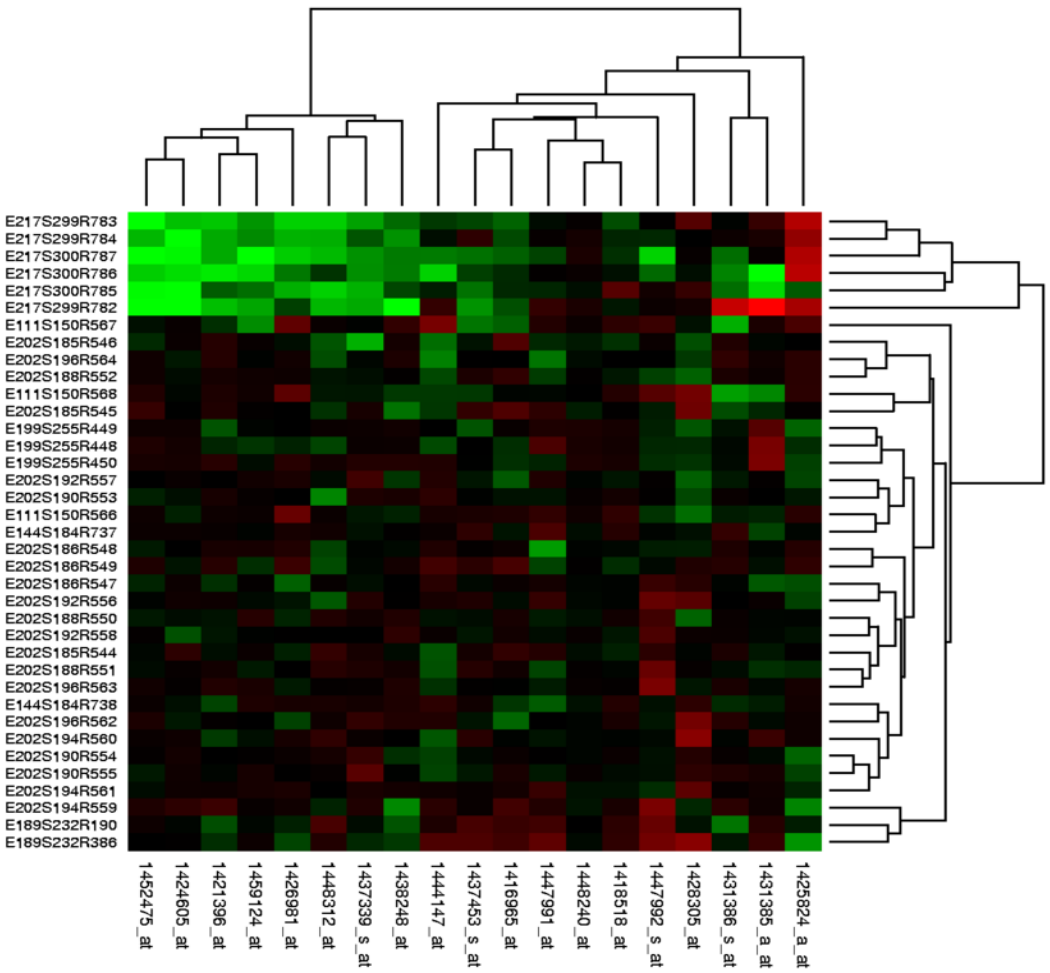


Figure 4. Gene expression matrix represented as heatmap form (Miguel Andrade, 2016)

Fielden and Zacharewski (2010) has addressed limitations of gene expression profiling techniques. Firstly, the sensitivity of our measurement is not good enough to detect subtle changes in gene expression level, even some changes are subtle, but they may affect strongly to the cell operations. Secondly, current knowledge about regulation mechanism is still incomplete but noised data is usually introduced during experiments by human mistakes. Finally, samples population in different environment context is still limited, therefore there are not enough observations to study about cell reactions in these contexts. In the scope of this thesis, one more issue will be discussed about current methods: they destroy the information about spatial distribution of mRNA in cells, which may have important meaning with gene expression process. Recent studies show the evidence that a mRNAs are usually distributed to some distinct cytoplasmic domains (Martin, 2009), it is different from the old hypothesis that believes cells operations is a chaotic process and molecules inside the cell moving freely and collide with each other almost randomly. In fact, in some situation directional flow of molecules can be observed in the cells. Let take a closer look at RNA splicing process, where introns segments are cut off from immature mRNA in nucleus of the cells. Notice that the environment in the cells is contains 70% water, and the buoyancy repulsion of water balances out with a fraction of gravity, creating a microgravity-like environment. When introns are folded and sliced away in microgravity-like environment, it can create a vector of force which direct exons-spliceosome complex toward a special sub-region in nucleus and increase probability of this mature mRNA to be transported to a particular cytoplasmic domain. Hence, the probability of which a protein X is produces in location Y of the cells, or a reaction between X and X' happened at location Y, may increase. Figure 5 shows the process of mRNA splicing in three steps. In the first step, spliceosomes

are formed by the combination of five subunits snRNPS. Each snRNPS binds to a special site in intron, but they do not bind in any exons. In the step 2, after binding to intron, the electrostatic forces pull snRNPS closer together and distort the structure of mRNA, the intron segments form a loop. The junction points between exon and intron are pulled closer and finally joined. Finally in step 3, the intron section is cut off; exon sections are connected to create final mature mRNA. A vector $\vec{v}$ give mRNA a direction and drive it to a specific subregion before it leaves nucleus to cytoplasm. Introns and snRNPS are released, they may have other functions as well.

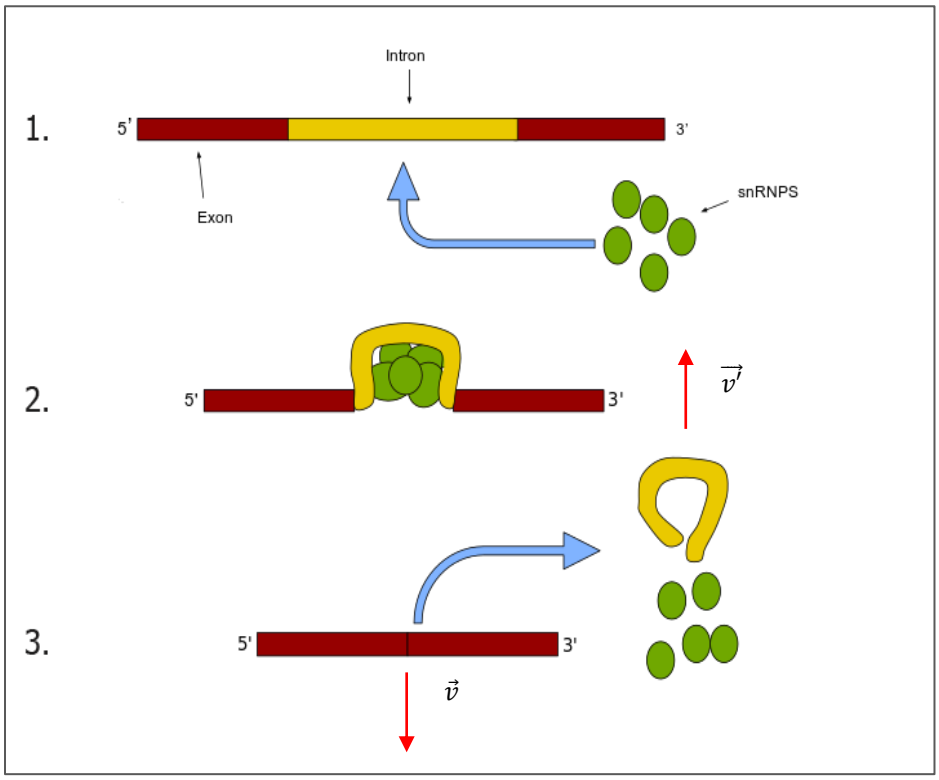


Figure 5. mRNA splicing (Source: Iinaba, 2013)

Direction, localization of mRNA, which affect the probability of events taking place in the cells, may have an important role in regulating cell activities. In 2002, Ilya Shmulevich and his research team proposed a probabilistic Boolean networks model to solve the limitation of conventional model. A new parameters called bias was introduced to the model, which stands for the probability of a function in the network to change its behaviors. The institution was that a system did not working basing on only one fixed set of functions, but the environments could alter the behaviors of several individual functions, in this case the functions set need to be updated. He then examined the relationships between bias value, in-degree K and the stable manner of a dynamic system. It was stated as: "Generally, a dynamical system is said to behave in an ordered or robust manner when similar initial conditions will lead to similar trajectories in the phase (state) space" (Shmulevich, 2002). The results shown this bias values strongly correlated with the stable manner of cells. Back to the discussion above, it is obvious that the direction effects from mRNA splicing could be one factor contributing to this bias. Observing in the figure 5, the probability of mature mRNA to move downward is higher than to move upwards; it tends to enter the bottom area of cytoplasm and have higher possibility expressing here rather in opposite direction. Therefore, it may hold an important meaning.

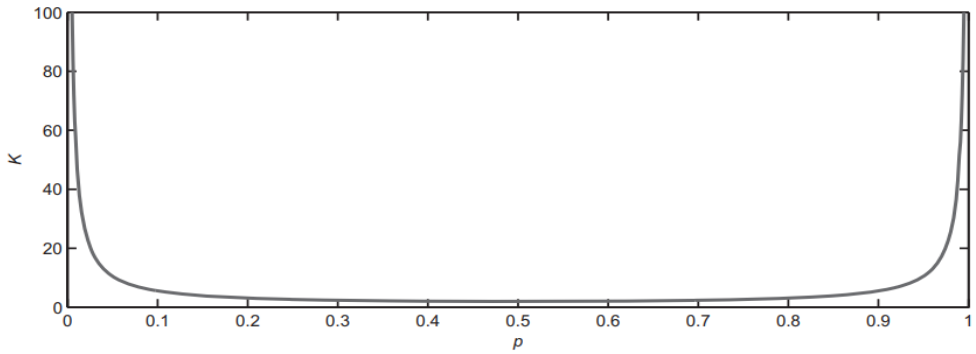


Figure 6. The critical value of a Boolean network. (Shmulevich, Dougherty, 2009)

Figure 6 shows the result from the research of Shmulevich (2009) and his team when they studied about the robustness of biology system despite of the disturbance from external factors. The critical value of a Boolean network is presented as the black line. Any combinations of p and K above the critical line will result in chaotic behaviors. Unfortunately, current methods of mRNA collection require cell disruption, which disturb the components in the cell. Information about mRNA localization then disappear and cannot use to analysis. In the future, this type of information should be considered and represented in the network for further analysis.

3.3 Boolean network model

Coming back to main purpose of the thesis, from the gene expression data and mechanism observed from the cells, a model can be built to mimic the operation of the cell and then to extract algorithms embedded in the genome. Suppose the genomes is made up of many fundamental units (a unit can be an exon segment, an intron segment, or a regulatory segment on DNA), each unit offers a fundamental function. Then fundamental functions are arranged in a logical order to create more complex functions, which are used to drive the biology system's behavior in a deterministic way. To discovery these algorithms hiding in the genomes, first these unit need to be identified accurately. Then it is necessary to find the right order of which these fundamental functions are arranged together. Finally, the system is put into surrounding context to analyze operating mechanism. Identifying fundamental unit can be done by DNA sequencing and classifying them into different categories. The second step is to discovery the arrangement between the functions of fundamental units and put them together in a network. Among current methods used to analysis genome, Boolean network are one of tools that can solve this problem. One interesting possibility of the Boolean network is that it can be used to detect the pattern of the cyclic data and infer the cause-effect relationships between the components of the given system. This ability will be represented in detail in the next parts.

3.3.1 Definition of Boolean network

A Boolean network is composed of two main components: a set of genes $V = \{v_1, v_2, \dots, v_n\}$ and a set of functions $F = \{f_1, f_2, \dots, f_n\}$. The gene v_i in the set V represents for a gene in real cells and the state x_i of this gene can take two value 0 or 1. When gene v_i is turn "ON", the corresponding value x_i = 1; when gene v_i is turn "OFF", the corresponding value x_i = 0. Each function f_i in F is a predict function, which maps n variables ∈ {0, 1} to single value of 0 or 1: $f: \{0,1\}^n \rightarrow \{0,1\}$. There are four fundamental Boolean functions: the AND operation which is denoted by "∧" symbol, the OR operation

denoted by "∨" symbol, the XOR operation denoted by "⊕" symbol and the NOT operation denoted by "–". From the description about the gene regulation process in last discussion, suppose a gene v_i is regulated by a set of genes $K_i = \{v_{k1}, v_{k2}, \dots, v_{km}\}$, then the value of gene v_i in time series is determined as equation below:

(2)

$$x_i(t+1) = f_i(x_{k1}(t), x_{k2}(t), \dots, x_{km}(t))$$

where $x_i(t+1)$ represents the value of gene v_i at time t+1, the $x_{k1}(t)$, $x_{k2}(t)$, …, $x_{km}(t)$ represents the value each gene in set K at time t. At any time t, a vector can be used which contain values $\{x_1, x_2, \dots, x_n\}$ of n genes $\{v_1, v_2, \dots, v_n\}$ to represent the state of cell at that moment X(t) = $[x_1(t), x_2(t), \dots x_n(t)]$, which is also called gene activity profile (GAP). From the way to create a state vector, it can be estimated that there are 2^n possible states of a system containing n genes, it is an enormous number which lead to challenge in computation (Akutsu,2008). From the above definitions, two graphs and a table can be built, including transition functions truth table, interaction graph of the network and state transition graph of the cells. The interaction graph represents the causal-result effects of genes to one, the state transition graph shows the dynamic of the system. Two other concepts need to be defined are the maximum indegree K, which is the largest number of connections toward a node in the gene set V. In fact, genes in biology system are not fully connected and in the network inference problem, iteration over the searching space is a common strategy (Kauffman, 2003). Therefore, using fully connected model and searching will be redundant, the better way is determining the maximum indegree K and restrict the searching space by this variable. The other concept is discretization level of the gene expression. In some circumstance, biology system may use a set of thresholds to determine their response and regulate their behaviors, and the discretization level is not necessary binary "ON" and "OFF". It can be "LOW", "MODERATE" and "HIGH" for example. To generalize this problem, a parameter d stands for discretization level of expression is introduced into the model. In this case, the set of predict functions F should have $f_i: \{0,1, \dots, d\}^n \to \{0,1, \dots, d\}$. In this document, the binary discretization level is used. The figure 7 demonstrates simple example about a fundamental function set $f_i: \{0,1\}^2 \to \{0,1\}$.

AND

x_1	x_2	$f(x)$
0	0	0
0	1	0
1	0	0
1	1	1

OR

x_1	x_2	$f(x)$
0	0	0
0	1	1
1	0	1
1	1	1

XOR

x_1	x_2	$f(x)$
0	0	0
0	1	1
1	0	1
1	1	0

NOT

x	$f(x)$
0	1
1	0

Figure 7. Truth table of fundamentals Boolean functions (Lai, 2023)

Let's examine an example about a Boolean network (BN) which include a set of five genes $n = 4$ and the maximum indegree $k = 3$. The set of function F is: $x_1 = (x_2 \wedge x_3) \vee x_4$, $x_2 = x_1 \wedge x_3$, $x_3 = \overline{x_3}$, $x_4 = \overline{x_2} \vee x_3$.There is total $2^4 = 16$ states that the system can hold in the example. Then state transition table and graph can be constructed as in figure 8 to describe the Boolean Network. There is an attractor in the Boolean network, it includes two states A = {0101, 1010}, the set of states {0000, 0001, 0010, 0011, 0100, 0111,1001 1111} are called the basin B of attractor A.

t				$t+1$			
x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
0	0	0	0	0	0	1	1
0	0	0	1	1	0	1	1
0	0	1	0	0	0	0	1
0	0	1	1	1	0	0	1
0	1	0	0	0	0	1	0
0	1	0	1	1	0	1	0
0	1	1	0	1	0	0	1
0	1	1	1	1	0	0	1
1	0	0	0	0	0	1	1
1	0	0	1	1	0	1	1
1	0	1	0	0	1	0	1
1	0	1	1	1	1	0	1
1	1	0	0	0	0	1	0
1	1	0	1	1	0	1	0
1	1	1	0	1	1	0	1
1	1	1	1	1	1	0	1

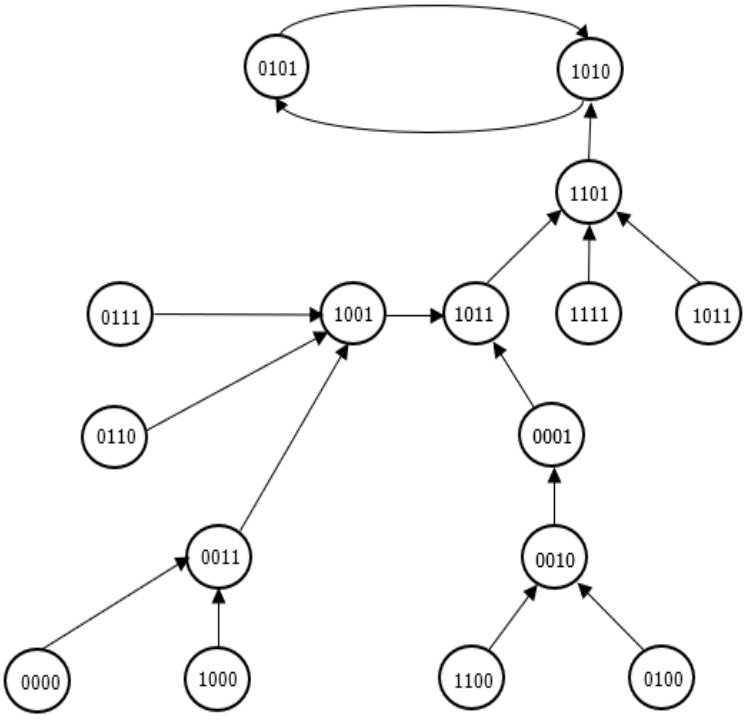


Figure 8. The transition table and the state transition graph. In the graph there is an attractor which contains two nodes 0101 and 1010. (Lai, 2023)

In previous study, to illustrate the relationship between genes in the system, the authors usually use a directed graph G (V, E) where each vertex represents for a gene and each edges shows the direct casual affect between two genes. However, this presentation method will lead to the loss of information about the order of influence of the regulated gene set on the target gene. The mathematic

issue discussed in chapter 2 shows that the order of operators is important to the result of a calculations. A graph which is described as above cannot represent this order. Let take an example from $x_1 = (x_2 \wedge x_3) \vee x_4$, the conventional graph will only provide information that x_1 are regulated by the set (x_2, x_3, x_4), but all information about operator type and their arrangement will be dropped. This causes ambiguity. If color codes is used to mark the operator type for edges: red for ∨ and green for ∧, once can rebuild six different arrangements from graph, and each arrangement will produce different result, for example $x_1 = (x_2 \wedge x_4) \vee x_3$ or $x_1 = (x_3 \wedge x_4) \vee x_2$. Hence, another way is to use Polish notation methods to fully represent the relationship between networks components and their operators. In figure 9A, the regulation graphs G (V, E) represents the simple causal-effect relationship between each node, in 9B a Polish notation graph shows both operators and order of between them.

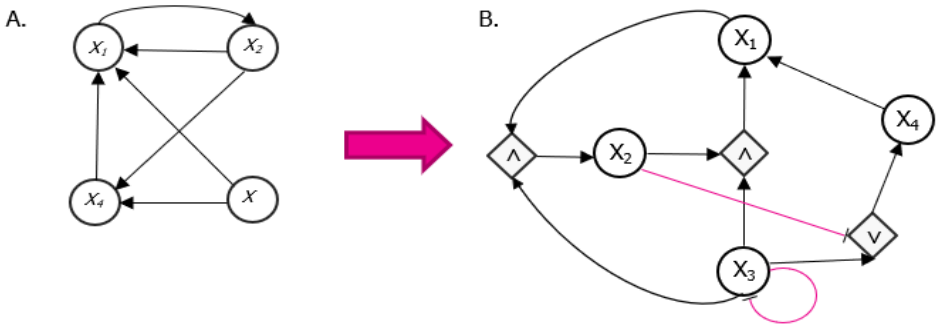


Figure 9: A. The regulation graphs G (V, E), B. In Polish notation form (Lai, 2023)

3.3.2 Characteristics of Boolean Network.

Boolean Network structure hold information about the flow of operators in the network. In biology scenario, it stands for the order of event happening inside the cells. Let's examine some important characteristics of the network.

- The existence of attractors: the attractors are cycles of a given Boolean the network. In general, because each evolving steps of the network is deterministic, a network can start from any random node, then gradually enters a set of fixed nodes called attractor A = $\{a_1, a_2, \dots, a_p\}$ Once the network enters the attractor, it will come back to the first node after p steps and be trapped in a loop forever, unless there is perturbation or intervention from environment. In biology scenario, attractors can be explained by the differentiate process of the cells. In fact, cells in the body share a common genome, but under different conditions they start to differentiate into a certain type of cell in the body. Other nodes, which are not attractors, but they have at least one pathway connecting to attractor, are called basin of this attractor. A network which has N node and K maximum indegree if N > K there will be at least one attractor.
- Canalizing functions: Living cells are considered a dynamical system, which a small change in any components of the system can lead to the large changes in the system operations and behaviors. However, cells usually represent robustness again perturbation. One reason that can explain for this capacity is the existence of canalizing functions. A canalizing function f is a function in which one or several variables can determine the output of f regardless the value of other variables. In more formal, given a function $f: \{0,1\}^K \rightarrow \{0,1\}$, if there is existence of $x_i \in \{x_1, x_2, \dots, x_k\}$ so that when $x_i = a$ then $f = b$ is always hold regardless the value

of other variables $\{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_k\}$. For example, the function $f = x_1 \vee (\overline{x_2} \wedge x_3)$, then f is a canalizing function because when $x_1 = 1$, the value of f will always be 1. Results from practical experiments have shown the evidence in eukaryotes to use canalizing functions to resist the chaos in their behaviors.

- Canalizing combination: From the example about canalizing functions, the problem can be extended: there exist a set of functions that act together to keep the value of a node's constant. For example, given a set of gen $\{x_1, x_2, x_3, x_4\}$ and a set of functions $\{f_1, f_2, f_3, f_4\}$. Assume $f_2: x_2 = \overline{x_1}$ and $f_3: x_3 = x_1 \vee x_2$, then regardless of form of f_1, f_4 the value of x_3 will be keep constant 1. Now for any reasons, if the system needs to turn gene x_1 OFF, or gen x_2 OFF, then gene x_3 is guaranteed to always be turned on (Schwab, 2020)
- The existence of communities in the graph: A fundamental characteristic of a network is community structure. A community of a graph is a subset of nodes which form dense connections together but form little connections with the nodes outside of the subset. In the biology case, organisms are usually composed from many organs; each organ performs certain functions. These functions are usually directly related to the characteristics of that internal organ and composed from many sub-functions; sub-functions closely support the main function. From this intuition, communities in the gene expression graph can form community which represents for the function of a certain organ. For example, a community of nodes which control the livers will connect densely together, but it can connect loosely with the nodes which control the kidney.
- Stem cell problem, convergence and divergence of the network: Stem cells are a very special case in biology. Immediately after fertilization, pluripotent stem cells are formed. Through differentiation processes, cells gradually lose potency and can only differentiate into higher lineage cells. However, a small number of stem cells are still found in tissues throughout the life of the organism. It means mitotic replication still occurs in stem cells, so there exists an attractor that reflect the self-replicating process of stem cells. On the other hand, all types of other cell start from stem cells, hence there also exist other edges that come from the initial state but not belong to stem cells self-replicating attractor, and they represent the process of differentiation into different cell lines. Another feature to notices is that the higher lineage cells are spreads from a stem cell and rarely spontaneously dedifferentiation back to the state of pluripotent stem cell. Therefore, the gene expression graph should have a divergence tree structure, with the root of the tree attaches with an attractor, and there are multiple branches spreading out from the root. In fact, it is impossible for a conventional synchronous Boolean network to form divergence structure in which a network spreads from a central root (model 1 to many), because each mapping function creates only single successor state from one or multiple original states (model many/1 to 1). There are two possible explanations for this divergence tree structure of a Boolean network. The first explanation is that some genes are affected by external environmental factors such as viscosity, pressure, concentration of substances in the extracellular fluids, and their ON/OFF values are also switched by these factors. In this scenario, the layers of these environment factors should be introduced to the model for further examination. The second explanation is the exist of internal stochastic processes of the cell, where genes are perturbed and randomly switched ON/OFF. Then it is possible to

create more than one successor states from a single original state. This problem can be well described by probabilistic Boolean network model.

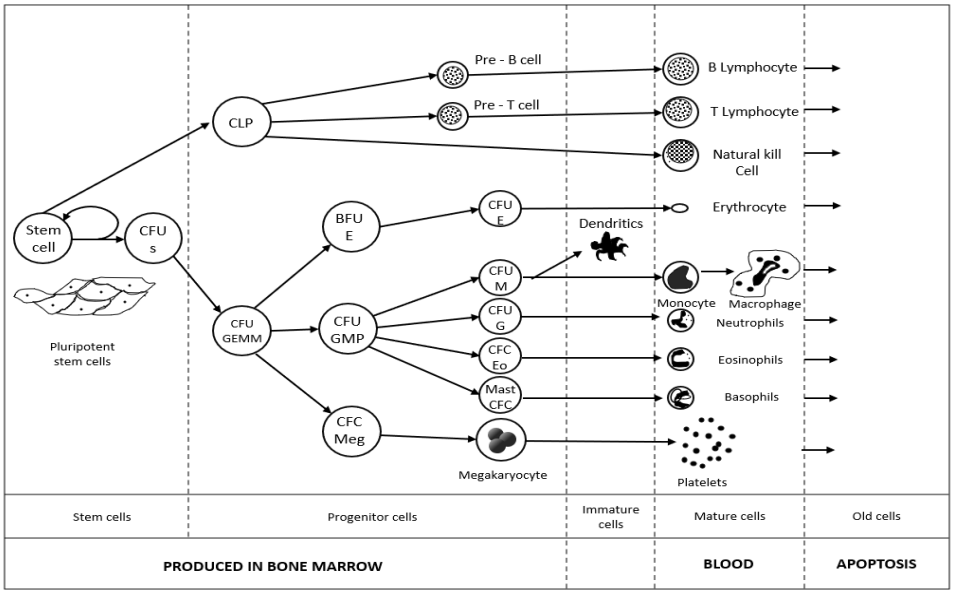


Figure 10A. The differentiation from pluripotent stem cells to different type of blood cells.

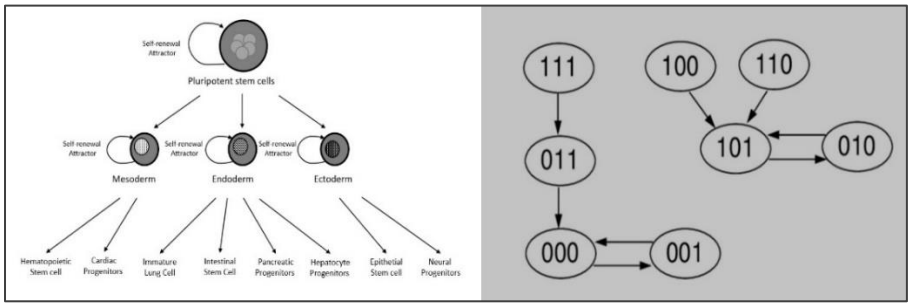


Figure 10B. Comparison between structure of a state transition graph in synchronous binary Boolean network versus reality observation. (Source: Lai, 2023)

Figure 10A and 10B demonstrate the comparison between structure of a synchronous Boolean network and the cell differentiation process. In the left graph of figure 10B, the network spreads from a central parent node to multiple child nodes. Assuming pluripotent stem cell is the initials state of the system, there are three successive states created from this state, they are Mesoderm, Endoderm and Ectoderm. This phenomenon corresponds to a one-to-many mapping function and it produces a divergence structure. However, in the right graph of figure 10B, from a single or from many different input states, only one output state is produced. Therefore, the system will gradually evolve and be trapped in two attractors (000,001) and (101,010), the transition graph structure now is a combination of two convergence blocks. This contradiction shows that the conventional synchronous Boolean network is incomplete and does not adequately describe the gene expression process.

- Nested loops and gene hierarchy system: two loops are said to have a nested relationship if they have a set of common vertices. To be able to distinguish the inner nested loop and the outer loop, it is necessary to assign a hierarchy score for each loop. The hierarchy score of a vertex can be calculated by the number of descendant vertices which can relate to it by at least one path. A vertex can be called a mother vertex if all other vertices can be reached by a path to it. Similarly, a hierarchy score of a loop is the number of all the descendant vertices of the loop's member

vertices. The hierarchy score implies the importance of a node or a loop, a change happening in the higher hierarchy vertex will lead to the change of more descendant vertices. For mother vertices, its change can cause a change at the global level, while the changes in other vertices cause changes at the local level. This can make sense when this property is used to analyze regulation graphs of Boolean network model. From observations in practice, the functions in the organism are divided by hierarchy. Let's take the example, sugar regulation functions can be divided into two levels: cellular level, and organ level. Since sugar levels regulation is essential activities that virtually every cell needs it, it is likely that sugar regulation genes could have a high position in the hierarchy system. By analyzing the hierarchy system of genes, the set of vital genes and vital loops can be found, and the necessity level of a particular gene function can be distinguished.

3.3.3 Different paradigms of Boolean Network

The conventional models of Boolean Network mentioned above are usually known as synchronous Boolean Network. There are two more paradigms. The first one is asynchronous Boolean Network models (Garg, 2008). Another paradigm is probabilistic Boolean Network (Dougherty, 2008). Both models can create a network having a divergence structure, which consistent with real-world events.

- Asynchronous Boolean Network: The above model is Random Order Asynchronous first introduced in 1997 by Harvey and Bossomaier. In contrast, asynchronous Boolean Network based on the assumption that genes take different periods of time to update. In the model, only one gene changes its state at a time. After a sequence of N updates, the whole set of N genes will be updated one time (N equals to the size of the gene set). The order of changing sequence is totally random. Hence, the asynchronous network has indeterministic characteristic. It also leads to a new concept of complex attractor (beside fixed node attractor and periodic cycles attractor), defined as a set of states A, which contain all possible states S and their successor states S'. Figure 11 show an example of a asynchronous Boolean network, which includes three nodes: $x_1 = x_1$, $x_2 = x_1$, $x_3 = x_1 \wedge x_3$. Gene x_1 takes four minutes to complete, the remnant gene take one minute to complete.

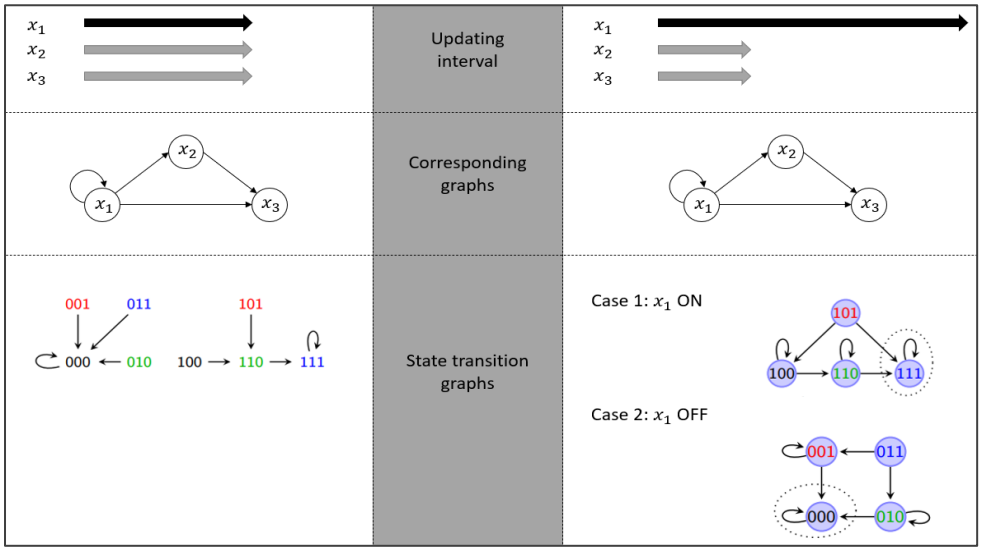


Figure 11: The synchronous Boolean network (left) and asynchronous Boolean network (right) (Matthew Macauley, 2016)

Although it overcomes the problem of converging structure, this model has some disadvantages that have been pointed out: the process of randomly selecting the update order leads to contradictions with actual events taking place in the cell, in particular the durations of biological processes. (Weidner 2021). Secondly, it cannot represent the dependence causal dependence between gene expression processes.

- Probabilistic Boolean Network: Probabilistic Boolean network is an extension of synchronous model, and hence it uses the same synchronous updating mechanism. The difference between probabilistic model and conventional model is the predict functions set F. In conventional model, there is only one fixed predict function f_i is assigned to gene x_i, hence the network is deterministic. In probabilistic model, it is assumed that there exists multiple predict functions, at this example is a list of l functions, $F_i : (f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(l)})$ assigned to gene x_i. At each step of the cell evolution, a function in the list is selected randomly, with the probability of $f_i^{(j)}$ selected is $c_i^{(j)}$ with $j \in \{1,2,\dots,l\}$. Then it results in $\sum_0^l c_i^{(j)} = 1$. The motivation behind the introduction of multiple sets of predict functions is from the uncertain of gene expression data and the effect from environment. In fact, intrinsic stochasticity of biological systems, exterior factors can alter chemical reaction in the cell, for example temperature can change the proteins shapes and conformations, then the function of this proteins can be modified, then function $f_i^{(t)}$ assigned to gene x_i at time t then can be switch to $f_i^{(t+1)}$ at the next step. Figure 12 shows the Boolean functions and dynamic of the system.

A.

	Boolean functions	$c_i^{(j)}$
$f_1^{(1)}$	$x_1(t+1) = x_2(t) \wedge x_3(t)$	0.3
$f_1^{(2)}$	$x_1(t+1) = \overline{x_2(t)} \wedge x_3(t)$	0.5
$f_1^{(3)}$	$x_1(t+1) = \overline{x_2(t)} \wedge \overline{x_3(t)}$	0.2
$f_2^{(1)}$	$x_2(t+1) = \overline{x_3(t)}$	1
$f_3^{(1)}$	$x_3(t+1) = \overline{x_1(t)} \vee x_2$	0.3
$f_3^{(1)}$	$x_3(t+1) = \overline{x_1(t)}$	0.7

B.

Realizations	$f_1^{(1)}f_2^{(1)}f_3^{(1)}$ 0.09	$f_1^{(1)}f_2^{(1)}f_3^{(2)}$ 0.21	$f_1^{(2)}f_2^{(1)}f_3^{(1)}$ 0.15	$f_1^{(2)}f_2^{(1)}f_3^{(2)}$ 0.35	$f_1^{(3)}f_2^{(1)}f_3^{(1)}$ 0.06	$f_1^{(3)}f_2^{(1)}f_3^{(2)}$ 0.14
000	011	011	011	011	111	111
001	001	001	101	101	001	001
010	011	011	011	011	011	011
011	101	101	001	001	001	001
100	010	010	010	010	110	110
101	001	000	101	100	001	000
110	010	010	010	010	010	010
111	101	100	001	000	001	000

C.

	000	001	010	011	100	101	110	111
000	0	0	0	0.8	0	0	0	0.2
001	0	0.5	0	0	0	0.5	0	0
010	0	0	0	1	0	0	0	0
011	0	0.7	0	0	0	0.3	0	0
100	0	0	0.2	0	0	0	0.8	0
101	0.35	0.15	0	0	0.35	0.15	0	0
110	0	0	1	0	0	0	0	0
111	0.49	0.21	0	0	0.21	0.09	0	0

D.

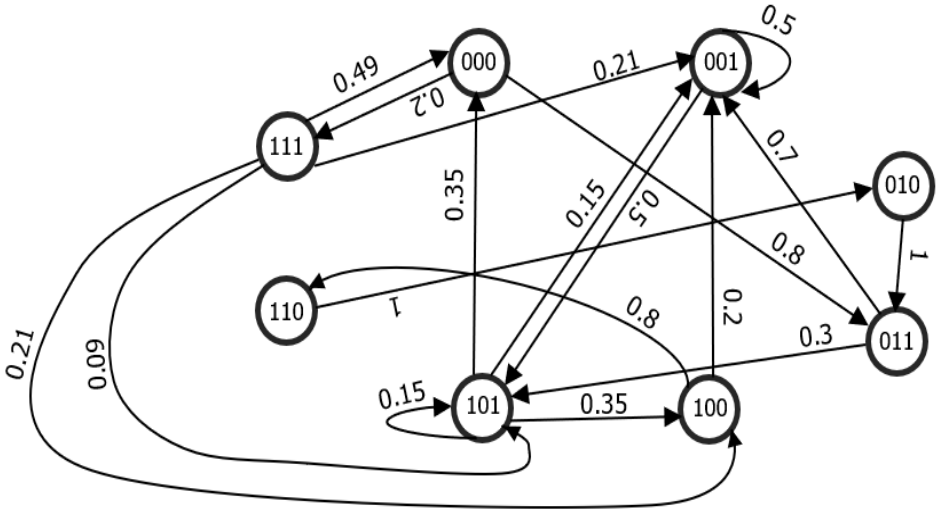


Figure 12: A – The function assigned to each gene along with their switch probability, B – Table shows the realizations and their transition states dynamic, C – The transition matrix a long with the probability of switching, D – states transition graph. (Lai, 2023)

In the scope of this thesis, because of limitation of the skills, experiments on the synchronous Boolean network will be reproduced and an idea to overcome its disadvantages is suggested.

3.3.4 Structure inference of Boolean network.

In the section 3.2, the gene expression data can be obtained by using decoding technique such as DNA microarray and high throughput next generation sequencing techniques. They provide large amount of data about gene expression in time series, which enable to observe the evolution of cells in timeline. If people can create a Boolean network which produce the pattern of events happening in the cells, compatible with data from the experiments, they will be able to reproduce the relationship and order between the functions in the cell. In other words, the arrangement of the functions set could be discovered from gene expression data, and it is an important algorithm.

Now assume different samples of cells in different states of cell phases are collected, and assume each cell have N genes. At each time, the state of cell is presented by a binary valued vector of size N, where each element in the vector stands for the state of one single gene (it is 1 if the gene is turned "ON", and it is 0 if the gene is turn of). At final a list of state vector $\{S_{t_1}, S_{t_2}, \dots, S_{t_m}\}$ from t_1 to

t_m can be collected. It is also known the maximum in-degree variable K and the fundamental regulation operators, then the purpose of is find a set of predict functions $F: f_1, f_2, \dots, f_n$ which can produce the best consistent data with experimental results. Notice that because the maximum in-degree variable is K, then each function in the set F have at most K variables.

$$\begin{cases} f_1: x_1 \square x_2 \square \dots x_{n-1} \square x_n \\ f_2: x_1 \square x_2 \square \dots x_{n-1} \square x_n \\ \dots \\ f_n: x_1 \square x_2 \square \dots x_{n-1} \square x_n \end{cases} \xrightarrow{\wedge, \vee, -} \begin{cases} f_1: x_1 \vee x_2 \wedge \dots x_{n-1} \wedge x_n \\ f_2: x_1 \vee \overline{x_2} \vee \dots x_{n-1} \vee x_n \\ \dots \\ f_n: \overline{x_1} \wedge x_2 \wedge \dots \overline{x_{n-1}} \wedge \overline{x_n} \end{cases} \sim \{S_{t_1}, S_{t_2}, \dots, S_{t_m}\}$$

The problem described above is an extension of the math issue in section 2.1. Instead of finding a solution for a single equation, the requirement in this case is find solution for a system of equations. Unfortunately, this problem seems to have received little attention and is rarely described in the existing literature, except from some ancient math problems. This may be due to the predefined order of combinations (default: multiply and divide first, add and subtract later), but this order is not necessarily correct in all practice realizations. It may be a problem near the boundary of current mathematical theory. Therefore, the consideration of mathematicians is necessary to solve problems scientifically. The number solution of the equation system could determine the type of Boolean network inference.

- Inference of the Boolean network: determine if there exist a solution of the equation system. If the answer is yes and there is more than one solution, this can be considered an inference the Boolean network problem.
- Identification of Boolean network: if the equation has only one solution, then the problem can be considered as identification of the Boolean network. The existence of unique solution is a desirable target when it comes to genomes analysis.

The simplest strategy to infer Boolean network problem is to iterate through all over the searching space and find the best fit networks which consistent with the data. The procedure to infer the network is:

- Prepare the input data as a list of dictionaries (X_i, Y_i) in which $X_i = S_t$ is the input of the system and $Y_i = S_{t+1}$ is the output of the system. This list is called D.
- Determine system variable including: the number of genes in the genome N, the maximum in-degree K, the discrete level of gene expression d, and finally the operator sets (in case of binary model, d=2 and the operators set are AND, OR, NOT and XOR (optional)).
- Generate all possible Boolean function f_i of a gene x_i which have at most K variables and store it in a tuple. The Polish notation structure is used to keep not only the information about variables but also the order of the variables. The progress is done for all N genes in dataset, the result tuples then combine into a dictionary for easily accessing later, this tuple is called G.
- Now from each tuple in list G, take one function, after that all possible predict function set F_j (also called realizations) are created.
- Start compare the result of each set F_j with the data in list D. Return the right network G.

Algorithm 1: *Inference of the Boolean network*
Input: *N, K, D, d, operation set, gene set V*
Output: *the prediction function set F*
1. Create tuples to store result;
$G \leftarrow \emptyset$;
$R \leftarrow \emptyset$;
$Result \leftarrow \emptyset$;
2. Create all possible Boolean function of gen x_i
For *all* x_i *in V :*
$L_i \leftarrow \emptyset$;
For *i = 0 to K:*
Create all possible Boolean function have i variables $f_j^{(i)}$;
$L_i \leftarrow L_i \cup f_j^{(i)}$;
$G \leftarrow G \cup L_i$;
3. Create all possible realizations
For *all* L_i *in G:*
Create all possible unique realizations r_i;
$R \leftarrow R \cup r_i$;
4. Compare
For *all* r_i *in* R *:*
If r_i *compatible with all* $(X_i, Y_i) \in$ D: *Result* $\leftarrow Result \cup r_i$;
Return *Result*

The iteration strategy is simple idea, but it leads to a big disadvantage of time complexity, and because of the enormous searching space, it may take a long time to compute the result. A possible way of reducing the complex of the problem is to use the support from laboratory experiment. The biologists can do laboratory test to check if there is any reaction between substances in the cell, from the results, the partial information about Boolean network can be collected. Now the problem reduces to inference of a network when knowing a part of it, and this problem is similar to graph completion problem. Another possibility way is to not create fully N x K connections in the model, instead of that a total 26 connections in the graph will be M connections, with M << N × K. The motivation behind this idea is that not every gene truly forms K connection with other K genes. In fact, a lot of them only create one or two connections with others, therefore a computing model iterating through all N × K possible connections is redundant. Similarly, the network inferring problem, the network identifying problem can also use iteration approach.

Algorithm 2: *Identification of the Boolean network*
Input: *N, K, D, d, operation set, gene set V*
Output: *the prediction function set F*
1. Create tuples to store result;
$G \leftarrow \emptyset$;
$R \leftarrow \emptyset$;
$Result \leftarrow \emptyset$;
2. Create all possible Boolean function of gen x_i
For *all* x_i *in V :*
$L_i \leftarrow \emptyset$;

For *i = 0 to K:*
Create all possible Boolean function have i variables $f_j^{(i)}$*;*
$L_i \leftarrow L_i \cup f_j^{(i)}$*;*
$G \leftarrow G \cup L_i$*;*
3. Identify the network
For *all* x_i *in V* ***do****:*
$F_{...} \leftarrow \emptyset$*;*
For *all* $f_j^{(i)} \in L_i$ *in G*
If $f_j^{(i)}$ *compatible with all* $(X_i, Y_i) \in$ D: $F \leftarrow Result \cup f_j^{(i)}$*;*
If F_i *has 0 or more than 1 element:* $Result \leftarrow \emptyset$*;*
Return *Result*

Another approach of network inference is to use genetic algorithm to solve problem. The algorithm mimics the evolution process to find the most fitting population of network. The information of a Boolean equation system can be encoded to a chromosome as a vector structure as shown in figure 13.

- The first step of the algorithm is population initialization. In this step, initial Boolean equation is generated by encoding information into a chromosome vector, then a random list of vectors is created. The number of populations is denoted P.
- A fitness metric is created to evaluate the compatible level of a Boolean network generations with the data observed from laboratory experiments. Different metric can be used to measure the fitness.
- Basing on the measurement from the last step, the chromosome vectors are compared and the vector with better fitness will be selected and passed to the next generation.
- After new generations is collected, the crossover between chromosomes is taken place. In this step sequence at position i of a selected chromosome are exchange with counterpart sequence in another chromosome.
- Then, mutation is introduced to the chromosome by randomly altering the value of a random element in each chromosome vector.
- Now the new chromosome population is evaluated by the fitness metric created above. If the fitness value is converging to a specific point and the difference between the fitness of two adjacent generations $\Delta fitness$ are below a given threshold, the set of chromosome population is returned. Figure 13 show the visualization of a chromosome. Each equation in the system is represented as a vector containing information about start point, target node, regulation nodes and logical operators set.

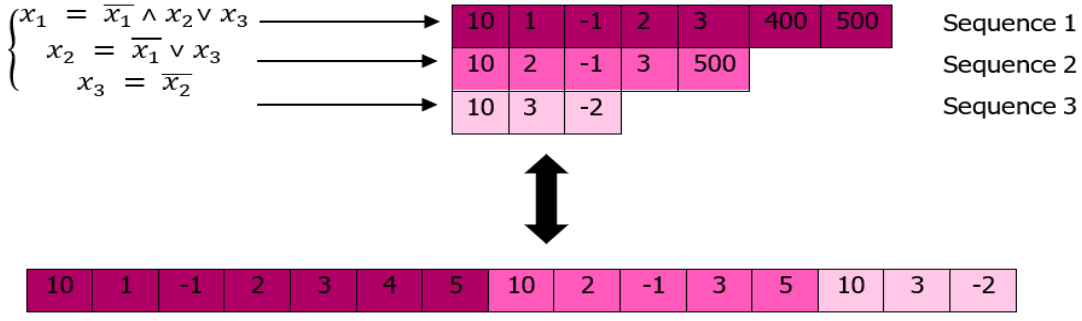


Figure 13: The encoding diagram for a Boolean equation system. (Shi, 2013)

The genetic algorithms can overcome the disadvantage of the conventional iteration algorithms about time complexity; however, it may not guarantee to give an accurate answer. The lack of a mathematical theory to solve problem of the system of operator equations is a difficulty, so it is very much in need of help from mathematical researchers.

3.3.5 Attractor detection problems

The definition of attractor is closed loop existing inside the Boolean network, once the system enters any nodes of this loop, it will stay there and cannot escape to any nodes outside the loop (unless there is internal perturbation or environmental influences). The problem of attractor identification is to find loop structure in a given network. Like the network inference problem, attractor identification problem has been proven to be NP-hard problem. Several methods have been proposed to solve the problem such as conventional recursive algorithms, SAT clauses reduction or Breadth-first search-based ordering algorithm, etc. In the thesis, the conventional recursive and circuit finding method for attractor detection are discussed.

- Conventional recursive algorithm: Given a set of Boolean function $F = \{f_1, f_2, \dots, f_n\}$, a set of nodes $V = \{v_1, v_2, \dots, v_n\}$ which give the vector state $X = \{x_1, x_2, \dots, x_n\}$ of the system. The strategy of recursive algorithms is to create a partial global state $X' = \{x_1, x_2, \dots, x_m\}$ with $m < n$, then partial global state is compared with partial global state $F' = \{f_1, f_2, \dots, f_m\}$ to detect any incompatibility between X' and F'. To start, the initial value is assigned to $m = 1$, and the initial state vector $X = \{\xi, \xi, \dots, \xi\}$ with n elements, the ξ symbol tells that the values of the elements in the vector X have not been determined at the time of initialization. Now, start iteration from gen v_1, first assigned the value $x_1 = 0$, three possible cases can happen. The first case is that the incompatibility between $x_1 = 0$ and f_1 is found, then the $x_1 = 0$ need to be changed to $x_1 = 1$ and continue to examine the partial global state with $m = 2$. The second case is that the compatibility exists between $x_1 = 0$ and f_1, then the value of x_1 is saved and continue to examine the partial global state with $m = 2$. Finally, if the compatibility between $x_1 = 0$ and f_1 cannot be known due to the value of all variable in f_1 are not determined yet (for example $f_1: x_1 = x_4 \wedge x_7$, but the value of x_4, x_7 are still ξ at the moment of iteration), then the value $x_1 = 0$ are temporary hold and algorithms continue to move to next step $m = 2$. Any incompatibility found in the following steps let to know that $x_1 = 0$ is not true and need to be switched to 1. Otherwise after increasing m to the value of $m < n$, and no incompatibility are found, then return the final state X and it is an attractor.

Algorithm 3: *Recursive algorithm for attractor detection*
Input: *N, F, m, V, X, p*
Output: *a set of attractors*
1. Variable initiation
$X \leftarrow \{\xi, \xi, \dots, \xi\}$;
$m \leftarrow 1$;
Result $\leftarrow \emptyset$;
2. Recursiveprocess(m,X,p)
If *m > n:*
If $f^k(X) \neq \forall k \in \{1, .., p-1\}$: *Result* $\leftarrow$ *Result* $\cup$ *X;*

```
    Return;
  For i = 0 to i = 1 do:
    x_m ← i;
    For j = 0 to j = m do:
      If f^p(X) [j] ≠ x_j : break;
      Else Recursiveprocess (m+1, X, p);
    Return;
```

- Circuit finding algorithm: an effective algorithm from the work of Donald B Johnson (1975) to detect circuits in graph are presented in algorithm 4.

```
Algorithm 4: Circuit finding algorithm
Input: N, G.
Output: a set of attractors
1. Variable declaration
    List of array A_k[N]; B[N] ← ∅
    Bool array blockedList[N] ← ∅;
    Int s; List tempStorage[ ]← ∅;
   List of array result ← ∅;
2. bool CIRCUIT (int v)
    bool f;
     UNBLOCK (int u)
        blockedList[u] ← false;
       for w ∈B[u] do:
          delete w from B[u];
          if blockedList[w] = true then UNBLOCK (int w);
    f ← false;
    tempStorage ← tempStorage ∪ v;
    blockedList[v] ← true;
    for w ∈ A_k[v] do:
       if w = s then
            result ← result ∪ tempStorage;
            f ← true;
       else if blockedList[w] = false then
            if CIRCUIT(w) = true then f ← true;
    if f = true then UNBLOCK (int v);
   else for w ∈ A_k[v] do:
       if v ∉B[w] then B ← B ∪ v;
   tempStorage ← ∅;
  CIRCUIT ← f ;
3. Iterate through the searching space to find all cycles
   tempStorage ← ∅;
   s ← 1;
   while s < N do:
      A_k ← adjacency matrix of strong component K with least
            vertex in subgraph of G composed from the set of
            vertex V_k = {s, s+1, …, n}
      if A_k ≠ ∅ then:
        s ←least vertex in V_K
        for i ∈V_K do:
            blockedList[i] ← false;
```

```
                B[i] ← ∅;
            updateResultVariable ← CIRCUIT(s);
            s+= 1;
        else s ← n;
    return result;
```

3.3.6 Causal relationship behind the model

The Boolean network model can also be used to represent the causal relationship not just only in the gene expression scenario, but also in reality events. From the evolution and patterns in state transition graph, the regulation graph can be rebuilt, in which the present of an event can lead to the "ON" or "OFF" state of other events. Therefore, the model can represent the cause-and-effect relationship between events in a system. Now suppose observations of a set of N events is recorded in time series, from that a set of vectors state $X = \{x_1, x_2, \dots, x_n\}$ at each point of time is constructed, where $x_i = 1$ means the event x_i happening at the time t, and $x_i = 0$ when it does not happen. After record a sufficient amount of data, a state transition graph can be rebuilt partially or completely. Now it is necessary to find a regulation graph that consistent with the observation, and then explore the causal effect relationships between the events from the observation.

The forerunner of this problem has been discussed by Yves CRAMA (CRAMA, 1988). In his study, a partially defined Boolean functions were used to discovery all possible functions which can create the same pattern with the observed data from experiment. The definition of this problem is closed to Boolean network, the difference between two problems is that Yves model helps to find causal relationships for a single event, in other word it solves one equation $f(x_1, x_2, \dots, x_n) \rightarrow y$, while the Boolean network model deals with a system of equations, which represents for multiple events happening in a closed system. There are two disadvantages if Boolean network are used to analysis the causal relationship of a system:

- In some special cases, it is possible to exist two separated subgraphs inside a mother graph, which have identical structure and number of nodes. In this case, if simple data of "ON, OFF" are used, it will cause ambiguous. Because two specific nodes can interchange their position between two subgraphs without changing the simulation result of the Boolean model. To differentiate between this coincidence correlation and truly causal-effect relationship, more information should be introduced into the model, for example introducing a variable to represent space-time interaction between two factors (x_i, x_j), because causal-effect event between two objects usually happens when there are some direct interactions happened at the same time and two objects often stay close enough in a space region. If there is a space-time connection between two factors (x_i, x_j), as well as their predict functions are consistent with the observation data, then there is high possibility they have causal-effect relationship. Another solution to examine truly causal-effect relationships is to generate all possible hypotheses Boolean networks and then use experiments to eliminate false hypotheses. Figure 14 shows an example demonstrates two networks which have six vertex and structure of two separated cycles. In a special case, when the updating process of the vertex in the network

coincidentally happens at the same time and takes a same duration to complete, the simulation result from observed data can be confused, because both network in the left and the right can produce identical result. In this case, experimental testing or introducing space-time information to the model is necessary.

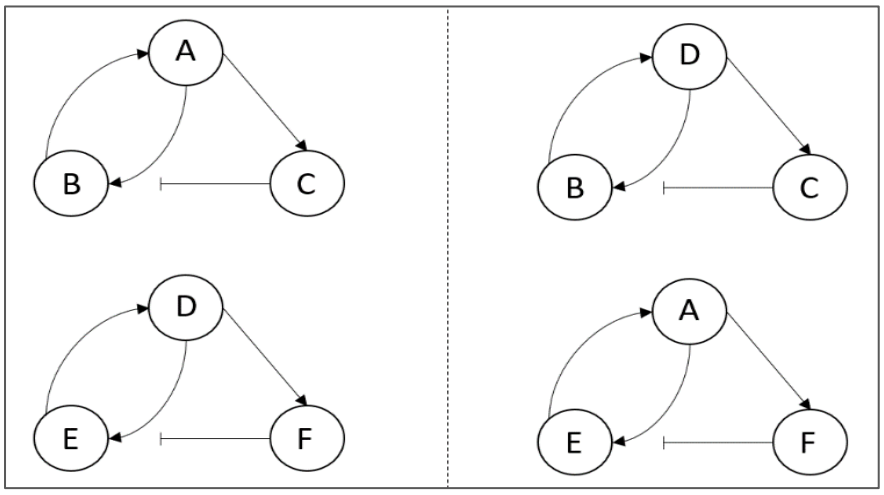


Figure 14: Error due to data coincidence when building network. (Lai, 2023)

- Using Boolean model to analyze causal-effect relationships inside a system usually requires large enough number of observations, and the size of state space increases exponentially with the number of components in the system. In fact, only a small part of whole system states can be recorded, hence the problem often becomes inferring a network from incomplete observations. To overcome the requirement of large observation, instead of taking the record for every node in a system, the system can be broken down to smaller clusters, and only necessary nodes of a single cluster is recorded. The reason to do that because functions relating to an organ/object tend to cluster together, each cluster can be considered as a meta node of the larger network. The connections inside a meta node are usually dense while the connections between meta nodes are usually sparse. Then it is easier to explore the information in a single meta node, then assemble it in the original network to find the sparse connections latter. Further examinations need to be done before conclusion.

3.3.7 An idea about model extension for further work.

The above presentations show the network analysis capabilities of the Boolean network models, but also reveal its shortcomings. The asynchronous model and probabilistic model are more consistent with reality observation than the synchronous model, but they cannot explain the driver of expressions. And reality shows that the operation of system does not depend only in their controlling code, but also in external factors, which the system responding with. From this perspective, by putting the gene expression network in its operating context, gene expression data can be modelized more accurately. In this scenario, the schema of the network could be described in figure 15. In the figure, the operational context may contain the following information: the presence of certain proteins, the concentration of solutes in the intracellular fluid, physical characteristics such as temperature surface tension, pH of intracellular fluid, etc. Operational contexts have the role of deciding which set of functions should be selected. Assuming the function set F1 is selected, the system will operate according to the rules of sub-network G1. The orange arrows in the graph represent the direct effect of

an operational context factor on a particular gene, such as the high temperature E2 that guarantee a gene in sub-network G1 to always be turn "ON". The blue arrow represents the feedback effect of gene expression layer to the intracellular environment, for example the activation or inhibition of certain genes to ensure homeostasis.

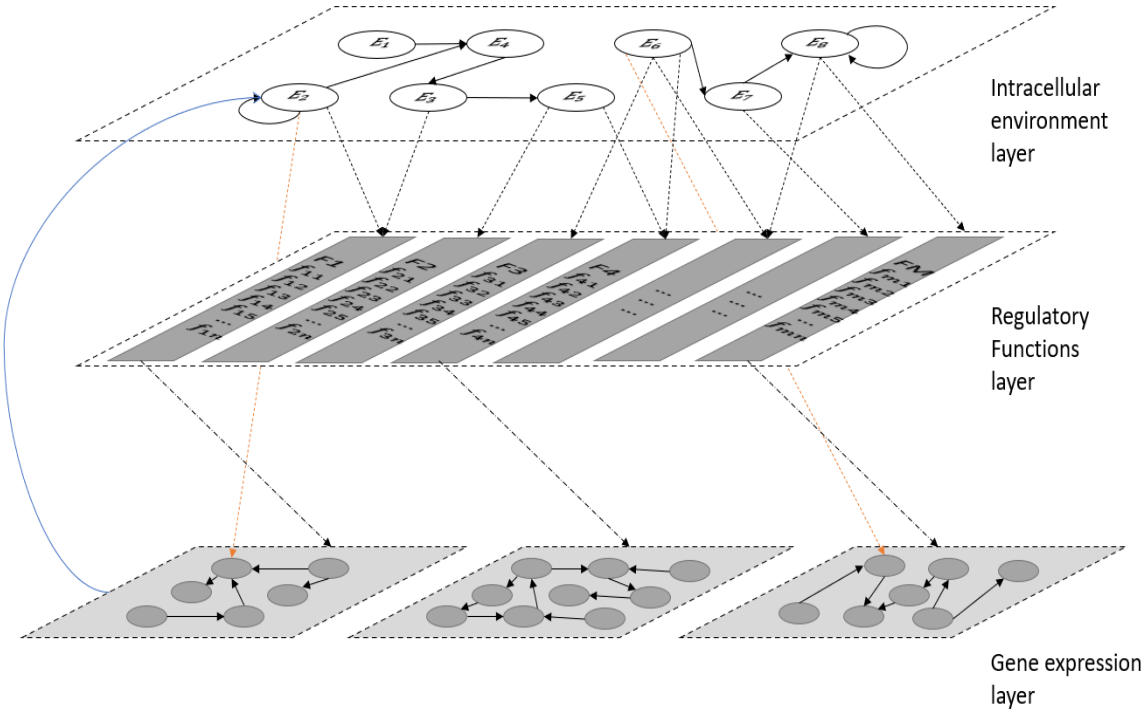


Figure 15: Schema of gene expression network with intracellular environment context. (Lai, 2023)

To conclude the first approach, analyzing gene expression data by Boolean network can be used to explore the arrangement of functional blocks embedded in the genome. This is the necessary information to support the understanding biological algorithms and then the of the operating mechanism of internal organs in the creature's body. The current shortcomings of this approach are that the current Boolean models still do not well describe biology processes and need to be adjusted in the future. The second problem is that current models usually ignore the effect of operating context on gene expression. To overcome these disadvantages, it requires the cooperations between people from different studying fields to adjust and build more accurate model.

4 NEURAL SIGNAL PROCESSING APPROACH

4.1 General introduction about the approach

From the discussion in Chapter 2, the phenomenon of paresthesia and the exist of electrical synapses suggests that cortical cells use patterns of neural signals to construct sensations. In vertebral animals, chemical synapses used in tandem with electrical synapses; thus, the type of neurotransmitter is also used by cortical cells in differentiating signal sources and constructing sensations. However, it is possible that the use of the patterns of neural transmission signal is still inherited from their ancestors. With this intuition, it is reasonable to think of a system, which record both initial neural signal sequence from sensory organs X(t,t+ξ) and signal from the cerebral cortex Y(t,t+ξ). After that if a mapping function Y = f(X) can be identified, and this function is relatively similar among individuals, then it is stronger evidence about common algorithms in the genome than the experiment in section 2. It may help to learn about sensory formation and then that knowledge can be used to enhance the abilities of current machines system. Thanks to the work of researcher and engineer generations, our technology is now advanced enough to make such devices. In this part of the thesis, author would like to represent an idea about implantable EEG device having structure like a stent (also called stentrode), which can be implanted at certain locations on arteries near nerves in human sensory organs such as eyes and skin. A long with it, the requirements, and challenges of designing such this system, and the technologies that can be used to overcome those difficulties also be discussed. Finally, a small experiment on an EEG signal acquisition system that works with the Arduino Mega is performed because it is feasible to implement and suitable for students' abilities. Critical comments and discussion are welcome to help finding the complete answer of this problem.

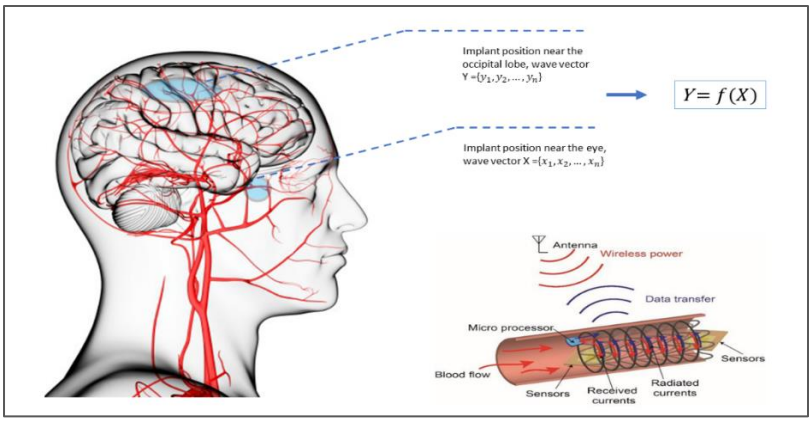


Figure 16: An example of stentrode implant. (Liu, 2019)

The figure 16 shows the implant position of the device is near to the ophthalmic veins, which have mean diameter of 2mm and near the position of optics nerve. The neural signal from eye cells could be recorded and transform into a vector $X = \{x_1, x_2, \ldots, x_n\}$. An EEG device is put near the position of occipital lobe, which receive and process signals coming from the eyes, the vector $Y = \{y_1, y_2, \ldots, y_n\}$ is obtained. Then data is analyzed to examine if there exists any function that map X to Y, and if it is similar between individuals. The stentrode is a minimally invasive method that avoids tissue damage. Due to its small size, it is not possible to deploy many signal reception channels, so for the applications which require high resolution records, it should be used in addition to other signal acquisition methods.

4.2 The structure of signal acquisition system

The Stentrode is a miniaturized data acquisition system, which is designed specifically for vascular implantation. Based on the existence of electric current in the cell, one can design devices to measure the potential of cell membranes and record the operations of the cells. All cells in an organism have a polar membrane, made up of a lipid bilayer and the proteins attached to it. When cells are in resting state, this membrane helps to maintain a stable voltage potential between the intracellular and extracellular fluids. When the cell goes to active states, the permeability of the cell membrane changes leading to instantaneous changes in electrical potentials, known as action potentials. The device consists of thin metal strips, under the influence of changes in the electromagnetic field on the cell surface, the induced current in the metal strip appears proportionally to the magnitude of the cell membrane potential. Because the amplitude of these current is usually tiny, it needs to be amplified, filtered noise signal, and finally convert from analog form to digital form so that computer can understand and extract the information contained in the signal.

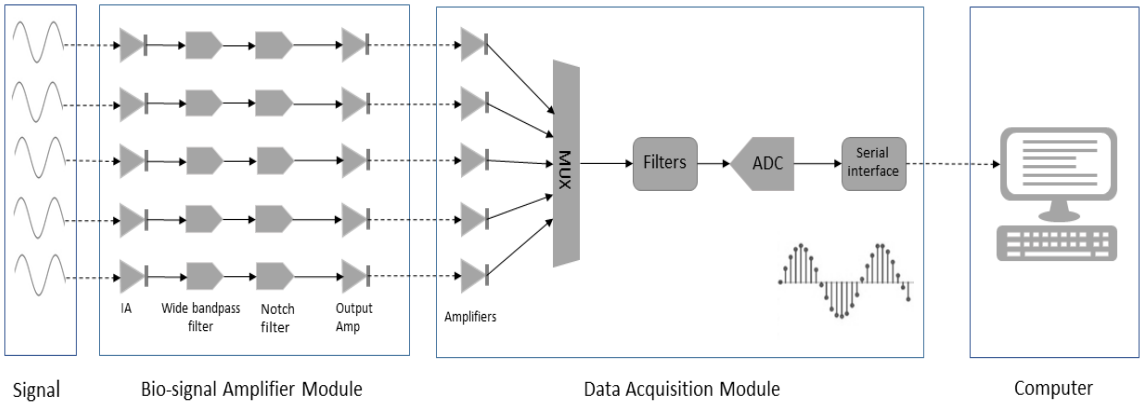


Figure 17: Schema of a signal acquisition system (Lai, 2023)

The acquisition system for signal from biology cells in figure 17 contains two main blocks, the first block is used to amplify the intensity of signal waves, the second block is a data acquisition module used to convert the analog signal to digital signal. A particular design of analog frontend in a biology data acquisition system contains four components: a differentiate amplifier, a wide bandpass filter, a notch filter and finally an output amplifier.

- Differentiate amplifier: the purpose to put a differentiate amplifier in this place is to suppress the common-mode noise. The common-mode noise come from two sources. First, when the current travel in the wire of an integrated circuit, it leads to the electromagnetic field variation, this phenomenon effects to metallics wire and components in the circuits by causing electromagnetic induction. Hence, some components especial the ground pin of the devices can easily pick up a few mV/V of noise signal. Second, the current leakage from another circuit into the ground wire of the mentioned circuit, which leads to ground potential noise. The output voltage of the differentiate amplifier is the product of the gain with the different between the voltage of inverting gate and non-inverting gate as described in the following equation: $V_{out} = A_d(V_{in}^{+} - V_{in}^{-})$. A popular schema of differentiate amplifier is WheatStone Bridge differentiate amplifier. By connecting one leg of the bridge to a reference voltage and another

to the brain signal measurement pins, the voltage output has a magnitude that is linearly correlated with the brainwave signal.

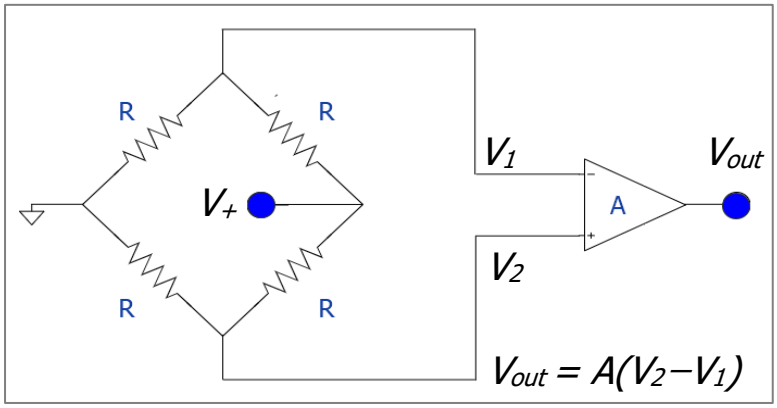


Figure 18: The schema of WheatStone Bridge differentiate amplifier. (Lai, 2023)

- Bandpass filters: After filtering out the common-mode noise, one or a series of band pass filters are used to eliminate the signal bandwidths which are outside of the signal range generated by the cells. The electroencephalogram signal has the frequency falling in the range of 0.5 – 150 kHz and the amplitude in the range of 5 – 300 μV, the electroneurogram signal has frequency range of 100Hz – 1kHz and the amplitude from 5 μV to 100mV. A bandpass filter can be composed by a high pass filter and a low pass filter connected in series. There are many methods to create such these filters: passive filter circuits using only passive components such as resistor, inductors and capacitors, or active filter used active components such as amplifiers.
- Notch filters: Notch filter is a special case of stop band filter, which removes a specific frequency or a band of frequencies. In this case, the frequencies from the power source need to be remove, which can interfere with the target signal. If a wireless power system is used, then frequency of the energy-carrying wavelength should be filter away. Notch filters can be designed with only passive components or with both passive and active components. There are many methods to design a notch filter, an example about a notch filter, which is a simple RLC circuit is presented below:

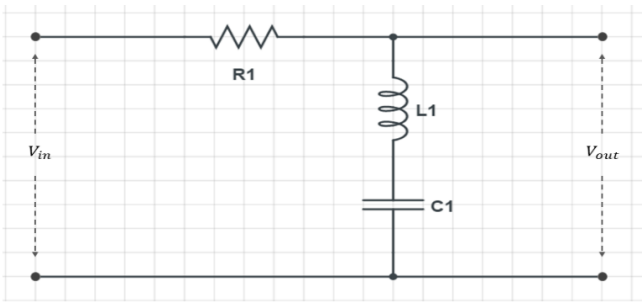


Figure 19: RLC circuit of a notch analog filter. (Lai,2023)

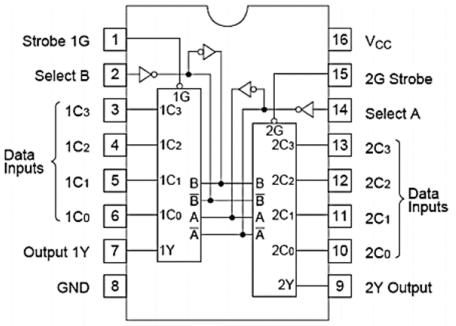
From the diagram above, output voltage is calculated by the equation $V_{out} = V_{in} \times \frac{(Z_{L1}-Z_{C1})^2}{\sqrt{R_1{}^2+(Z_{L1}-Z_{C1})^2}}$. To eliminate a specific frequency from signal, let the V_{out} at this frequency equal to 0, that mean $Z_{L1} = Z_{C1}$. From the formulas for impedance of inductors and capacitors, the equation becomes $\frac{1}{2\pi fC} = 2\pi fL \leftrightarrow f = \frac{1}{2\pi\sqrt{LC}}$. By choosing a suitable value of L and C, target frequency can be filtered out from the wave spectrum.

- Output amplifier: Finally, the signal is amplified again before being sent to the data acquisition module. None-inverting amplifier is used to do this task, because of its high input impedance, it can also act as an isolation between the bio-signal amplifier module and the data acquisition module. The gain of none-inverting amplifier is given by the equation $A_V = 1 + \frac{R_2}{R_1}$.

After the raw signal is collected and processed by analog frontend module, it is feed to data acquisition module. Digitization and interfacing take place to ensure that the computer can understand the signal. Three fundamental components of a data acquisition module are multiplexer, analog to digital converter and serial interface circuit.

- Signal conditioning unit: this block contains a combination of amplifiers and filters which help transform the signal into a suitable form for processing purposes.

  Multiplexer: Multiplexer circuit, also known as multiplexer, multiplexer (Multiplexer-MUX) is a type of combinational circuit that allows selecting between many parallel input lines (input channels) then forward signal to 1 output (output channel). The input channel selected is decided by the multiplexer's switch. The MUX acts as a multi-position switch controlled by a code. This code is a binary number. Depending on this combination of binary numbers, only one input is selected and allowed to be output at any time. Typically, the MUX can usually switch 2^n input channels to 1 output channel, for example 2 to 1, 4 to 1, 8 to 1, etc. The example below demonstrates structure and truth table of a 4 to 1 multiplexer 74LS153. Figure 20 shows the multiplexer 74LS153 structure. It is designed with eight data channels, and two multiplexer unit. When the G port get the value 0, it enables respective data selector gates. When the value of G port is 1, the circuit is closed then the output Y equals to 0, regardless the value of data input ports. When the value of G port is 0, the data port is selected by data selector gates A and B as described in the table. The output value Y then equal to the value of data port selected.



Strobe (Enable)	Selection Inputs		Data Inputs				Output
G	B	A	C0	C1	C2	C3	Y
1	-	-	-	-	-	-	0
0	0	0	0	-	-	-	0
0	0	0	1	-	-	-	1
0	0	1	-	0	-	-	0
0	0	1	-	1	-	-	1
0	1	0	-	-	0	-	0
0	1	0	-	-	1	-	1
0	1	1	-	-	-	0	0
0	1	1	-	-	-	1	1

Figure 20: Circuit schema and truth table of multiplexer 74LS153. (Fairchild, 1986)

An advantage of using multiplexer in data acquisition system is to reduce the cost and complexity. Each ADC unit in a signal acquisition system can only process information for one signal channel at a time. In case there are many signal channels to be processed, it will be expensive to use an ADC for each channel. By using multiplexer, then one ADC unit can handle several signal channels, hence it helps saving installation and operating costs.

- Analog to Digital converter: The electrical signals collected from sensor are usually in continuous form; in contrast the computer work with binary number system 0 and 1. To ensure the computer understand the information from sensor, data need to be digitalized. ADC unit is the devices which

handle this task. ADC continuously takes sample of original signal in certain sampling rate, convert the value of signal to binary form. There are many ways to do signal digitalization, fundamental design that can be listed such as: direct-conversion ADC, pipeline ADC, algorithmic ADC, integrating ADC, so on and so forth. The conventional direct-conversion ADC will be presented here. In the direct-conversion architecture, a set of 2^{n-1}comparators are connected in, one gate of each comparator is connected with the input voltage source, another gate is connected with reference voltage set by a series of resistors. This system can be used to convert N-bit resolution. The advantage of direct-conversion methods is fast speed because the signal can be processed parallel on multiple comparators, however due to the large number of resistors and comparators used in the design, it causes the problems such as bulky size and large power consumption when operating. Figure 21 show a design of flash ADC (Kutre, 2018). Notice that flash ADC is composed from repetitive blocks of components. The electromagnetic induction on the conductors can add noise or errors to the signal.

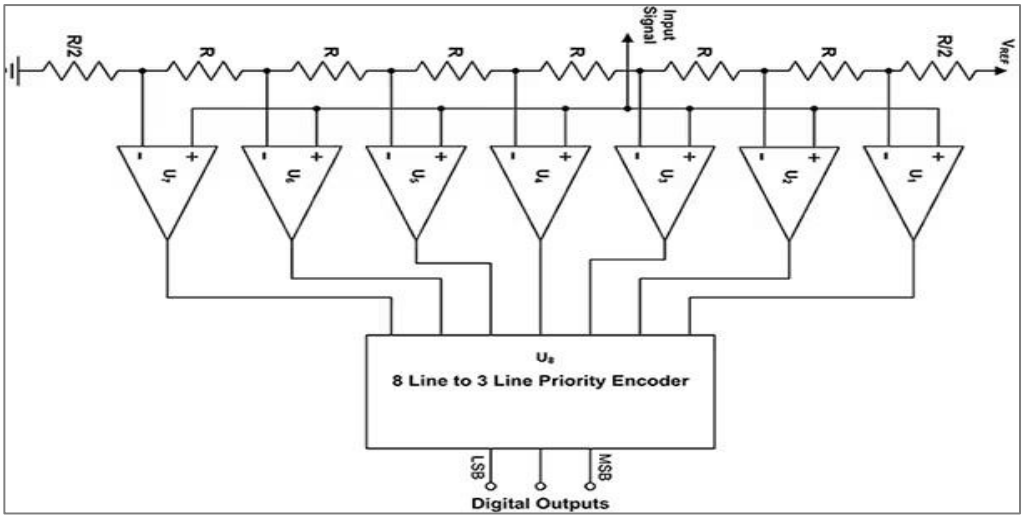


Figure 21: The schema of a conventional flash ADC (Kutre, 2018)

Another popular design of ADC converter is the pipeline architecture. This architecture the requires the participation of 3 components: the sample-and-hold circuit (H/S unit), a small flash ADC and a DAC. The signal processing process is usually divided into two smaller stages. In the first stage, analog signal is sampled and encoded into binary form by flash ADC and hold by H/S unit. Then recorded binary value is converted back to analog signal by DAC and be subtracted from the original analog signal. The residue after subtraction is amplified and used for the next stage. The advantages of pipeline architecture to flash architect are compact size, energy saving, higher resolution and less of noise. Along with it, its disadvantages are delay phenomenon and more complicated calculation process.

- Computer interface circuit: Depending on the type of ADC and the ending device, an interface circuit may be needed to convert the data from the ADC into a form suitable for the computer. The data after being converted need to be compatible with end devices on the following characteristics: using the same communication protocol (serial, parallel, I2C, etc.), suitable transfer rate, port, bus, and card compatibility.

4.3 Advances in micro-device manufacturing and stentrode recording array.

The problem of stent designing is a difficult problem, requiring a lot of knowledge, therefore the content of this section will present the problem based on the knowledge from the previous studies of scientists. Their achievements in creating microscopic electronics will facilitate successful stentrode fabrication.

4.3.1 Vascular stents

Vascular stents are a small device used for arterial stenosis treatment. With their compact size, vascular stents can be implanted deep in the body, in blood vessels, without the need for an open surgery. Therefore, it is minimally invasive, helping to avoid damaging surrounding tissues during treatment. Recent advances in materials and sensor research have allowed to integrate small sensors into stents, creating smart stents suitable for picking up signals from brain cells. The purpose of creating these stents is to bring them closer to the locations near the sensory organs, collecting signals from the desired organs. Then the signal from the cortex that controls that sense organ is also recorded and the relationship between the two signals is examined. The stent can act as a signal acquisition system, but the size needs to be small enough to fit the size of the blood vessel. To be safely implanted into blood vessels, vascular stents need to meet the following requirements: diameter of the stent ranges from 2 - 15mm depending on the implanting site, stable structure and minimum deformation after implantation, biocompatibility with blood vessel wall. There are three material category (Pan,2021) that is used to fabricate stents:

- Bare metal stent: stainless steel and NiTi alloy are most popular metals material used for stent manufacturing. The advantage of metal is high duration, high support strength for the vessels. Stability of stents after implantation depends on the designed structure. However, metal stents usually have low biocompatibility with blood vessels, which can lead to an inflammatory response, scarring of the vascular tissue, and stenosis. Therefore, metals or alloys should be carefully selected to avoid the above complications. In addition, the selected material should have corrosion resistance properties because of highly corrosive characteristics of environment inside the vessels.
- Drug Eluting Stent: The proliferative phenomenon that occurs after stent implantation is caused by irritation at the contact site between the stent and the vessel wall, causing thickening of the vessel wall. To solve this problem, the scientists developed stents that has a metal frame but is coated with a layer of material that is biocompatible with blood vessels (typically polymers) and contains anti-proliferative drugs. This solution ensures to retain the durability and vascular support properties of bare metal stents but helps to reduce the complications of vascular stenosis. After the layer of coated material and drug dissolves, the metal framework remains in the blood vessels and can cause stenosis in a later time. Materials commonly used to make frames include stainless steel, cobalt-chromium alloy; NiTi alloy, the anti-proliferative drugs include sirolimus, paclitaxel and everolimus, the polymers used as drug delivery vehicle are Poly L-lactic acid, Racemic polylactic acid, so on and so forth.
- Biodegradable materials: Biodegradable stent was developed to address the disadvantages of bare metal stent and drug eluting stent. They are mainly manufactured from polymers or

metal alloys, which have capable of self-decomposing in the body of an organism. Some polymers which can used for biodegradable stent including poly-L-lactic acid, polylactic acid, poly-glycolic acid, Racemic polylactic acid, the metal alloy can be used including the alloys of iron (Fe), Magnesium (Mg) and Zinc (Zn) with other metals such as Lithium (Li), Calcium (Ca), Manganese (Mn), etc. After being implanted into vessels, their structure remains stable for 3 to 6 months, then is gradually absorbed in the body and disappears completely after about 18 to 24 months. This characteristic helps to prevent inflammatory reaction, thrombosis. The strength of those materials is usually lower than the metal frame and some of them have fast duration attenuation. A problem when using the degradable alloys is that the metal ions released in the blood can be toxic with the surrounding tissue and body, hence metabolizable metals are usually used.

An important characteristic of the stent is its structure, this characteristic will determine the deformation of stents when they are implanted, and their stableness inside the blood vessels. There are many designs of stent, the most popular design is to use multiple individual rings connected by several bridges, or by weld between struts and crowns; another popular design is to use a mesh of metal. Mechanic properties of stents which are usually considered are radial stiffness, Young elastic modulus and tensile strength.

- Radical stiffness: this is the most important property. Because the stent surface will be subjected to pressure from the blood vessel wall after implantation, so its structure needs to have the suitable stiffness to avoid deformation. The radical stiffness is calculated by the ratio of the radical force to the radical deflection: $K_s = {}^{F_r}/_{u_r}$. Calculations are often performed by scientists using computer simulations combined with actual measurements.
- Elastic modulus: Beside the radial pressure exerted by the vessel wall, stents can stretch, compress, or bend during implantation, or under effects of blood vessels movements. Therefore, the elasticity of the stent is also an important property. A compatible elasticity will ensure a stable shape for the stent regardless contractions of blood vessels, but not too stiff to avoid damaging the vessel tissues.
- Tensile strength: is the limit of endurance of the material when being pulled, a force exceeding this limit will cause the object to break.

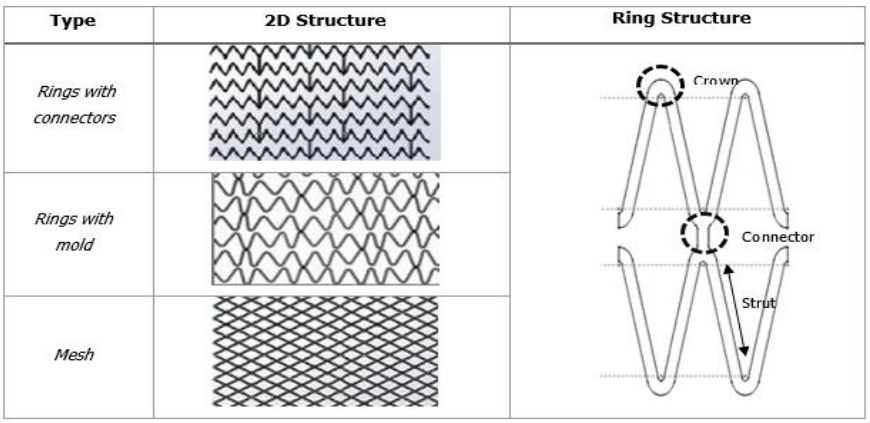


Figure 22: Left: popular structures of implantable stent; Right: ring structure with crown, connector, and strut. (Watson, 2017)

4.3.2 Micro-antenna integrated with vascular stent.

Recently, scientists have achieved some success in turning a stent into an antenna, using wireless signal transmission methods to connect to other devices. The purpose of these experiments is to monitor and detect restenosis in early stage. This technology can be also applied on the stentrode as well. The antenna is an open harmonic oscillator circuit, in an LC circuit, if the capacitors are deviated and kept away from the inductor, the electromagnetic waves can spread to the outer space. By integrating a LC circuit or designing a stent as an open circuit, it can be turned to an antenna. The three fundamental problems of creating an endovascular antenna stent are size, power problem and ability to transmit waves through tissues in the body. To be able to implant in the vessels near sensory organ of animals, the stent diameters need to fall in the range between 2 to 5mm (Chen,2022). This constraint not only limits the size of electronic components integrated with the stent, but also limits the transmit power of the stent. Besides, for the antenna stent signal to reach the external receiver, the broadcast power needs to be sufficient, thus requiring sufficient energy. The furthest distance which an antenna-stent system signal can penetrate through body tissues was observed in an experiment (Chow, 2009), with a device implanted 3.5cm below the chest and captured signal has a magnitude of 32-35dB at 10 cm. The researchers have proposed two main following solutions to solve the energy problem for the system. First solution is to use wireless powering technology such as Bluetooth or Radio Frequency Identification (RFDI) to transfer energy. An external coil is place near the implanting site, when the system turns on, electromagnetic waves are transmitted to the stent, causing electromagnetic induction, and generating an electric current to help the smart stent work. The second solution is to harvest energy from animal's body. The energy sources that can be harnessed in the blood include chemical energy, kinetic energy from blood flow, and electromagnetic energy. However, the energy output of these sources is low. Current efforts include creating devices with even smaller energy consumption. Some of solution has been proposed by researchers:

- Passive resonance LC antenna: Simple resonance circuits including a capacitor and an inductor can be turn to a simple antenna. Similarly, the stent can be manufactured to become and inductor, then it is connected parallel with a compact capacitor to form an antenna. The antenna can perform both functions including converting incident electromagnetic wave to electric current and transmitting signal. First, the signal from brain cells is converted to electric current in EEG probes, then it becomes a voltage source for the stent antenna. Since the current around the cell varies with time, the voltage from probe also varies over time, and causes oscillations in the LC circuit. These oscillations finally are converted to electromagnetic wave and transmitted.

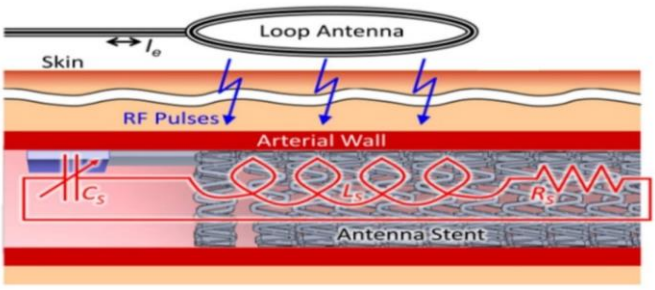


Figure 23: The schema of LC antenna integrated with stent. (Brox, 2015, copyright: © 2015 IEEE)

Some antenna parameters include frequencies and bandwidth, gain, impedance, directivity, etc. In the figure 23, the impedance of the antenna stent can be calculated by the equation $Z_{stent} = R_{stent} + \omega L e^{j\frac{\pi}{2}} + \frac{1}{\omega C} e^{j\left(-\frac{\pi}{2}\right)}$. The good electrical conductivity of the blood is also a factor which reduce resonance of the circuit and reduce the transmitter power. To avoid this phenomenon, the metal frame of the stent can be coated with a waterproof polymer.

- Miniaturized packaged antennas: In this approach, stead of using stent as inductors, a complete antenna package with antenna and radio frequency integrated circuit is integrated into stent. The package diameter can be reduced to 6 mm, and it is mounted at the edge of the stent by using delay locked loop design. However, this design is still cumbersome if compared to the first solutions and till use an external wire loop to transfer the power. Another problem which needs to be considered is the robustness of the components mounted in the stent. Attaching a chip to a stent has two problems: first, the chip can be damaged during the process of transferring the stent to the implant site; Second, the connection between the chip and the stent may rupture causing obstructions in the blood vessel.
- Active stent antenna: Unlike the LC circuit that uses only passive components, the active antenna stent circuit integrates both an integrated circuit (IC) and a sensor. The use of active components brings both benefits and challenges. In terms of benefits, the integrated circuit IC allows the system to perform additional functions such as filtering, digitizing, or amplifying the signal, thereby improving signal transmission quality, and reducing noise. On the downside, using multiple components increases power consumption, the components also need to be miniaturized to a reasonable size. ICs are now being miniaturized thanks to advances in lithography technology. In several recent studies, scientists have built ultrathin IC devices, printed onto polymers that are biocompatible with vascular tissue.

4.3.3 Electrode probes and manufacturing materials.

Electrode is one of the important parts of the neural signal acquisitions system. The non-invasive EEG brainwave measurement system is a signal acquisition system, which uses hemispherical, planar disk, or needle-shaped electrodes on the head area to pick up nerve signals. The commonly used material to make this type of electrode is silver (Ag) and can be coated with silver chloride (AgCl) salt, because of its high conductivity, which makes silver sensitive to receiving electromagnetic waves and high stability. The electrodes in this scenario can be active or passive device and can hold components as there is no constraints of size. As for implantable electrodes, they must meet more stringent requirements. Small size and stretchability are the first requirements. Current technology has made it possible to print electrodes onto small packages, which can be applied to the fabrication of smart stents. There are three characteristics of an electrode's materials (Yang, 2020) should be considered:

- Electrical property: Impedance is an importance property of electrodes. In printable electrodes, the metal strips are printed on a polymer bases and isolated by these layers, hence it forms a micro capacitor inside the electrodes. Hence the impedance of electrodes is the sum of metal lines resistance and inductive reactance of the capacitors. The formular to calculate

the resistance is $R = \frac{\rho L}{S}$ and the reactance is $Z = \frac{9\times10^9\times4\pi d}{\omega\varepsilon S}$, when the size of electrode decreases, its resistance will increase and add noise to the recorded signal. Electrochemical performance and conductivity are also criterions for material selection. Materials with good electrochemical performance are sensitive to signals from the cells.

- Compatibility: because the device is implanted inside the body, it can interact with surrounding tissue and can be possible to cause damages. Therefore, the implantable devices need to exist in harmony with tissue and cause minimum disturbance to living organs. The electrodes should be made a non-toxic material and do not trigger immunological response such as allergy or inflammatory response. Good flexibility is also a desirable property, components made of metal or silicon often have great stiffness, which can tear and damage tissue, causing scar tissue. The use of materials with the compatible stiffness and elasticity will help avoid these complications.
- Durability and mechanical property: The electrode work in an environment containing body fluids with many corrosive ions, if designed for long-term use, it should have reasonable corrosion resistance and durability. The electrode array implanted in brain tissue usually requires materials with low Young modulus, in the range of 3.5 - 10 kPa. To meet this requirement, ultra-thin metal strips are printed and packaged between thin polymer materials such as perylene-C, polyamide, or SU-8. In the scenario of stentrode, if an electrode array is selected to integrate with the stent, these materials and design can be used.

Today's flexible electrodes are mainly composed of three main components: the electrode material, the substrate material, and the package agent (Yang, 2020). The materials which have suitable and properties and can be used to fabricate the electrode are:

- Metals: inert metals are often used to make electrodes because of its good conductivity. Due to the high ionization energies, they tend not to reaction with other substances, therefore, it has high durability and is not toxic to cells. Commonly used metals include gold (Au), silver (Ag), platinum (Pt), iridium (Ir), tungsten (W). The disadvantage of metal materials is that the young modulus is much higher than that of tissues in the body, downsizing the electrodes will cause the resistance to increase. Attempts to address this include three dimensional (3D) nanostructured Au microelectrodes (Fairfield, 2018), in some experiments, one can use this method to reduce the size of the electrode to a diameter of 5 μm; or using alternative materials such as black Pt (Zhang et al., 2015), so on and so forth.
- Carbon-based materials: New carbon-based materials with good electrical conductivity and good mechanical performance including graphene, carbon fibers, carbon nanotubes can be used to fabricate electrodes. In general, a desirable feature of this group of materials is their high biocompatibility, flexible and mechanically strong. However, similar to metallic materials, they still have the problem of high values of Young modulus, which can eventually cause a rejection reaction.
- Conductive Polymers: Polymers with conjugated backbone (–C=C–C=C–C=C–) have properties that differ from conventional polymers in the ability to conduct electricity. A long the back bond, the sigma bond (σ) has stronger attraction because it concentrates electrons directly

between them; while the pi bond (π) has weaker and less strongly localization because it attracts electrons distributed above or below the molecule's plane. This process distorts the electron orbitals and allowing them to move freely between conducting band and valence band. However, electron mobility in pure polymers is low, mobility only increases when there is the presence of addition ions, also known as doping process. Many conductive polymers have been produced using this phenomenon, of which polymers that can be used as electrodes include PEDOT:PSS, PPy, poly(aniline) (PANi), poly(thiophene), etc. The weakness of this material is that it is less conductive than metals and graphene (Weiyang,2021). Among of them, PEDOT:PSS has good conductivity and flexible enough to not be rejected by tissue, suitable for electrode fabrication.

- Metal liquid: Gallium and its alloys are gaining more attention because of their metallic properties, but they melt at room temperature. In contrast with other liquid metals like cesium or mercury, gallium, has a high boiling point and is therefore not volatile, low toxicity and biosafety. This special property allows gallium usage in fabrication of implantable and wearable conformal electronics device. However, gallium is quite rare and expensive, the use of alloys of gallium with tin or indium helps to reduce the cost and the melting point of the alloy. The electrical conductivity of gallium is higher than that of organic polymers but slightly less than silver, low impedance. Also, its rigidity is compatible with tissues in the body. Table 1 shows a list of materials which can used to manufacture the EEG electrodes.

Table 1: The properties of several stent materials

Materials	Conductivity (S/m)	Young Modulus (GPa)
Gold	4.25×10^7	79
Platinum	9.43×10^6	172
Gallium	1.8×10^6	0.008
Graphene	104 – 105	2400
Polyaniline (PANi)	500	2 – 4
PEDOT:PSS	4.6×10^5	0.8 – 2,4

Besides the issue of material selection, the structure of the electrodes is also of interest to researchers. Depending on the implant purpose, the electrodes could have different shapes and sizes. For stents, there are two options to integrate with electrodes. The first is to design a small stretchable electrode array and attach it to the stent, the second is to turn the stent itself to an electrode array. If stretchable electrodes are used, some of the structures that have been for implanting purpose are linear strip, ultra-thin planar and mesh. Linear Strips is used when the electrodes are penetrated deep into the nerve tissues. The size of a linear strip electrodes has now been reduced to less than the distance of the brain microvasculature to the nerve cells (approximately 15 μm) and it can be integrated with the stents. Ultra-thin plana is designed to be attached on the surface of brain tissue, it is also suitable for grafting around a stent. The plana electrode are made by printing the electrodes onto the surface of the polymer layers. The size of a single probe on the planar electrode can now be reduced to 4 μm. Hence many probes can be printed on a small area, improving the resolution of the electrode. Finally, the mesh structure is created to be injected deep into the brain tissue or place on the cortex surface. For the usage on brain surface, the device is equipped with a matrix of tiny needles. For injectable purpose, the

device has a soft mesh structure. Due to the ability to penetrate deep into the structures of the brain, so it has high spatial resolutions and could record signals from different layers and regions of the brain. The main layouts of mesh structure are finger mesh, 3D mesh and Kirigari mesh (Zhou, 2021).

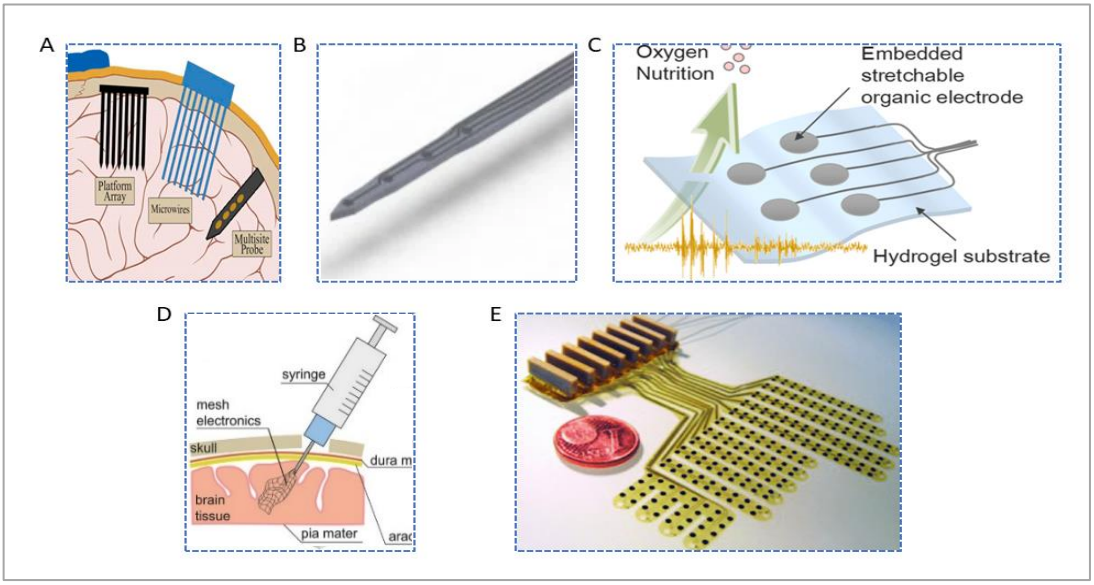


Figure 24: The structure of electrodes.

Figure 24 demonstrate different structure of implantable electrodes. Picture A is an example about linear strip structure, including three layers: platform array, microwires and multisite probe (Weltman,2016). Picture B shows a Michigan style probes design (Weltman,2016). The ultra-thin planar structure with array of electrodes packaged between the layers of polymers is presented in picture C (Oribe, 2019). Picture D shows an injectable electrode having a mesh structure (McGlynn, 2021)- Finally,a mesh structure with bulky IC electrode is represented in picture E (Ershad, 2018, copyright: © 2018 IEEE).

4.3.4 Implantable integrated circuit

Much effort has been made in designing integrated circuits (ICs) small enough to be implanted in the body. Depending on the construction of the electrodes used, the ICs will be differently designed to interface the signal received from this electrodes array into a form of information that can be used by the computer. Integrated circuit systems can have filters if analog filtering is used. If digital filtering is used, the integration into the circuit can be reduced, further reducing the size of the IC. However, this can increase the time and computational complexity of the decoder. In this section, the aim is to present the general schema of the neural signal receiver circuitry and describe in more detail a simple signal amplifier circuits and filter components that can be used in the IC of EEF system. In general, a neural signal acquisition circuit consists of 4 main components: neural electrodes array, readout architecture, digital signal processing block (optional) and antenna. The changes in the cell voltage induce a small current or voltage in probes of electrodes array and this current is transmitted to readout architecture. The signal can be filtered (optional) and then converted to digital form (compulsory). After that, it can it is passed through a digital processing block for further filtering, noise processing, etc. Finally, it is interfaced and packaged depending on the selected communication protocol in communication module and then transmitted by antenna.

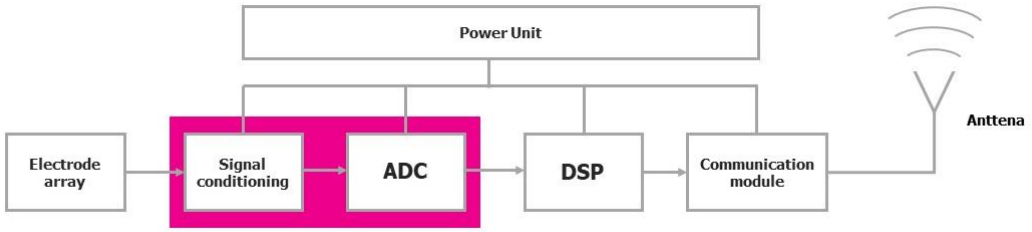


Figure 25: Schema of brain signal acquisition system. (Dasari,2013)

In the signal conditioning block, one can use an amplifier to amplify the signal, then the signal will pass through the series of filters to get rid of frequencies outside the frequency range of biological waves. A simple circuit of amplifier and a band pass filter are chosen to analyze and practice.

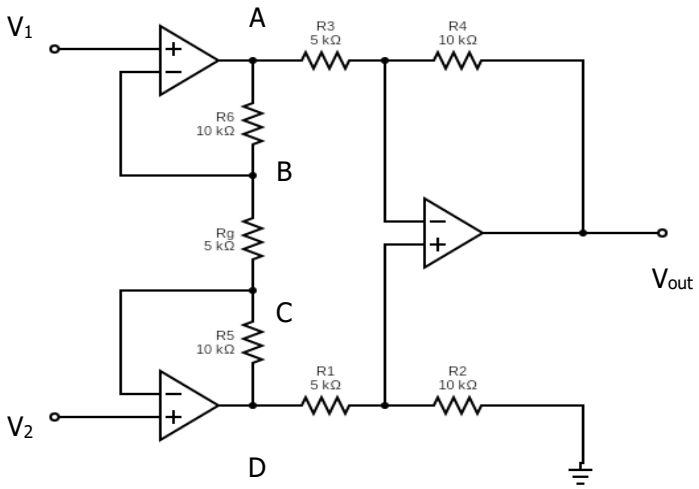


Diagram 1: Instrumentation Amplifier for signal from electrodes (Saptono, 2016)

The figure demonstrates an instrumentation amplifier, which is composed of one differential amplifier and two buffering algorithm amplifiers. The input V_1 and V_2 is put through algorithm amplifiers IC1 and IC2, then the outputs of IC1 and IC2 become the input of differential amplifier IC3. The instrumentation amplifier provides a high input impedance so that the voltages do not drop significantly because of divider effect. In addition, the high common mode rejection ratio and gain of this circuit make it good option to collect signal from sensors. The output voltage of circuit can be calculated as the equation below:

$$V_{out} = (V_1 - V_2)\frac{R_2}{R_1}(1 + \frac{2 \times R_5}{R_G}) \quad (3)$$

$$A = \frac{R_2}{R_1}(1 + \frac{2 \times R_5}{R_G}) \quad (4)$$

When $R_1 = R_2 = R_3 = R_4 = R_5 = R_6 = R$, the equation to calculate the gain of the circuit is:

$$A = 1 + \frac{2R}{R_G} \quad (5)$$

The frequency range of brain waves is in the range of 0.5Hz to 100Hz, signals outside this range can be filtered out. Active or passive band-pass filter circuit can be selected for circuit design. The active band-pass filter circuit uses the same principle as the passive band-pass filter circuit, the difference is

that the output signal will be amplified by an algorithmic amplifier. There are many ways to design an active bandpass filter, this circuit below demonstrates a bandpass filter which is made up of a low pass filter circuit and a high pass filter circuit.

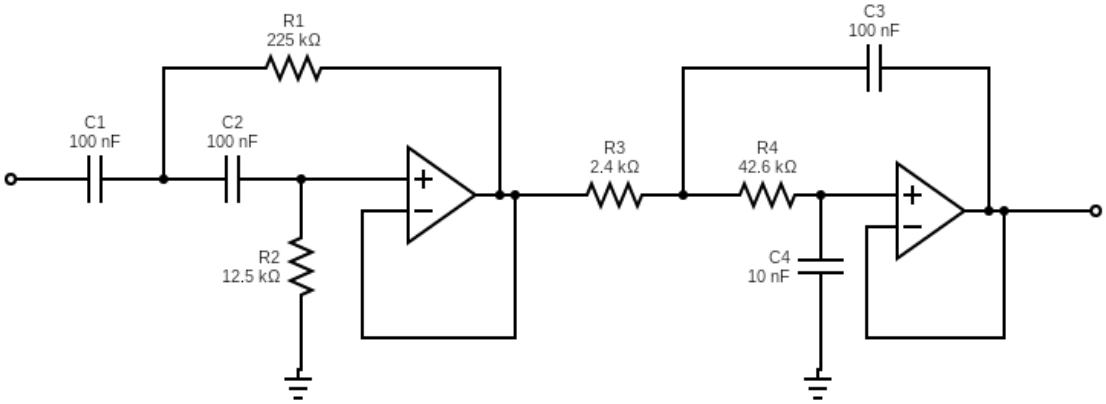


Diagram 2: A band-pass filter for EEG signal (Lai, 2023)

The first block of the circuit is a Sallen-key high pass filter circuit. The second block is a Sallen-key low pass filter circuit. The cut off frequencies of each circuit are calculated by the equation below:

- For the high pass filter:

$$\frac{V_{out}}{V_{in}} = \frac{{}^{K}/_{R_1C_1R_2C_2}}{s^2 + s\frac{R_1C_2 + R_2C_2 + R_1C_1(1 - K)}{R_1C_1R_2C_2} + \frac{1}{R_1C_1R_2C_2}} \quad (6)$$

$$fc_1 = \frac{1}{2\pi\sqrt{C_1C_2R_1R_2}}$$

- For the low pass filter:

$$\frac{V_{out}}{V_{in}} = \frac{K}{s^2(R_3C_3R_4C_4) + s(R_3C_4 + R_4C_3 + R_3C_3(1 - K)) + \frac{1}{R_3C_3R_4C_4}} \quad (7)$$

$$fc_2 = \frac{1}{2\pi\sqrt{C_3C_4R_3R_4}} \quad (8)$$

The two important concerns with implantable ICs are size and power consumption. Attempts to miniaturize IC in recent years have enable to produce integrated circuits small enough to be implanted. The size of implantable can be shrunk to 1 mm^3 with the power consumption 39.4 mW (Sutardja, 2017). Another IC with a smaller power consumption of only 22 μW, using 1V and recording 32 channels is also developed (Zou, 2010). The miniaturization of devices has made it possible to implant devices in locations close to the receptors. Signals from the sensory organs can thus be observed and further analyzed for their relationship with the activity of the associated cortical regions.

4.3.5 Conclusion about implantable device

The advances outlined in previous sections show that the current technology has enable to build a small enough neural signal acquisition system, which can be integrated in a vascular stent. The device can be connected by wire or wirelessly. These devices can capture neural signals with a high degree of spatial resolution, and they can be implant in the desirable locations. Placing the devices near sensory nerves or receptors will minimizes signal overlap phenomenon from other tissues and organs with neural signal. The received signals then could be analyzed using models such as support vector

machines or artificial neural networks to derive the necessary information. In fact, these devices have been successfully developed and are currently used in the treatment of epilepsy and some other neurological diseases, also known as stentrode (Oxley, 2020). It could be modified for the purpose of deciphering neural algorithms.

4.4 Experiment on signal acquisition system

This part of the thesis will present an experiment on a single-channel signal acquisition circuit. The signal acquisition circuit will consist of three main circuit blocks: an instrumental amplifier circuit, a bandpass filter circuit to filter out signals outside the frequency range of brainwaves and a circuit that removes the 50Hz signal from the source. Main used devices include Arduino Mega 2560, general purposed amplifier UA741, resistors and capacitors from Savonia university electronics laboratory. The choice of circuit that includes an active element is because the signal strength of the EEG or EMG waves is usually quite weak, their amplitudes are usually between 0.5 and 100 V, so these signals need to be amplified several times thousand times before being recorded. This also brings trade-offs in terms of size and power consumption. Hardware requirement for constructing the circuit is listed in the table 2.

Table 2: List of components used in the experiment.

No	Components	Quantity	Description
1	Operator Amplifier LM741C	7	The general purposed amplifier with high gain.
2	Arduino Mega 2560	1	The integrated circuit uses ATmega2560. In this experiment, the Arduino Mega is used to convert analog signal to digital signal.
3	Resistors 5KΩ	3	Used to build instrumental amplifier
4	Resistors 10KΩ	5	Used to build instrumental amplifier
5	Resistors 82KΩ	1	Used to build T-notch filter
6	Resistors 330KΩ	2	Used to build T-notch filter
7	Resistors 150KΩ	1	Used to build T-notch filter
8	Resistors 220KΩ	2	Used to build band-pass filter
9	Resistors 110KΩ	1	Used to build band-pass filter
10	Resistors 47KΩ	1	Used to build band-pass filter
11	Capacitor 10nF	3	Used to build T-notch filter
12	Capacitor 20nF	1	Used to build T-notch filter
13	Capacitor 100nF	3	Used to build band-pass filter
14	Electric board	1	AW board model C50
15	Wires	25	male to male, female to female

Issues to consider when choosing components for EEG circuits include sampling rate, frequency response, input range, impedance, and common mode rejection ratio.

- Sampling rate: The frequency range of the EEG wave is in the range of 0.5 - 80 Hz, based on the Nyquist sampling theorem the sampling frequency of the device needs to be at least twice as large as 80Hz. Thus, 160Hz is the minimum frequency that needs to be achieved to record the information carried in the wave in its entirety. The ATMEGA 328 analog to digital converter operates at sampling rate of 15kHz and meets the requirement.

- Frequency response: When choosing an Op Amp for the circuit, note the following parameters: bandwidth, resolution, input range. Since the Op Amp's internal components consist of capacitors connected to resistors and transistors, high frequency signals will be attenuated when passing through the Op Amp circuit. The cutoff frequency of LM741 at 3dB is 1Mhz, so the entire neural waveband is within this range. Another phenomenon that occurs on an amplifier is saturation. If the signal strength is above the saturation point of the Op Amp, it cannot be recorded correctly. When choosing Op Amp for EEG circuit, it is necessary to choose amplifier with saturation point above 50 mμV.
- Impedance: The EEG signal is usually weak and to avoid losing signal, a high impedance amplifier should be selected. The reason of this selection is because of the high impedance between the contact of electrode and skins, then high input impedance plays a role in conserving the transferred power from the original signal and protect the signal against noises. With the electrode resistant range from 1kΩ to 1MΩ, the impedance of amplifier should be at least 100MΩ.
- Common mode rejection ratio: From the above section, a differential amplifier should have a high CMRR to eliminate the common mode voltage. The EEG circuit require the CMRR at least 80dB at 50/60Hz.

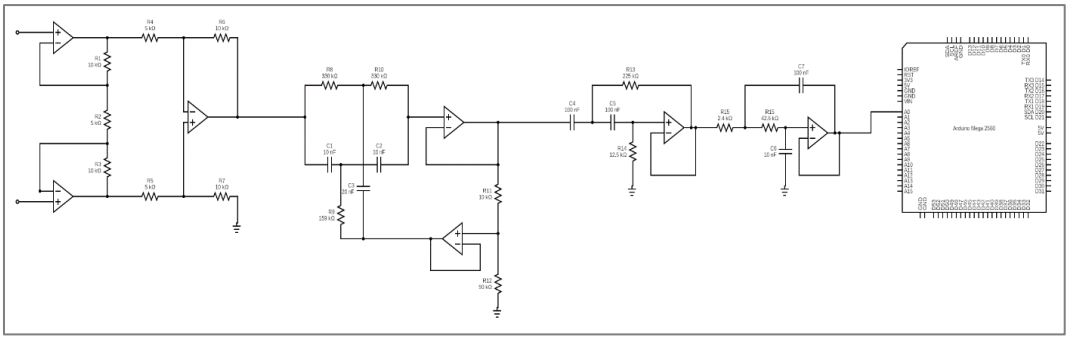


Figure 26 A: Diagram of the circuit. (Lai, 2023)

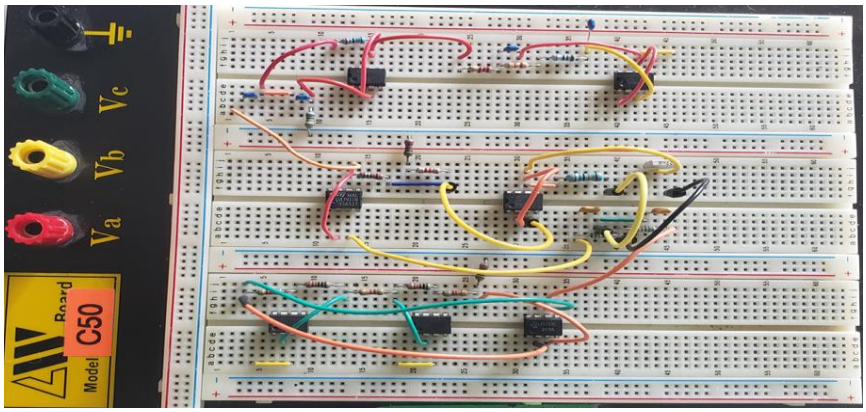


Figure 26 B: Assembled circuit. (Lai, 2023)

The first block of the circuit is instrumental circuit which have the diagram described in the diagram 1. The requirement for instrumentation Amplifier circuit in this scenario is low DC offset, high open loop gain, high CMRR, high input impedance and low noise. Given the resistors value $R_1 = R_G = 5$ and $R_2 = R_5 = R_6 = 10$, then the gain of the circuit can be calculated as below:

Consider the current I flow from the node A to node D:

$$I_{AD} = \frac{V_{out1} - V_{out2}}{R_5 + R_6 + R_g} = \frac{V_{out1} - V_{out2}}{2 \times R_5 + R_g} \quad (9)$$

Because of the high impedance of the amplifiers, it can be considered that there is no current follow between the input pin of the op amp, therefore a voltage potential appears across the non-inverting pins of the op-amp. Therefore, the current through the resistor R_s can be calculated as:

$$I_s = \frac{V_1 - V_2}{R_g} \quad (10)$$

Since $I_{AD} = I_s$, the following equation can be derived:

$$\frac{V_{out1} - V_{out2}}{2 \times R_5 + R_g} = \frac{V_1 - V_2}{R_g} \quad (11)$$

The two output signals of amplifiers 1 and 2 become the inputs of the 3rd amplifier and are amplified according to the ratio R_2/R_1, so the output gain of the system then should be:

$$A = \frac{R_2}{R_1}\left(1 + \frac{2 \times R_5}{R_G}\right) = \frac{10}{5}\left(1 + \frac{2 \times 10}{5}\right) = 10 \quad (12)$$

Thus, the EEG signal obtained after subtracting the noise will be magnified by 10 times. The second block of the circuit is an active twin T notch filter, with the center filtered frequency is 50Hz. Passive twin T notch filter is a circuit consisting of two simpler sub-circuits - a high pass filter circuit and a low pass filter circuit - connected in parallel. This is a phase shifting circuit, which is take the input signal, change the phase of the signal in each sub-circuit and then add the signal together. As the phase difference between high pass filter and low pass filter is 180°, it eliminates a specific frequency and provide null output at this frequency. The passive twin T notch filter has a disadvantage that the signal below the cut off frequency always have amplitude smaller than the signal above the cut off frequency and quality factor cannot be changed. To overcome this problem, active components are introduced to the circuit. In EEG experiment, the goal of this filter block is to filter away the 50Hz frequency, using k = 90% and the bandwidth is 20Hz. The notch frequency can be calculated by the equation:

$$f_n = \frac{1}{2 \times \pi \times \sqrt{R_8 \times R_9 \times C_{8\times}C_9}} = \frac{1}{2 \times 3.14 \times \sqrt{318300 \times 159200 \times 10 \times 10^{-9} \times 20 \times 10^{-9}}} \approx 50000\ (Hz) \quad (13)$$

Considered the quality factor Q and feedback fraction k, its formular is:

$$Q = \frac{f_n}{bandwidth} = \frac{1}{4 \times (1 - k)} \quad (14)$$

As $k = R_{12}/(R_{11} + R_{12}) = 0.9$, the bandwidth of the filter is estimated equal to 20Hz. Finally, the ratio between the input and the output signal is calculate by the equation:

$$\frac{V_{out}}{V_{in}} = \frac{s^2 + \frac{1}{R_9 C_3}}{s^2 + \frac{1}{R_9 C_3} 4\left(1 - \frac{R_{12}}{R_{11} + R_{12}}\right)s + \frac{1}{R_9 C_3}} \quad (15)$$

The third block of a circuit is bandpass filter described in the diagram 2. The analytical formulas for this type of circuit have been presented in Section 4.3.4. In this experiment, the purpose of this filter design is to eliminate the frequency outside of the range 20Hz to 500Hz, it helps to obtain the neural signal from nerves structures surrounded by muscle tissue.

4.5 Result and analysis

4.5.1 Circuit simulation by using LTspice

The component circuits of the EEG circuit were simulated individually using LTspice software to test their own frequency response. Then all the circuits are connected to check the final output. The results of the simulation program are presented in figure 27.

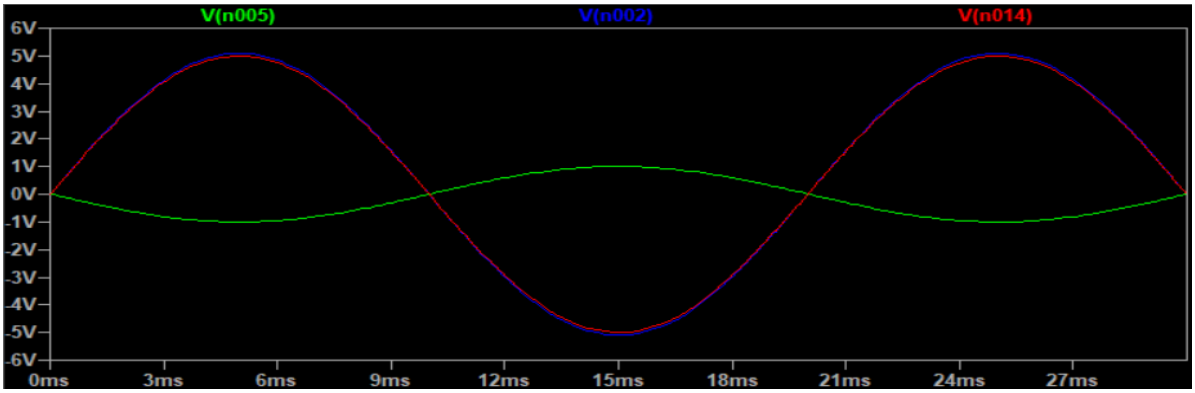


Figure 27: The simulation result of instrumentation amplifier. (Lai, 2023)

In the setup, the instrumentation circuit is connected to two AC sources with voltages of 5.1V and 5V respectively. The difference between two voltage sources is 0.1V. The output signal of the circuit has a value of 1V. This result is consistent with the calculations in the previous section, which is that the gain value of the circuit is 10 times larger than the difference between the two input sources. The phase of the output signal is shifted 180 degrees relative to the input signal. This is because at the peak and trough position of the two input waves, the potential difference between them is the largest, while at position 0 the difference is 0.

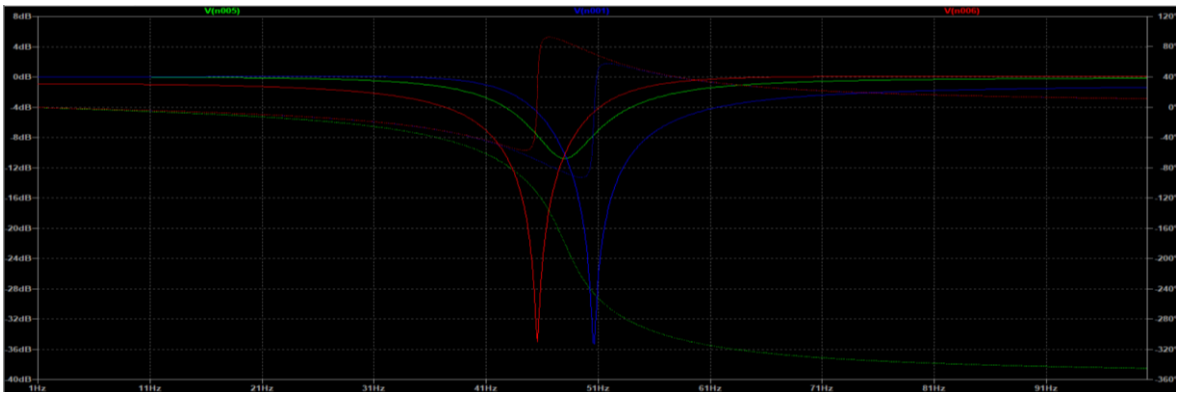


Figure 28: The frequency response of twin-T notch filter. (Lai,2023)

In the figure 28, the red line represents the response at the low pass filter branch, blue line represents the response at the high pass filter and the green line for the output signal. The phase shifting phenomenon between high pass and low pass branch can be observed, which lead to the cancellation

point of the output signal. The analysis result of twin T-notch filter shows its center frequency is 48.03Hz with the magnitude is decreased by 10.78 dB and the phase is shifted -180.3°. The filter in the experiment does not filter exact the frequency 50 Hz since there are no resistors and capacitors of the same value as in the calculation, the resistors and capacitors with approximate values are selected for assembly. 3dB level occurs at two points 41.4 Hz and 55.75 Hz.

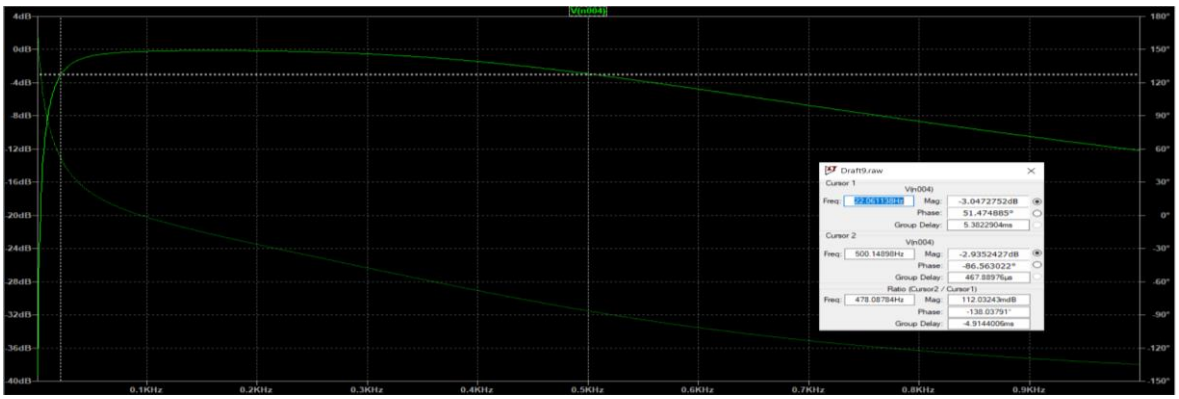


Figure 29: The frequency response of bandpass filter using Sallen-Key topology. (Lai, 2023)

The meets the requirement to remove all the wavelengths out of the range of 20Hz - 500Hz. The 3dB points occur at frequencies 21.7 Hz and 501.2 Hz. The DC gain of the filter is 2.515.

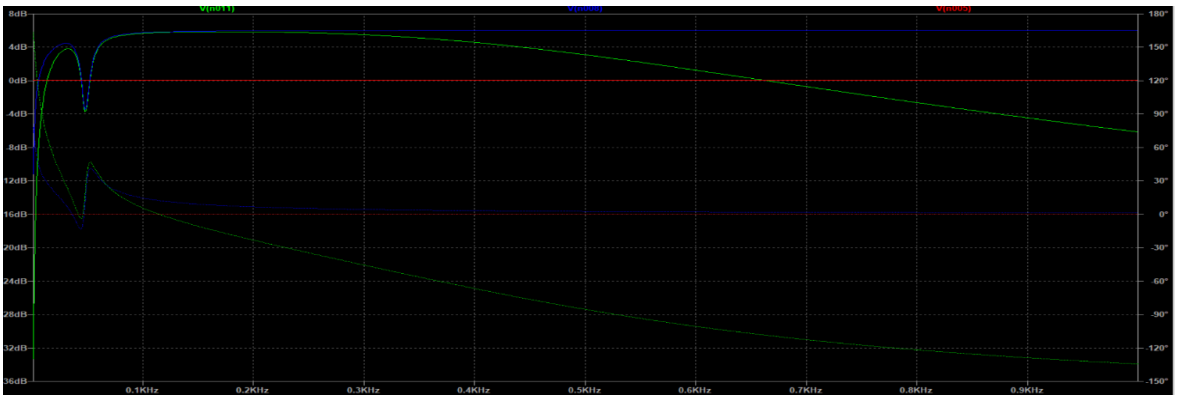


Figure 30: The frequency response of EEG filter. (Lai, 2023)

The frequency response analysis model of the whole circuit shows the combination of the above results. It is worth noting that the 50Hz frequency has not been completely filtered out, this is due to the lack of suitable resistors, as well as the resistance of the capacitors and amplifiers that have not been included in the calculation model. This will affect the actual measurement results.

4.5.2 Circuit assembly and spectrum analysis with Arduino Mega

The EEG signal receiver circuit after being assembled will be connected to the Arduino Mega 2560 via analog pin A0. Fast Fourier transform FFT is efficient algorithms to perform discrete Fourier transform (DFT), this is a particularly useful algorithm in case the ATM2560 microprocessor because of small RAM size and moderate computational speed. Program used third party library which implements the FFT algorithm invented by Cooley Tukey. Signal of each component block is analyzed separately before the signal of assembled circuit is analyzed. The experiment results show instrumentation amplifier functions as designed. When a potential difference occurs between the two electrodes, the signal is

amplified and displayed as oscillations displayed on the screen. Meanwhile, the notch filter does not work as expected, when the 50Hz noise wave has not been removed from the power source. With high amplitude, this frequency overwhelms the signals received from the electrode, it sometimes causes the wanted signal to be drowned in the noise signal. To confirm the above cause, the result then is examined again on the oscilloscope. The results show that the wavelengths in the frequency range 49Hz-51Hz still have the high amplitude. Therefore, it shows that the design of the notch filter needs to be further modified to obtain the neural signals clearly. Spectral analysis also shows that the bandpass filter works as expected with the frequency range from the filter's output in the range of 10Hz - 506Hz, which is an acceptable result.

Figure 31: Spectrum analysis result with Arduino Mega 2560 (Lai, 2023).

Picture A in figure 31 shows the signal is level when there are no connections with two electrodes. Picture B shows the signal appearing on screen when two electrodes are attached to skin. In picture C, spectrum analysis result of the signal is represented, and picture D shows the histogram of frequencies which have the highest peak in the frequency spectrum over time. The results of spectrum analysis show that the signal received from the electrodes mounted on the back of the hand has a frequency ranging from 40Hz - 180Hz, the highest frequency when the hand is at rest is 95Hz - 100Hz, while the frequency is highest when the hand is at rest. The highest number of recorded hand movements is 120Hz - 128Hz. When changing analog input pins, noise can be detected from adjacent pins.

5 CONCLUSION

The observations presented in the thesis are hints about the existence of logical algorithms embedded in the genome. If this hypothesis is proven correct, nature may once again make people humble about our current understandings. There are many possible explanations for the expression of these algorithms from genes to behavior. First, these algorithms depend entirely on the final product of the gene - the proteins themselves. Under physical and chemical properties, the constraints and limitations between proteins and their interactions cause information in DNA to manifest into traits and behaviors. The second explanation is the interaction between proteins and the DNA's intermediate product which is RNA, the binding or reaction between types of RNA and protein contributes to the transition expression from DNA to external behaviors (because RNA is usually short-lived) (Mofatteh,2020). The last possibility is that the protein directly interacts with the DNA, at which point the DNA becomes the framework for the protein to read information and express algorithms to the outside environment. In a study the difference between determinate and indeterminate tomato cultivars, it was found that the protein binds to DNA and uses it to determine growth patterns (Pnueli, 2001). At the present, there are not enough observations to rule out any of these three possibilities, so the mechanism of algorithms expression from genes to behavior remains a question with no complete answer. The best way to deduce algorithms in DNA is to directly observe cells while they are alive and performing their activities. However, with current techniques, this idea is difficult to implement. Most of our observations are made when cells have been destroyed or separated from their operating environment. Two approaches are discussed in the topic; however, this thesis has not completely solved the problems posed in these two approaches. Following future works could make this research more robust:

- Gene expression approach: There are three unresolved problems in this approach. Firstly, about the sampling technique, the current sampling methods almost disrupt the cell structure, so the spatial distribution information is lost. This the spatial distribution has a certain role in gene expression. The second problem is the mathematical theory to solve the system of operator vacancy equations. Due to the lack of this theory, problem solving methods often use iteration strategy in very large searching spaces. Hence it is inefficient and not feasible. The last problem is that there are many features of the graph which have not been known yet. Each feature of the graph is a clue of the arrangement of functions in a system. Therefore, the deeper the understanding of these characteristics, the more effectively information about the system can be exploited.
- Neural signal acquisition approach: In the thesis, the first step of algorithm discovery process in the nervous system was presented. That is finding an implantable device to observe and record the activity of neurons. However, once the signals are received, these data need to be analyzed, with the final goal to find the *F* functions that the brain uses to process signals from the sensory organs and build perceptions from these signals. Although part of the problem, this problem has not been discussed in the thesis.

In conclusion, the logic algorithms are still hidden in the creature's DNA. A well understanding of these logics would be keys to understand the origin of biological intelligence and of life. With the right consideration and prudence, one day the current challenges can be overcome to unravel these algorithms and apply them on variety of fields.

REFERENCES

Alessandro Sergi. 2009. *Quantum Biology.* Journal Atti della Accademia Peloritana dei Pericolanti - Classe di Scienze Fisiche, Matematiche e Naturali. vol 87. [Accessed 15 December 2022]. Available from: https://arxiv.org/abs/0907.1968.

Chen, J.C., Kan, P., Yu, Z. et al. 2022. *A wireless millimetric magnetoelectric implant for the endovascular stimulation of peripheral nerves*. Nat. Biomed. Eng. vol 6. pp 706–716. [Accessed 5 April 2023]. Available from: https://www.nature.com/articles/s41551-022-00873-7.

Christopher Sutardja. 2017. *RFID Reader Design for Neural Implants*. Berkeley: University of California. Thesis. [Accessed 20 September 2022]. Available from: https://escholarship.org/uc/item/4zq1v17c.

Crama, Y., Hammer, P.L. & Ibaraki, T. 1988. *Cause-effect relationships and partially defined Boolean functions*. Ann Oper Res. vol 16. pp 299–325. [Accessed 20 July 2022]. Available from: https://orbi.uliege.be/bitstream/2268/138872/1/CauseEffectRelationships.pdf

D. S. Brox, X. Chen, S. Mirabbasi and K. Takahata. 2016. *Wireless Telemetry of Stainless-Steel-Based Smart Antenna Stent Using a Transient Resonance Method*. IEEE Antennas and Wireless Propagation Letters. vol 15. pp 754-757. [Accessed 7 April 2023]. Available from: https://ieeexplore.ieee.org/document/7219394

Donald B Johnson. 1975. *Finding all the elementary circuit of a graph*. SIAM Journal on Computing. vol 4. 1. Page77-84. [Accessed 20 October 2022]. Available from: https://www.cs.tufts.edu/comp/150GA/homeworks/hw1/Johnson%2075.PDF

E.R. Dougherty, I Shmulevich, 2010, *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*. United States: Society for Industrial and Applied Mathematics. [Accessed 5 January 2022]

E. Y. Chow, Y. Ouyang, B. Beier, W. J. Chappell and P. P. Irazoqui. 2009*. Evaluation of Cardiovascular Stents as Antennas for Implantable Wireless Applications*. IEEE Transactions on Microwave Theory and Techniques. vol 57. no 10, pp 2523-253. [Accessed 7 April 2023] Available from: https://ieeexplore.ieee.org/document/5247020.

Eghbalnia, Hamid & Wilfinger, William & Mackey, Karol & Chomczynski, Piotr. 2020. *Coordinated analysis of exon and intron data reveals novel differential gene expression changes*. Scientific reports. vol 10. 15669. [Accessed 15 February 2023] Available from: https://www.nature.com/articles/s41598-020-72482-w.

Elliott TA, Gregory TR. 2015.*What's in a genome? The C-value enigma and the evolution of eukaryotic genome content*. Philos Trans R Soc Lond B Biol Sci. [Accessed 15 February 2023] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4571570/.

Enzo Grossi, Massimo Buscema. 2007. *Introduction to artificial neural networks*. European Journal of Gastroenterology & Hepatology. vol 19. pp 1046-54. [Accessed 5 January 2023]. Available from: https://www.researchgate.net/publication/5847739_Introduction_to_artificial_neural_networks.

Fairchild Semiconductor Cooperation. Inc. 1986. 74LS153 Datasheet Dual 1 of 4 lines data Selectors/Multiplexers, USA. [Accessed 10 April 2023]. Available from http://eeshop.unl.edu/pdf/DM74LS153.pdf.

Fajriyah, Rohmatul. 2021. *Paper review: An overview on microarray technologies*. Bulletin of Applied Mathematics and Mathematics Education. 1. 21. [Accessed 15 February 2023]. Available from: https://www.researchgate.net/publication/350862726_Paper_review_An_overview_on_microarray_technologies

Felix M Weidner and others. 2021. *Capturing dynamic relevance in Boolean networks using graph theoretical measures*. Bioinformatics. vol 37. no 20. pp 3530–3537. [Accessed 15 January 2023]. Available from: https://academic.oup.com/bioinformatics/article/37/20/3530/6275260

Fukushima K. 1980. *Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*. Biol Cybern. vol 36. no 4. pp 193-202. [Accessed 1 May 2023]. Available from: https://www.rctn.org/bruno/public/papers/Fukushima1980.pdf.

Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G. 2008. *Synchronous versus asynchronous modeling of gene regulatory networks*. Bioinformatics. 24(17). [15 March 2023]. Available from: https//: doi.org/10.1093/bioinformatics/btn336.

Govindarajan R, Duraiyan J, Kaliyappan K, Palanisamy M. 2012. *Microarray and its applications*. J Pharm Bioallied Sci. 4(Suppl 2). [Accessed 15 February 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3467903/.

Graham R. Fleming, Gregory D. Scholes, Yuan-Chung Cheng. 2011. *Quantum effects in biology*. Procedia Chemistry. vol 3, no 1, pp 38-57. [Accessed 15 December 2022]. Available from: https://www.sciencedirect.com/science/article/pii/S1876619611000507.

Jo BS, Choi SS. 2015. *Introns*: *The Functional Benefits of Introns in Genomes. Genomics Inform*. vol 13. no 4. pp 112-8. [Accessed 15 February 2022]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4742320/

Karl R. Gegenfurtner and Lindsay T. Sharpe. 1999. *Color Vision: From Genes to Perception*. England: Cambridge University Press. [Accessed 2021].

Kauffman S, Peterson C, Samuelsson B, Troein C. 2003*. Random Boolean network models and the yeast transcriptional network.* Proc Natl Acad Sci USA. [Accessed 15 February 2022]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC299805/

Kortmann, Jens & Narberhaus, Franz. 2012*. Bacterial RNA thermometers: Molecular zippers and switches*. Nature reviews. Microbiology. vol 10. pp 255-65. [Accessed 5 February 2022]. Available from: https://www.researchgate.net/publication/221708753_Bacterial_RNA_thermometers_Molecular_zippers_and_switches

Kutre, Tejaswini & Patil, Sujata & Kore, Sheela. 2020. *Design of MUX based Flash ADC for Reduction in Number of Comparators*. IOP Conference Series: Materials Science and Engineering. pp 925. [Accessed April 2023]. Available from: https://iopscience.iop.org/article/10.1088/1757-899X/925/1/012063/meta

Leong, Mei Chee, Dilip K. Prasad, Yong Tsui Lee, and Feng Lin. 2020. *Semi-CNN Architecture for Effective Spatio-Temporal Learning in Action Recognition*. Applied Sciences. vol 10. no. 2. pp 557. [Accessed 5 January 2023]. Available from: https://www.mdpi.com/2076-3417/10/2/557

Liu CH, Chen SC, Hsiao HM. 2019. *A Single-Connector Stent Antenna for Intravascular Monitoring Applications.* Sensors (Basel). vol 19. no 21. [Accessed 15 April 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6864709/.

Mark R. Fielden , Tim R. Zacharewski. 2001. *Challenges and Limitations of Gene Expression Profiling in Mechanistic and Predictive Toxicology*. Toxicological Sciences. vol 60. no 1. pp 6–10. [Accessed 20 February 2023]. https://doi.org/10.1093/toxsci/60.1.6

Martin KC, Ephrussi A. 2009. mRNA localization: gene expression in the spatial dimension. Cell. vol 136. no 4. pp 719-30. [Accessed 5 March 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2819924/

Matthew Browne, Saeed Shiry Ghidary. 2020. *Application of Convolution Neural Network for Digital Image Processing*. Engineering International. vol 8. No 2. [Accessed 5 February 2023]. Available from: https://abc.us.org/ojs/index.php/ei/article/view/592

McGlynn, Eve & Nabaei, Vahid & Ren, Elisa & Galeote-Checa, Gabriel & Das, Rupam & Curia, Giulia & Heidari, Hadi. 2021. *The Future of Neuroscience: Flexible and Wireless Implantable Neural Electronics*. Advanced Science. vol 8. No 10. pp 1002. [Accessed 25 April 2023]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/advs.202002693.

Michael L. Goldberg, Leland H. Hartwell, Janice A. Fischer, Leroy E. Hood. 2021. *Genetic: From Genes to Genomes*. 7th Edition. McGraw-Hill Education. USA. [Accessed 2022].

Mofatteh M. 2020. *mRNA localization and local translation in neurons*. AIMS Neurosci. vol 7. no 3. pp 299-310. [Accessed 15 March 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7519968/

Oribe, S., Yoshida, S., Kusama, S. et al. 2019. *Hydrogel-Based Organic Subdural Electrode with High Conformability to Brain Surface*. Sci Rep 9, 13379. [Accessed 5 April 2023]. Available from: https://www.nature.com/articles/s41598-019-49772-z.

Oxley, Thomas & Yoo, Peter & Rind, Shafiquat & Ronayne, Stephen & Lee, C & Bird, Christin & Hampshire, Victoria & Sharma, Rahul & Morokoff, Andrew & Williams, Daryl & MacIsaac, Christopher & Howard, Mark & Irving, Lou & Vrljic, Ivan & Williams, Cameron & John, Sam & Weissenborn, Frank & Dazenko, Madeleine & Balabanski, Anna & Opie, Nicholas. 2020. *Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis: First in-human experience*. Journal of NeuroInterventional Surgery. vol 13. [Accessed 7 February 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7848062/

Pan C, Han Y, Lu J. 2021. *Structural Design of Vascular Stents: A Review.* Micromachines (Basel). vol 12. no 7. [Accessed 5 April 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8305143/

Philipp L Antkowiak. 2021. *Engineering DNA Data Storage via Low-cost DNA Synthesis and an Automated Storage Architecture*. Germany: ETH Zurich University. Thesis. [Accessed 7 September 2022]. Available from: https://www.research-collection.ethz.ch/handle/20.500.11850/527005.

Pnueli L, Gutfinger T, Hareven D, Ben-Naim O, Ron N, Adir N, Lifschitz E. 2001. *Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering*. Plant Cell. vol 13.

no 12. [Accessed 9 September 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139482/

Righetti F, Narberhaus F. 2014. *How to find RNA thermometers. Front Cell Infect Microbiol*.. vol 4. pp 132. [Accessed 5 February 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4166951/

Sakharkar, Meena & Chow, Vincent & Kangueane, Pandjassarame. 2004. *Distributions of exons and introns in the human genome*. In silico biology. vol 4. pp 387-93. [Accessed 10 January 2023]. Available from: https://www.researchgate.net/publication/8491627_Distributions_of_exons_and_introns_in_the_human_genome.

Saptono, Debyo & Wahyudi, Bambang & Irawan, Benny. 2016. *Design of EEG Signal Acquisition System Using Arduino MEGA1280 and EEGAnalyzer*. MATEC Web of Conferences. vol 75. [Accessed 9 April 2023]. Available from: https://www.researchgate.net/publication/312140464_Design_of_EEG_Signal_Acquisition_System_Using_Arduino_MEGA1280_and_EEGAnalyzer.

Schwab JD, Kühlwein SD, Ikonomi N, Kühl M, Kestler HA. 2020. *Concepts in Boolean network modeling: What do they all mean?* Comput Struct Biotechnol J. vol 10. no 18. pp 571-582. [Accessed 10 January 2023]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7096748/

Shi, Ning, 2021, *Boolean network inference and control using metaheuristic algorithms*. England: University of Birmingham. Thesis. [Accessed 15 September 2022]. Available from: https://etheses.bham.ac.uk/id/eprint/11522/1/Shi2021PhD.pdf

Tatsuya Akutsu. 2008. *Algorithms for Analysis, Inference, And Control of Boolean Networks*. Japan: Kyoto. [Accessed April 2022].

Vahid Elyasigomari. 2017. *Analysis of microarray and next generation sequencing data for classification and biomarker discovery in relation to complex diseases*. England: University of London. Thesis. [Accessed 5 February 2023]. Available from: https://qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/25817/Elyasigomari_V_PhD_final_280717.pdf?sequence=1.

Veerendra Dasari. 2013. *EEG Acquisition System on Mobile Platform*. United States: Western Michigan University. Thesis. [Accessed 5 May 2023]. Available from: https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=1128&context=masters_theses.

Watson T, Webster MWI, Ormiston JA, et al. 2017. *Long and short of optimal stent design*. Open Heart. vol 4. [Accessed 10 March 2023]. Available from: https://openheart.bmj.com/content/4/2/e000680

Weltman A, Yoo J, Meng E. Flexible. 2016. *Penetrating Brain Probes Enabled by Advances in Polymer Microfabrication*. Micromachines (Basel). vol 7. no 10. pp 180. [Accessed 8 March 2023]. Available from: https://pubmed.ncbi.nlm.nih.gov/30404353/.

X. Zou, W. -S. Liew, L. Yao and Y. Lian. 2010. *A 1V 22μW 32-channel implantable EEG recording IC*. IEEE International Solid-State Circuits Conference - (ISSCC), San Francisco, CA, USA, pp. 126-127. [Accessed 25 April 2023]. doi: 10.1109/ISSCC.2010.5434024.

Yang W, Gong Y, Li W. 2021. *A Review: Electrode and Packaging Materials for Neurophysiology Recording Implants*. Front Bioeng Biotechnol. vol 14. [Accessed 12 April 2023]. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2020.622923/full.

Yi, Ying & Chen, Jiaxu & Selvaraj, Madeshwaran & Hsiang, York & Takahata, Kenichi. 2019. *Wireless Hyperthermia Stent System for Restenosis Treatment and Testing with Swine Model*. IEEE Transactions on Biomedical Engineering. [Accessed 20 February 2023]. Available from: https://pubmed.ncbi.nlm.nih.gov/31449000/

Yiming, Dong & Sun, Fajia & Ping, Zhi & Ouyang, Qi & Qian, Long. 2020. *DNA storage: Research landscape and future prospects*. National Science Review. vol 7. pp 1092-1107. [Accessed 7 September 2022]. Available from: https://academic.oup.com/nsr/article/7/6/1092/5711038

Zhou, Yuhao, Bowen Ji, Minghao Wang, Kai Zhang, Shuaiqi Huangfu, Huicheng Feng, Honglong Chang, and Xichen Yuan. 2021. *Implantable Thin Film Devices as Brain-Computer Interfaces: Recent Advances in Design and Fabrication Approaches*. Coatings. vol 11. no 2. pp 204. [Accessed 12 March 2022]. https://doi.org/10.3390/coatings11020204

APPENDIX 1: Copyright license.

The thesis used visual materials from other documents, which require to include the licenses under electronic forms properly. This appendix represents required compliance certificates and licenses from the author or issuer:

- Figure 16 in this thesis used the material from the document of Liu, 2019. The picture is remixed with other materials in this paper. The license can be found at https://creativecommons.org/licenses/by/4.0/.
- Figure 21 in this thesis used the material from the document of Kutre, 2018. The license can be found at https://creativecommons.org/licenses/by/3.0/.
- Figure 22 in this thesis used the material from document of Walton, 2017. The license can be found at https://creativecommons.org/licenses/by-nc/4.0/.
- Figure 23 in this thesis used the material from Brox, 2015 with the copyright from © 2015 IEEE, in the regulations on the permission to reuse materials for thesis/dissertation.
- Figure 24 in this thesis used the material from Weltman, 2016. The pictures from original paper are used partially and combined with other materials in figure 24. The license can be found at: https://creativecommons.org/licenses/by/4.0/.
- Figure 24 in this thesis Figure 24 in this thesis used the material from Oribe, 2019. The picture from original paper have combined with other materials in figure 24. The license can be found at: https://creativecommons.org/licenses/by/4.0/.
- Figure 24 in this thesis used the material from Ershad, 2018 with the copyright from © 2018 IEEE, in the regulations on the permission to reuse materials for thesis/dissertation.