DESIGN AND IMPLEMENTATION OF CMOS NEURAL NETWORKS

A Dissertation

Submitted

By

M.SACHIN KUMAR

(Reg.no.11306921)

To

Department of Electronics and Communication Engineering

In partial fulfillment of the Requirement for the

Award of Degree of

MASTER OF TECHNOLOGY

IN

VLSI DESIGN

Under the guidance of

Mr. TEJINDER SINGH
Assistant Professor

School of Electronics Engineering

Lovely Professional University

Phagwara-144402, Punjab (India)

(May 2015)

CERTIFICATE

This is to certify that M.Sachin Kumar has completed M.tech dissertation titled **"DESIGN AND IMPLEMENTATION OF CMOS NEURAL NETWORKS "**under my guidance and supervision. To the best of my Knowledge, the present work is the result of his original investigation and study. No part of the dissertation has ever been submitted for any other degree or diploma.

The dissertation is fit for the submission and the partial fulfilment of the conditions for the award of M.tech in VLSI Design.

Date: Signature of Advisor

Name: Mr Tejinder Singh

UID: 17324

DECLARATION

I hereby declare that the Dissertation-II report entitled **"DESIGN AND IMPLEMENTATION OF CMOS NEURAL NETWORKS"**, is an authentic record of my own work carried out as the requirements for the award of degree of Master of Technology in VLSI Design at Lovely Professional University, Jalandhar under the guidance of **Mr. Tejinder Singh,** Assistant Professor, Department of Electronics and Communication Engineering, during January to May, 2015.

Dated:

M.SACHIN KUMAR

(11306921)

Objective of the Dissertation-II is satisfactory / unsatisfactory

Examiner I **Examiner II**

ACKNOWLEDGEMENT

This is a humble effort to express my sincere gratitude towards my mentor who has guided and helped me to explore the field of artificial neural networks, which is the base of my Dissertation undertaking.

Problem formulation is a major milestone for a student while working on Dissertation. As such this subject was a challenge for me and was an opportunity to prove my caliber. Being a beginner, I faced many problems, which would have frustrated me. I am highly grateful and obliged to my mentor.

It would not have been possible to see through the undertaken field of study without the guidance of Mr**. Tejinder Singh.** It was purely on the basis of his experience and knowledge that I am able to clear the theoretical and technical hurdles during the analyses and exploration of the selected field of study.

I would like to thank the **Project Approval Committee members** for their valuable comments, **Mr. Bhupinder varma, DOD (School of electronics), Ms. Cherry Bhargava, HOD (Dept. of ECE)**, the **faculty members of VLSI department** for the valuable suggestions and **my friends** who supported in lab work and I would also like to thank **Lovely Professional University** for facilitating me to undertake this study.

M.SACHIN KUMAR
(11306921)

ABSTRACT

This thesis work mainly concentrates the infiltration of advanced electronics into the field of biomedical sciences, which leads to the generation and development of neuromorphic designs. These neuromorphic designs will become a pavement for a greater breakthrough in the history of technology. These designs are made to synchronize with that of the functionality of human brain and forms as one of its part. These designs has got their basis from the artificial neural networks which are being designed with the help of CMOS transistors. These CMOS neural networks are made capable of performing complex computations, adoptions and learning using analog domain with the help of massive connectivity. As we know, how complex is a human brain with a vast network of neurons, which take part in many bodily activities and in fact control many of the motor mechanisms such as walking, swimming and so on. To include these dynamic behavior of the biological neuron is the challenging part, as it has a time varying response to the provided stimulus. Along with the neurons, synapses also play a vital role, as they are the one through which a communication between neurons takes place. Synapses are the one which converts the chemical signals into electrical pulses which is fed to another neuron. So more studies including the signal transmission of neurons through synapses are required.

This thesis demonstrates the way in which biological systems operate in analog domain by making use of CMOS neural networks. Whatever may be the design it should replicate the biological neuron properties, but it is impossible to mimic such complex nature of the human brain. So the work insights on the design which closely tries to mimic few of the biological neuron properties.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	–	Artificial neural networks
CMOS	–	Complementary metal oxide semiconductor
CNN	–	CMOS neural networks
NAF	–	Neural activation function
DNAF	–	Derivative of neural activation function

CHAPTER-1
INTRODUCTION

1.1 What Are Artificial Neural Networks?

Artificial neural-networks are simply an artificial implementation of human brain and it becomes the main motivation for neural networks as the human brain functions entirely in a different prospect from that of the conventional computers. In general, human brain is not only highly complex but also a non-linear system and as well as a parallel computer. The brain is a collective structure of massively connected neurons, in the same way the ANN's are made up of artificial neurons. In the most general form, a neural network is a machine that is being designed in order to model the way in which the human brain accomplishes a certain task. The ANN is usually implemented with the help of electronic components or it can also be simulated using any software. As previously discussed, neural networks closely replicate the human brain, it means all the functions that are being done by a typical human brain should be able to done by the designed artificial neural networks.

The design of digital computers have also been motivated by the functions of human brain, alike human brains, they also performs a large number of functionalities and computations within a fraction of seconds. Then what is the need of developing artificial neurons? Compared to conventional computers, ANN's have many advantages, like they can handle functionalities even now-a-days computers cannot. Unlike computers, ANN's are not programmable, and can learn adaptively by many learning processes, such that there is no need of reprogramming it to perform different functions. So the study of artificial neural networks is motivated by their similarity to successfully working biological systems, which in comparison to the overall system, consists of very simple but numerous nerve cells that work collectively in parallel and have the capability to learn. Humans can perform their own functions by self-learning, in the same way the artificial neurons learn by learning algorithms such as back propagation algorithm. One result from this learning procedure is the capability of neural networks to generalize and associate data.

1.2 Biological Neuron and Activation Function:

The neurons are the most basic information processing cells of human brain. All the movements, thoughts, memories, feelings, sensations and so on are the result of the signals that are passing through massive network of neurons. A neuron is nothing more like a switch with an information input and output. This switch will be activated if there are

enough stimuli from other neurons hitting the information input. Then, at the information output, a pulse or a corresponding response is sent to, other neurons.

The main components of neurons are: dendrites, the one which receives input electrical signals and transfer it through them into the cell body, shown in Fig.1.1. Axons, which transfers the electrical signal off the cell body to other neuron for further processing. Synapses are the basic functional units that guides or mediates the information and interaction between the neurons. They are the points at which a unidirectional conduction of a signal from presynaptic to the postsynaptic membrane takes place. The synapses strengthen each and every signal and stretch it out in time so the postsynaptic potential has larger amplitude and a longer duration than the presynaptic potential. Chemical synapses can either preserve the electrical sign of the potential or reverse it by the chemical mediation of the synapse. Synaptic transmission begins with the action potential hitting the terminal of the axon. A synapse transduces electrical signal to equivalent chemical signal and then chemical signal to electrical signal.

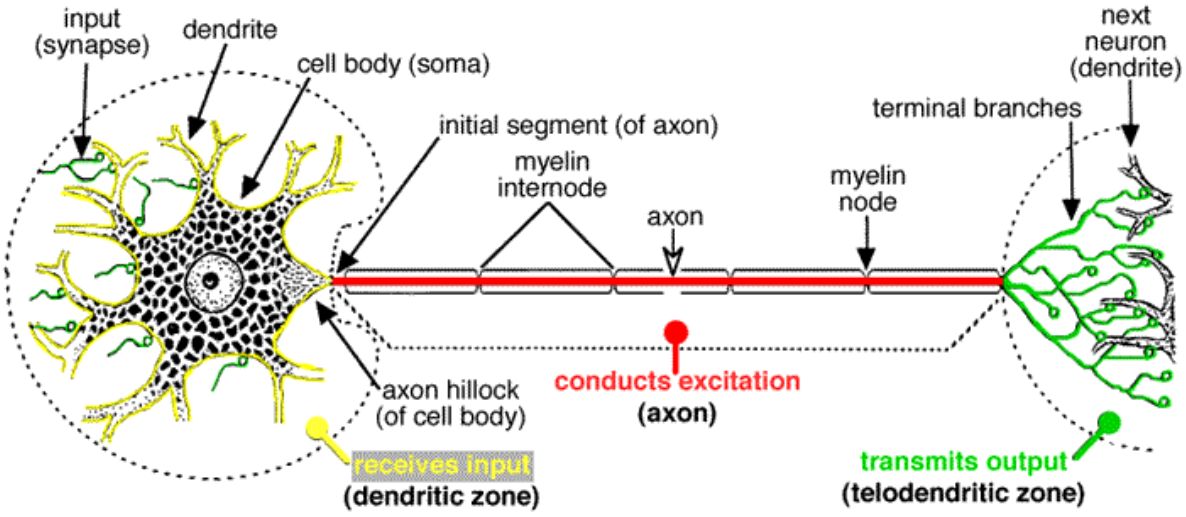


Fig.1.1: biological neuron

Due to a variations of ion concentrations inside and outside the cell, there is a difference in electric potential across the membrane, called the resting membrane potential, which is -65mV. The maintenance of the resting potential is important for the functioning of a neuron, by which it decides whether or not to transmit information. External signal impinging on the neurons at synapses creates disturbances in the cell potential called graded potential. When an impinging signal lowers the internal potential below its resting potential, then the neuron is said to be hyper polarized, and it is depolarized when the potential is above the resting potential.

Action Potential:

Neurons continuously gathers inputs from other neurons all along their dendrites at points of contact called synapses. As previously mentioned, these inputs take the form of small electrical disturbances that are referred to as post synaptic potentials. Near axon hillock, the ion channels exists in considerably high density, these are highly susceptible to minor perturbations. These ion channels continuously monitor the cell body potential such that when the cell potential exceeds a threshold about –40mV, the neuron triggers an action potential shown in Fig.1.2 that is passed down to axon towards a synaptic terminal i.e. the cells "listen" to the neuron. As mentioned above, at the axon hillock, electrically specialized gated sodium and potassium channels are found. Electrically gated means, the proteins that regulate these channels are sensitive to deviations in the membrane potential from its resting value of –65mV. Now as long as the cell is hyperpolarized or depolarized nothing unusual happens. However, when the depolarization is sufficiently large enough to push the membrane potential above –40mV, the specialized ionic channels open up, permitting the free passage of specific ions in and out of the axon.

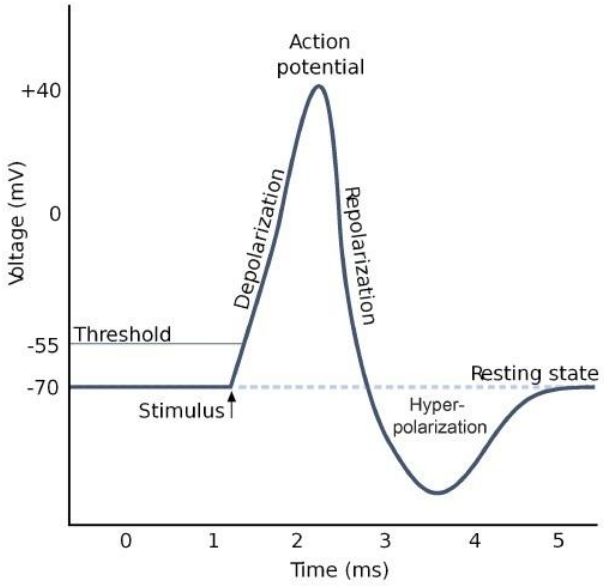


Fig.1.2: Initiation of action potential

CHAPTER-2
SCOPE OF THE STUDY

Over the past few decades, numerous studies have been performed on artificial neural networks. Many researchers have proposed different neuron models, but however most of the proposed neural networks were limited to simple neuron models in which the complex dynamic behavior of biological neurons such as transmission of signal between neurons through synapses and the time varying response of the neuron. In 1952, Hodgkin's time varying neuron model approach was one of the first most attempts to model and realize a real neuron. An equivalent electrical circuit model, which represents the active membrane potential of neuron in the form of differential equations, has been proposed. Since then various advancements took place in the proposals of neuron models, such as theoretical analysis have been reported and applying these results to many of the commercial neural network software programs have been done. Later came the well-known integrate and fire (I&F) neuron model of a silicon neuron published by Mead in 1989, which uses several MOS transistors and capacitors. Sooner or later many types of integrate and fire neuron models have been reported, such as multi-conductance based neuron model using low pass filters and multipliers. Despite of these many advancements, still it is not easy to mimic a human neuron because of its non-linearity and complex dynamic behavior. Therefore we need few more experimental studies which include the time varying neuron's signal transmission via synapses.

Out of all the models of neurons, the perceptron model which is also referred to as the mathematical model of neuron yields approximately precise results of the biological neurons. Generally neural networks are employed when a particular problem has no algorithmic solution or when a problem is very much complicated such that it cannot be solved by existing algorithms. ANN's can be used when there exists no definition for a problem, but the input samples and the corresponding outputs are available. To understand intelligence and to develop artificial intelligent machines or computers, we need to study the brain and the neurons, and how neurons collectively work to solve different problems. In that case, the corresponding neural implementation will be generally larger and less accurate than the direct algorithmic solution.

We are going to compare an artificial neuron with biological neuron shown in Fig.2.1. We can observe that, they both take inputs, uses weights and generates an output. Artificial neural networks comprises a large number of parallelly operating computational

blocks. These blocks are nothing but the neurons, which are the fundamental processing elements with a certain number of inputs and a single output that branches into collateral connections, which forms as the input for other neurons. Generally they perform a nonlinear function (activation function) on the sum of the inputs. These neurons are massively interconnected through their weight strengths and their interconnections are typically referred to as synapses which controls the influence of the neurons on others neurons. The processing done by the synapse is typically a multiplication between the neuron outputs and synaptic weights. Each neuron's output level depends on the outputs of the connected neurons and on the synaptic weights.

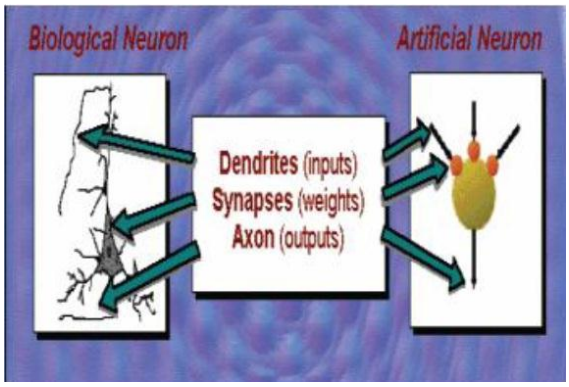


Fig.2.1: Comparison of biological and artificial neuron

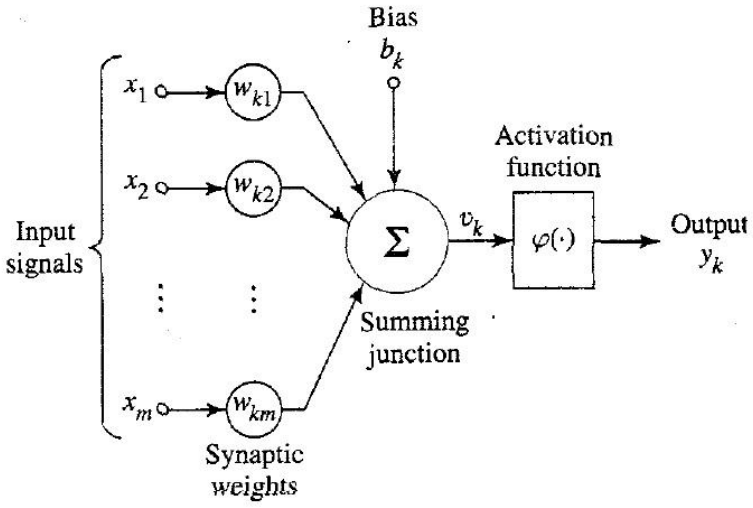


Fig.2.2: basic model of neuron

2.1 Basic Artificial Neural Network Block Diagram:

Generally neurons are the fundamental information processing units for the operation of a neural network. The perceptron is a mathematical model of a biological neuron. We model this phenomenon in the form of a perceptron by evaluating the weighted sum of the inputs to represent the total strength of those signals, and by applying an

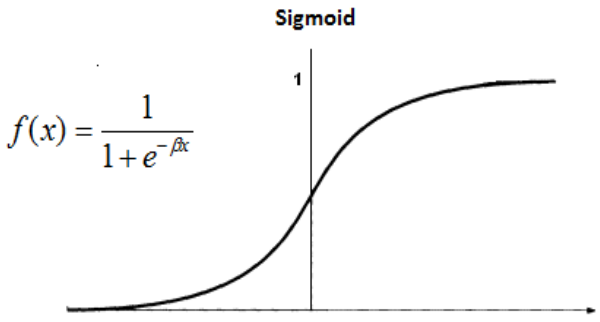
activation function on the sum of input signals and weights we can determine its output. As in biological neural networks, this output is fed to other perceptron. The Fig.2.2 shows a model of neuron, which forms the basis for designing of neural networks.

The three basic components of artificial neuron are:

1. The connecting links referred as synapses that are characterized by a weight or strength of its own. A signal x_j at the input of synapse j connected to the neuron k is multiplied by the synaptic weight w_{kj}.
2. An adder for summing all the input signals, weighted by the respective synapses of the neuron.
3. An activation function to limit the amplitude of the neuron output. It is used for mapping the inputs and the outputs**.**

Neural Activation Function:

Every neuron in a neural network has an activation function which evaluates the output of neuron with respect to given input. Neurons are nothing but a switch which output a 1 when they are sufficiently activated, and a 0 when they are not. The activation function in the biologically inspired neural networks is usually an abstraction which represents the rate of action potential firing in the cell. In its simplest form, activation function is binary that is, either the neuron is firing or not. There are different activation functions used for the functioning of artificial neural networks, such as tan sigmoid function, linear function, Gaussian function, log sigmoid function etc. One of the activation functions most commonly used for neurons is the sigmoid function.



2.2 Network Architectures: In general we may identify two different classes of neural architectures which are discussed below.

2.2.1 Single-layer feed forward networks:

The neurons are organized in the form of layers, in layered neural network. In its simplest form, a layered network has source nodes which are the input layers that projects onto an output layer of neurons, but it cannot be vice versa. Simply, this network is strictly a feed-forward or acyclic type. The below figure represents such network with four nodes. Such a network is called feed forward single layer network, with the designation 'single layer' referring to the output layer of the computation nodes. Here we do not take the count of source nodes as no computation is performed there.

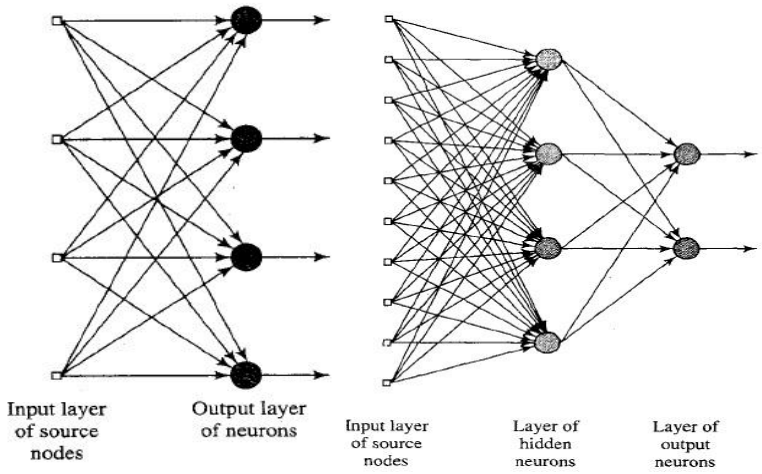


Fig.2.3: feed-forward network with a single layer and multiple layer of neurons

2.2.2 Multiple layer perceptron:

The second class of layered neural networks can be distinguished with the presence of one or more hidden layers, whose computational nodes are called hidden units or hidden neurons shown in Fig.2.3. The function of hidden units is to intervene in between the provided external input and the overall network output. By introducing one or more hidden units, the network is enabled to extract higher order statistics, particularly when the size of input layer is large. The source nodes of the network supply respective elements of the activation pattern, which comprises the input signals fed to the neurons in the next layer or the first hidden layer. The output signals of that layer are used as the input to the further next layer and so on for the rest of the network. Typically the neurons in each layer of the network have as their inputs, the output signals of the preceding layer only. The set of all the output signals of neurons in the final output layer of the network constitutes the overall response of the network to the activation patter fed by the source nodes in the first input layer.

CHAPTER-3

OBJECTIVES OF THE STUDY

The work containing the implemented circuit should possess the following design considerations for its improved functionality, they are as follows:

- The model circuitry should have a compact size, it means the area must be very less such that when an architecture is made with such models it should occupy much less area because brain is a massive interconnection of neurons, in order to replace them, the designed model should be as compact as that of the biological one.
- By considering the design specifications of already proposed models, we are modifying an existing model with comparably lesser supply voltage.
- Low power electronics must be worked out to realize basic operations required in artificial neural networks.
- The implemented model is a mathematical model of neuron so that it should possess almost similar characteristics of biological neuron while processing.
- Signals transmitted between neurons must be voltages such that the power losses in the conductive path are reduced.
- The neuron models must faithfully reproduce the dynamics of individual neuron, such as the action potential.
- The magnitude of the electrical currents produced by the circuit should be same order of magnitude as biological neuron.
- The delays pertaining to the provided input stimulus should be comparably less.
- The functionalities offered by the design should closely replicate to that of neurons in real time, such as generating different responses corresponding to different stimulus.

CHAPTER-4
REVIEW OF LITERATURE

B. M. Wilamowski, J. Binfet, and M. O. Kaynak et.al,[1] This paper mainly discusses about the use of neural controller over fuzzy controllers. The fuzzy controllers are easy to implement. Even though the ANN's are complex and harder to design, but they provide much less errors compared to fuzzy controllers. By using a simple shichman-hodge MOS transistor in strong inversion mode for this purpose, we can generate an activation function. This differential pair transistor is a transconductance circuit. The circuit generates two types of current outputs, one is positive and the other is negative. Both the positive and the negative output currents may have flow of currents in both the directions. All the positive weights of the circuit are connected to the positive outputs and all negative weights are connected to the negative outputs. Another problem that arises with that of conceptual neuron circuit is, the input is a voltage and the output is a current. So there is a need to modify the circuit such that the input would also be a current.

Drawback: Implementing size is large (1.5µm technology) and enlarged errors are obtained by this model, but much better than fuzzy controllers.

R. WOJTYNA¤ and T. TALA´SKA et.al,[2] This paper overall deals with the implementation of low power consuming neural networks. In general, a typical neural network is an interconnection of large number of artificial neurons, which consumes more power for their operation. So this paper discusses the criteria, in which the neurons consume power only during active state and no power during inactive state. This paper discusses the use of analog short term memories which store neurons adaptive learning procedures during learning phase and uses them in recall phase. They also make use of additional digital circuit to refresh the analog memory weights. They are developing a static model for neural networks in which we use transconductance synapse circuit and trans-resistance neuron circuit with the help of 0.35µ CMOS technology. This is an improvised model of neuron circuit; this improvement probably relies on the secondary output.

Drawback: size of implementing technology is 0.35µm.

Fitzhugh-Nagumo model et.al,[3] This paper presents a CMOS VLSI implementation of oscillating biological neuron model, known as Fitzhugh-Nagumo model. This model is simplification of Hodgkin and Huxley's model. The Hodgkin and Huxley's model is also known as conductance based model. They have introduced a CMOS Fitzhugh-Nagumo

neuron model, which is intended for biological system emulations. The knowledge obtained by concatenating several of these artificial neurons improves the understanding of biological neuron interactions. The potential to implement such neuron type which almost mimics the biological neural networks and consequently it has been expected that these implementations should provide better performance statistics than other neural networks.
Drawback: not practically implemented due to larger area (2µm)

Yasuhiro Ota and Bogdan M. Wilamowski et.al,[4] This paper provides information about CMOS hardware implementation of voltage mode pulse coupled neural network architecture. The implementation methods adopted in this paper shows inherent tolerance to fault and high speed. This model represents a biological neural network that uses a pulse stream signaling and analog summation and multiplication. Main feature of the proposed neuron circuit is that the structure is very much compact which exhibits all the prerequisite properties of biological neurons. A simple neuron circuit structure with synaptic weight summation and multiplication is described in this paper. The concept neuron initiates reactions with a provided external stimulus, and generates a stream of electrical pulses. The electrical impulse provided to the circuit is current. The synaptic weights in the current mode neuron circuit are controlled with the help current mirrors at the circuit output with proper W/L ratios.

F. J. Pelayo,1 E. Ros - X. Arreguit - A. Prieto1 et.al,[5] This paper presents a very efficient manner of simplifying interconnections in VLSI neural systems, such that the analog signals are represented by means of frequencies of asynchronously produced short pulses, but using temporal multiplexing of addresses of firing neurons on fast digital buses to reduce the number of interconnections required. The proposed circuit also provides a general approach to reproduce the continuous dynamics of few mathematically described neural models. Similar neuron models can be designed with simpler circuits than those proposed; emphasis has been put on implementing process independent functions and accurate control of time constants. Due to the way the input spikes are processed and time multiplexed, several synapses of a neuron can be implemented by single synaptic circuit, provided spike time is preserved. The adjustable weights can be implemented by using current mirrors of variable ratios. For linear synapses, the adjustable voltage references are implemented by using analogue memory cells.

Izhikevich-Neuron Model et.al,[6] This paper deals with the low power implementation of Izhikevich neuron model. The main building blocks used for this model are two log

domain filters of first order. This paper also discusses different current mirror models such as: classical mode (CM), source shifted mode (SSCM) and active diode connection mode (ACM) to reduce the power consumption. One of those filters uses an active diode connection mode in order to reduce the currents levels, which yields in lesser power consumption. Out of the three discussed modes, classical mode is not preferred due to the poor DC performance of the circuit at current levels below 1pA. So SSCM and ACM modes are used to obtain better DC behavior.

Izhikevich model is not only produces rich dynamical behavior but also are computationally efficient in providing easier electronic design. Basically used two log domain VLSI implementations of Izhikevich model are, one is simple model of spiking neuron, which reproduces the rich behavior of biological neurons, including spiking, and mixed mode firing patterns, post inhibitory spikes and bursts, spike threshold variability, sub threshold oscillations and resonance, bistability of resting and spiking states are presented. The other is sub threshold VLSI implementation of simple neuron model, which uses log domain circuit utilizes MOS transistors operating in sub threshold region.

Yasuhiro Ota, Bogdan M. Wilamowski et.al,[7] This paper introduces a CMOS hardware implementation to realize weighting and summing signals by current controlled oscillator for pulse coded neural networks. They also presented a novel design and pulse stream neural cell implementation with summing capability and synaptic weighting. They achieved summation and synaptic weight multiplication by adjusting proper aspect ratios of output current mirror transistors in the neuron circuit. A computational style which mimics a biological neural system using pulse stream signaling and analog summation and multiplication is described in this paper.

Drawback: this model consumes very high power which makes it non feasible for its application in real time.

Il Song Han et.al,[8] this paper is mainly based on Hodgkin and Huxley neuron model and it introduces a circuit with controlled conductance for silicon synapse and neuron, a reference synapse circuit for controlled conductance and a balanced structure of synapse, they presented these circuits with improved functionalities. From electrical equivalent of neuron and HH model as basis a neuron block is implemented with the help of voltage controlled conductance circuit along with the asynchronous spike firing neuron with three synapses of controlled linear conductances. With synaptic spike currents as inputs,

asynchronous spike response of the neuron is being observed and achieved membrane potential through neuron capacitor's potential, conductance control for ionic conductance, firing pulses with the refractory period.

Drawback: fails to explain whether the obtained responses pertains to that of the response of biological neuron or not.

S.Veni, B.Yamuna et.al,[9] this paper clearly explains the requisite properties of CMOS neural networks such as gilbert multiplier, operational transconductance amplifier and current mirror. They presented neuron as a combined block of multiplier, integrator, activation function block and a buffer stage. Folded gilbert multiplier and linear tunable operational transconductance amplifier serves the purpose of multiplier and integrator respectively. This paper utilized sigmoid function circuit to generate activation function for neuron. By combining all these blocks, a neuron architecture is implemented. They discussed each and every block's transient and DC responses.

Drawback: uses higher power supplies for its operation and the neuron architecture response may not have the resemblance of original neuron response.

Neeraj Chasta, Sarita Chauhan, Yogesh Kumar et.al,[10] this paper mainly focuses on the implementation of neural network architecture with on chip learning using analog VLSI technology. For implementation, four quadrant gilbert multiplier, neural activation function generator are used. It also comprises of analog adders along with the tan-sigmoid function circuit with CMOS transistors operating in sub-threshold region. They also introduced the back propagation algorithm and advanced weight storage techniques into the neuron architecture. Further they used the implemented architecture for signal compression and decompression applications.

Drawback: fails to present the prerequisite functionality of biological neuron, which generates the action potential for provided stimulus.

V. Suresh Babu, M.R. Biju et.al,[11] this paper introduces novel circuit configurations to realize neural activation function and its derivative with low power dissipation. The presented circuits realize neural activation function along with its derivative with programmable characteristics. The proposed configurations explain the generation of tan sigmoid and log sigmoid functions and their respective derivatives with the help of the same circuitry by adding a few pairs of transistors and varying the control voltages. Slope and the threshold values of neural activation function and its derivative are continuously

varies and the results are obtained. The proposed circuit uses CMOS differential pairs to obtain NAF and asymmetry in transistors of differential pair to obtain its derivative.

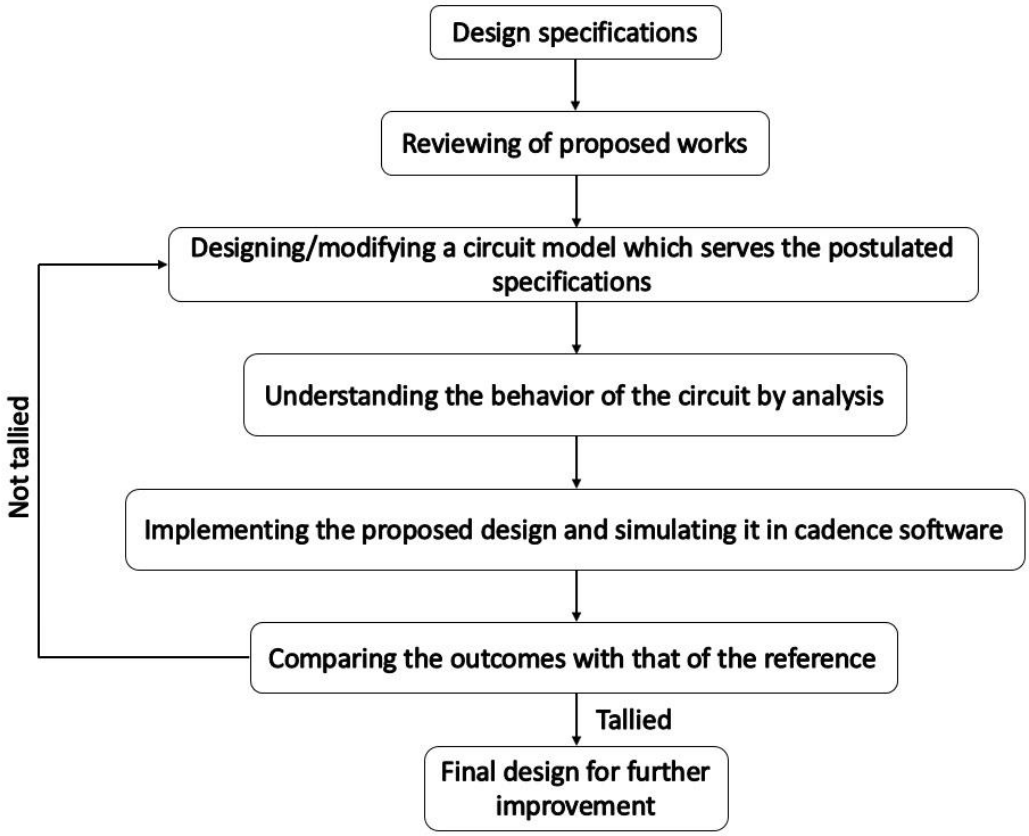
Ethan Farquhar, Paul Hasler, Senior Members, IEEE et.al, [12] this paper presents a circuit model which generates action potentials along with the channel currents of biological neurons. It generates a response by adopting the knowledge of the numerous physical similarities between biological channels and silicon channels. Based on the ion flow in the biological neuron and comparing it with the electron flow in the MOS transistor, this paper presents two channel model circuit which are sodium and potassium channels. These sodium and potassium channels are implemented using four transistors, two capacitors and two transistors, one capacitor. This circuit operates in real time, runs in sub-threshold mode and uses power supply to that of real biology. The proposed circuitry generates action potential after sodium circuit releases a large hyperpolarizing event. This response is consistent with biology and is referred to a post inhibitory rebound.

Drawback: the capacitors used in the circuit are large enough and still have the current levels to be accurately measures. There is a great tradeoff between the channel currents and the capacitors.

Mr.Maulik B.Rami, Prof.H.G.Bhatt, Prof.Y.B.shukla et.al,[13] this paper introduces a model of artificial neuron which comprises of a multiplier block and adder followed by NAF generator. The implementation of those blocks include analog multipliers, operational amplifiers in different configurations. This paper also discusses generating of various activation functions such as step, linear and sigmoid signals. They formulated the whole neuron block by amplifying the outputs of the multiplier circuits and adding them using an op-amp and feeding the obtained signal into the activation function generator. Different W/L ratios of transistors with respect to different circuits are provided along with the simulation results.

Drawback: fails to discuss the overall response of neuron architecture for different sets of stimulus and the output obtained by their circuit does not resembles the action potential of human neuron.

CHAPTER-5
RESEARCH METHODOLOGY



- Design specifications include the designing of neural network based on CMOS transistors. The circuit model should closely resemble the functionality of biological neuron as exact resemblance is still a hypothesis. The circuitry should possess a compact size and should offer lesser delay and power, it should respond to different provided stimuli, in the same manner as the biological neuron responds to various kinds of stimuli from different parts of the body within a fraction of seconds.
- Different models of artificial neurons has been proposed over the past few decades out of which only few complies the dynamic behavior of human neuron. The well-known neuron models include Hodgkin and Huxley model, Fitzhugh-Nagumo neuron model, Integrate and fire neuron model, Izhikevich Neuron Model and so on. In order to meet the required specifications, the present work mainly focuses on the modifying of the perceptron neuron model.
- The perceptron neuron model is a simple mathematical model which demonstrates the functionality of neuron as the function of summated input products. To achieve

this, a gilbert cell serves the purpose of the multiplier which products the inputs with provided weights. For the circuit to operate as neuron we need an activation function with which the neuron trigger the responses. There are many activation functions out of which sigmoid function is commonly used. The multiplier serves as synapses whose output currents are converted to respective voltages and later fed to the activation function generator.

- Understanding the behavior of the circuit is a very crucial part, through which we can have a clear idea of how the inputs are being processed by the circuits and we can also observe the behaviors of transistors involved in the circuit. For the operation few of the transistors are maintained in saturation while the rest of them are forced to run in sub threshold for low power applications. The present circuitry makes use of a power supply of 1.8V according to 180nm technology.
- After designing the neuron model we have to check it for its response to provided inputs and weights. We have to ensure that the neuron model should at least have a resemblance of the original neuron response. If in case it's not met, then again have to modify the design such as the aspect ratio of transistors, weights, and the input stimulus. Unless the designed model mimics the biological neuron we can't step to further process.
- Once the proposed neuron model has got the required response then we check its functionality and its response to various kinds of stimulus.

CHAPTER-6
PRESENT WORK

The present thesis work has taken its basis from the perceptron model of the neuron which takes its inputs from the sensory units and multiply it with the adjustable weights so as to generate a final neuron response as the function of the product of inputs and weights. The designed artificial neuron computes a linear combination of its inputs and applies a sigmoid function to the result. The main use of sigmoid functions is to introduce non linearity in the model. A reason for its popularity in neural networks is because it satisfies a property between derivative and itself such that it is easy to compute. Derivation of sigmoid function is generally employed in learning algorithms. The present design is a simple mathematical representation of biological neuron, and we made use of analog components such as multipliers, adders, and differential amplifiers in order to realize the perceptron model of neuron. The basic block diagram is shown below.

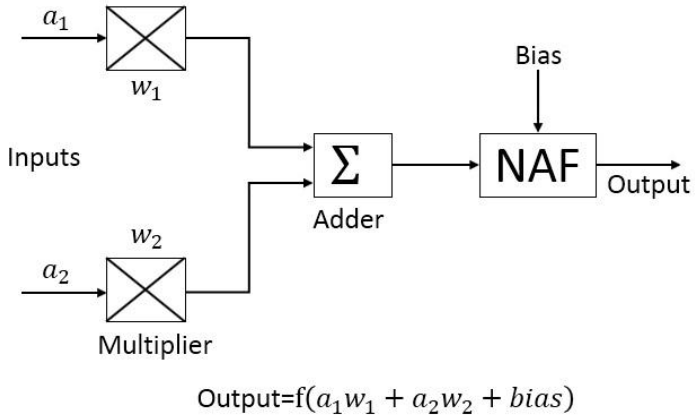


Fig.6.1: perceptron model of neuron using analog components

The multipliers used above represents the synapse, whose basic function is multiplication. Linear multiplications which is depicted by the mathematical model is area expensive in analog integrated circuits. So as a result, simpler circuits that exhibits only approximate multiplication behavior is often preferred. The multiplier is a differential pair which generates differential current output with respect to differential voltages. All these current are collected at the summing nodes that is common to all the multipliers (synapses) and are converted to corresponding voltage signals which is then further supplied to neural activation function generator (NAF). In order to generate a activation function, in particular, sigmoid function, a simple differential amplifier well serves the purpose. In case of training the neuron model, we slightly modify the differential amplifier to generate the

derivative of the activation function. Finally we get the output of the artificial neuron as the action potential, the obtained response may not be the exact replica of that of biological neurons response, but it closely resembles the action potential with approximate amplitudes.

Now we shall discuss each block of the neuron model along with its analysis. The fundamental blocks of the proposed neuron model are gilbert cell multiplier, differential amplifier. The purpose of adder is served by just connecting the differential output nodes of the multiplier.

6.1 Gilbert Cell Multiplier:

Analog multipliers has wide range of applications in signal processing, neural networks, phase detectors, frequency doublers. Gilbert cell multiplier has two important properties, one is that the small signal gain of the circuit is the function of its tail current and the other is the two transistors in a differential pair provides a simple means of steering the tail current to one of its output destinations. In this part we are going to discuss how a gilbert cell can work as a multiplier along with its analysis and operation using Fig.6.2.

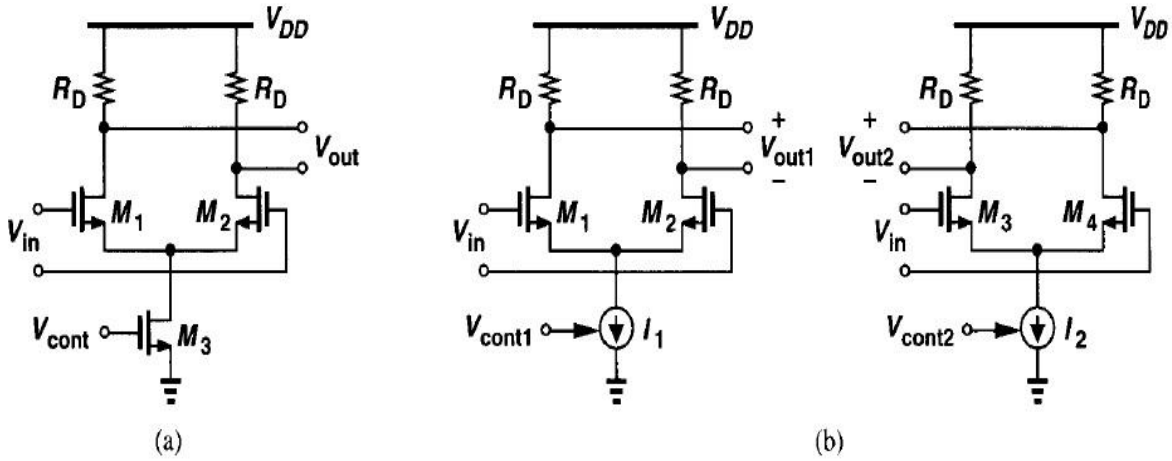


Fig.6.2: (a) simple variable gain amplifier, (b) variable gain provided by two stages

If suppose we need to construct a differential pair whose control voltage varies its gain, this can be accomplished by below circuit, where the control voltage defines the tail current and hence the gain. In that topology, the overall gain $A_v = V_{out}/V_{in}$ varies from zero to maximum value, i.e., if no current flows through the transistor M_3, given by the device dimensions and voltage headroom limitations. This circuit gives a simple example of a *variable gain amplifier* (VGA). Variable gain amplifiers find its application in systems

where the signal amplitude may experience large variations and hence requires inverse changes in the gain.

Now if suppose we need an amplifier whose gain can be continuously varied from a negative value to positive value. For that purpose, we need to consider two differential pairs that can amplify its inputs by opposite gains, which can depicted from the above figure. We now have $V_{out}/V_{in} = -g_m R_D$ and $V_{out}/V_{in} = +g_m R_D$, where g_m represents the transconductance of each and every transistor which is in equilibrium. If the currents through the transistors M_1 and M_2 vary in opposite directions, so do the gains V_{out1}/V_{in} and V_{out2}/V_{in}.

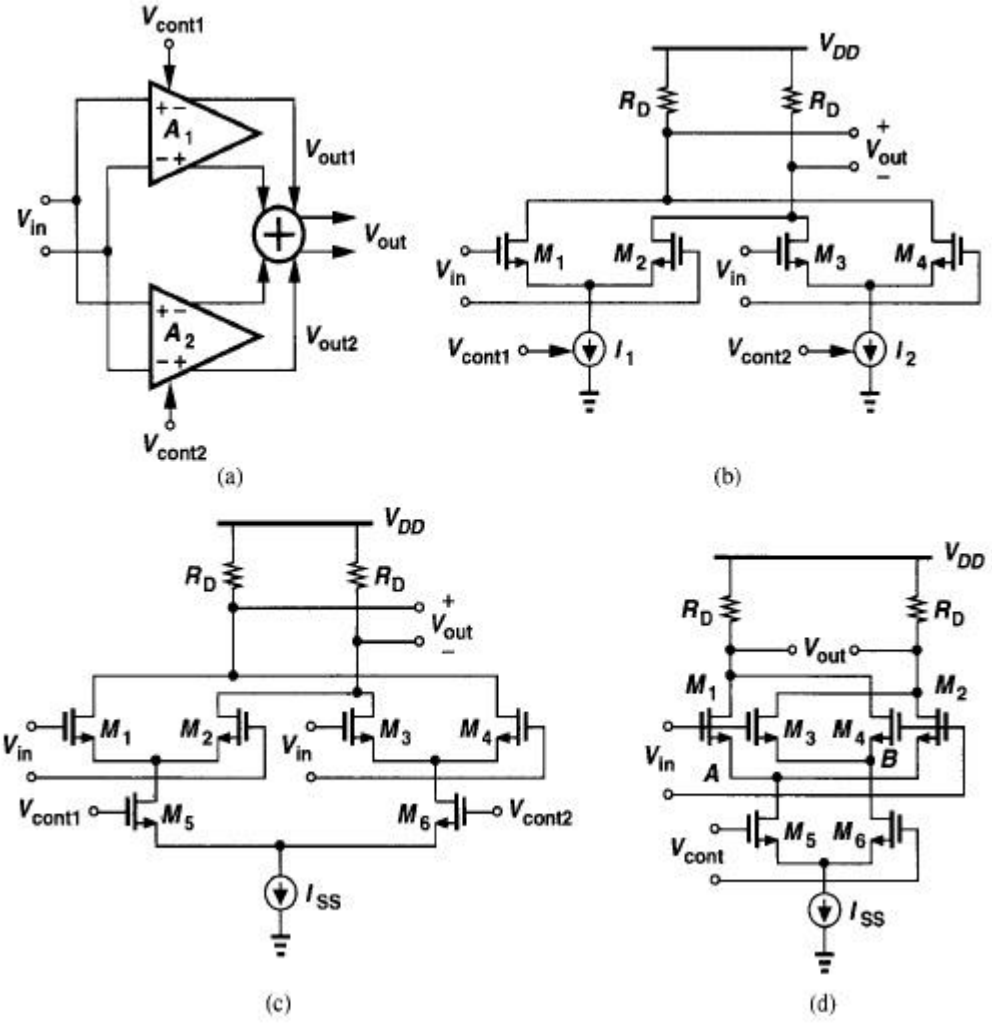


Fig.6.3: (a) summation of output voltages of two amplifiers, (b) summation in the current domain, (c) use of M_5 and M_6 to control the gain, (d) Gilbert cell.

But how the output voltages V_{out1} and V_{out2} can be combined into a single output? It can be accomplished by using below topology shown in Fig.6.3, where the two voltages can be

summed, resulting $V_{out} = V_{out1} + V_{out2} = A_1V_{in} + A_2V_{in}$, where the gains A_1 and A_2 are controlled by control voltages V_{cont1} and V_{cont2}, respectively. The actual implementation is quite simple, since $V_{out1} = R_DI_{D1} - R_DI_{D2}$ and $V_{out2} = R_DI_{D4} - R_DI_{D3}$, we have $V_{out} = V_{out1} + V_{out2} = R_D(I_{D1} + I_{D4}) - R_D(I_{D2} + I_{D3})$. Thus, rather than just adding V_{out1} and V_{out2}, we are simply shorting the drain terminals to sum the currents and subsequently generate the output voltage. If the current through the transistor M_1 ($I_1 = 0$), then $V_{out} = g_mR_DV_{in}$ and if the current through the transistor M_2 ($I_2 = 0$), then $V_{out} = -g_mR_DV_{in}$. For $I_1 = I_2$, the overall gain drops to zero. In the circuit shown in the above figure, V_{cont1} and V_{cont2} must vary the currents I_1 and I_2 in opposite directions such that the overall gain of the amplifier changes monotonically. We need to know what circuit can vary its currents in opposite directions. It can be accomplished by a differential pair, leading to the topology shown in the above figure. For a typically large value of $V_{cont1} - V_{cont2}$, all the tail current is steered to one of the top differential pairs, and the gain from input to the output will be at its most positive value or negative value. If $V_{cont1} = V_{cont2}$, then the overall gain will be zero. The circuit having such characteristic feature is known as *gilbert cell*, which is used in many analog and communication systems. While designing, the transistors $M_1 - M_4$ are identical and so are M_5 and M_6 and symmetry between them should be maintained.

When a gilbert cell is used in a cascade structure, it consumes a greater voltage headroom than a simple differential amplifier pair. The reason is that the two differential amplifiers $M_1 - M_2$ and $M_3 - M_4$ both are stacked on the top of the controlling differential transistor pair. In order to understand this point, suppose the differential input V_{in} has a common mode level V_{CMin}. As a result we have, $V_A = V_B = V_{CMin} - V_{gs1}$ where $M_1 - M_4$ are assumed to be identical transistors. For the transistors M_5 and M_6 to operate in saturation region, the common mode level of V_{cont}, V_{CMcont}, must be such that $V_{CMcont} \leq V_{CMin} - V_{gs1} + V_{th5,6}$. Since $V_{gs1} + V_{th5,6}$ is almost equal to one overdrive voltage, so we conclude that the control common mode level should be lower than the input common mode level at least by that value.

6.2 Gilbert Cell as Multiplier:

For the proper functioning of the circuit we need to maintain the transistors in saturation region, the condition for a transistor to be in saturation is $V_{ds} > V_{gs} - V_{tn}$ and $V_{gs} > V_{tn}$. Now we apply the input signal in the complementary form such that, $V_1 = -V_2$ and $V_{c1} = -V_{c2}$. By varying the input signal we obtain an output of the magnitude,

$$I_{out} = \frac{2\sqrt{2}R_d V_1 V_c k_n}{R_{out}}$$

$$I_{out} \propto V_1 V_c, \text{ where } k_1 = \frac{2\sqrt{2}R_d k_n}{R_{out}}$$

From the above equation we can say that the output current is the product of input voltages, by which we can say that the gilbert cell acts as a multiplier.

In the below shown circuit, we are providing a bias current of 30μA for its operation. The circuit is powered with a supplies voltage of 1.8V as per 180nm technology shown in Fig.6.4. For all the transistors to run in saturation their aspect ratios must be adjusted. The W/L ratios of the transistors for which they are operating in saturation are given in below table. The control voltages are actually used to adjust the overall gain of the circuit and the transistors with provided control voltages operate in linear region i.e., $V_{ds} < V_{gs} - V_{tn}$ such that they drive the transistor pairs above them into saturation. The symmetry of the circuit must be taken into consideration while adjusting the W/L ratios of the transistors shown in table 6.1.

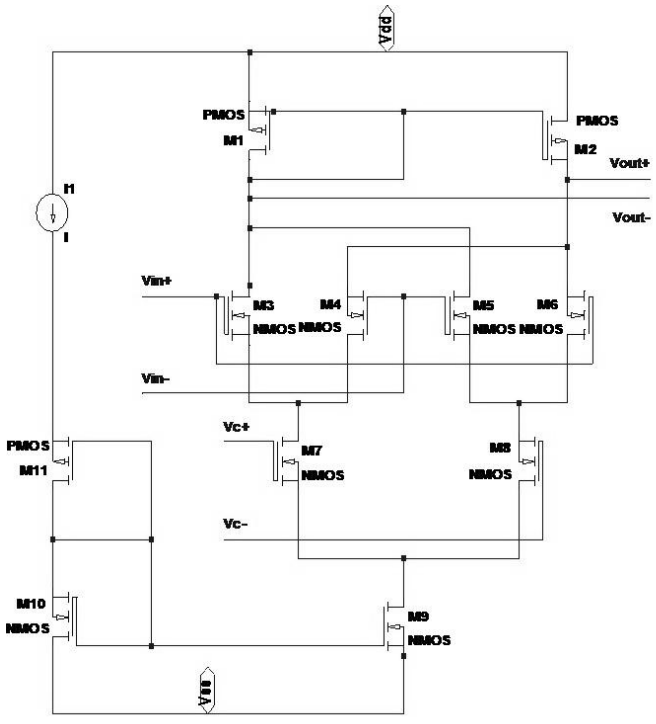


Fig.6.4: Gilbert cell as multiplier

We provide an input signal range of 100mV peak to peak, assuming that it is the output transmitted from other neurons whose ideal action potential is of 100mV range. Here the gilbert multiplier works as a synapse whose output will be in the transconductance mode i.e., current signal, which is to be converted into equivalent voltage signal and transferred

to the activation function generator. The expected waveforms of the gilbert cell multiplier shown in Fig.6.5, which is an amplitude modulated form of input signals, provided modulating signal and carrier signal frequencies as 50 KHz and 5 KHz respectively.

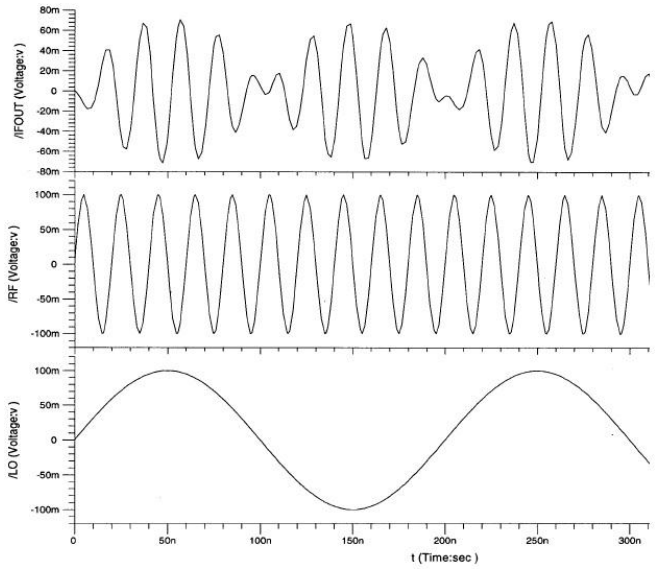


Fig.6.5: expected waveform of gilbert multiplier

Table 6.1: W/L ratios of multiplier

Transistors	W/L ratios
M_1, M_2	2µ
M_3, M_4, M_5, M_6	0.87µ
M_7, M_8	1µ
M_9, M_{10}	1µ
M_{11}	2µ

6.3 Neural Activation Function Generator:

Tan sigmoid is used as the neural activation function for the present design. In order to generate that function we just need a differential amplifier along with its modification for differentiation output. The same circuit should be capable of producing both the NAF and its derivative. In this work we consider two models of differential amplifier.

1. Differential amplifier as neural activation function generator
2. Modified differential amplifier as NAF along with differentiated output

Now we will discuss the functioning of differential amplifier and its analysis by considering the behavior of MOS differential pair as a function of differential input voltage.

To understand the exact functioning of the differential amplifier let us study the small signal behavior of differential transistor pairs. The Fig.6.7 shows the application of small signals

V_{in1} and V_{in2} and also assume both the transistors M_1 and M_2 are operating in saturation region. Now we have to calculate the differential voltage gain $V_{out}/(V_{in1} - V_{in2})$.

From large signal analysis we have the overall voltage gain as,

$$A_v = \sqrt{\mu_n C_{ox} I_{ss}\, W/L}\, R_D$$

At equilibrium, each transistor carries approximately equal current which is $I_{ss}/2$, this expression decomposes to $g_m R_D$, where g_m is the transconductance of transistors M_1 and M_2. We need to obtain the same gain through small signal analysis, so two different methods are employed with each method providing better insight into the circuits functioning.

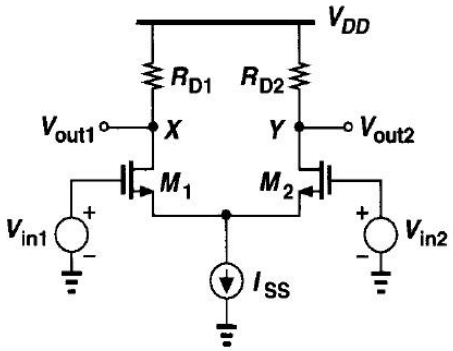


Fig.6.6: Differential pair with small signal inputs

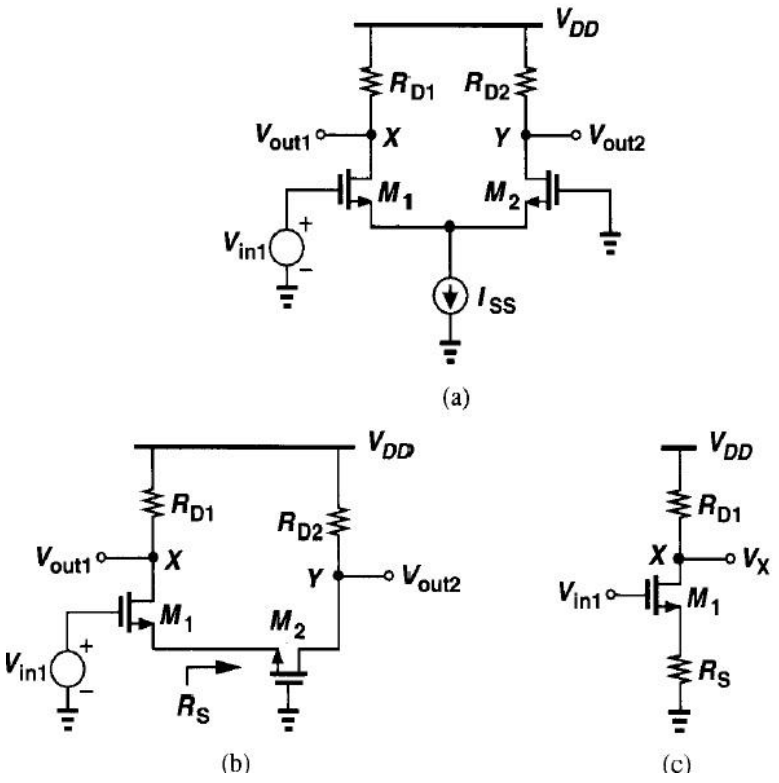


Fig.6.7: (a) differential pair sensing one input signal, (b) circuit viewed as common source stage degenerated by M_2, (c) equivalent circuit of (b)

Initially we assume the resistances $R_{D1} = R_{D2} = R_D$. To first method can be accomplished by driving the above differential pair with two independent signals and the output is computed as superposition. Let us assign V_{in2} to zero and we will find the effect of V_{in1} at X and Y. in order to obtain V_X, we note that the transistor M_1 forms a common source stage with a source degeneration resistance which is equal to the impedance seen through the source of transistor M_2. By neglecting the channel length modulation and the body effect, we have $R_s = 1/g_{m2}$ and

$$\frac{V_X}{V_{in1}} = \frac{-R_D}{\frac{1}{g_{m1}} + \frac{1}{g_{m2}}}$$

In order to calculate voltage at node Y, V_Y we note that the transistor M_1 drives M_2 as a source follower and we replace V_{in1} and M_1 by a thevenin equivalent shown in Fig.6.8.

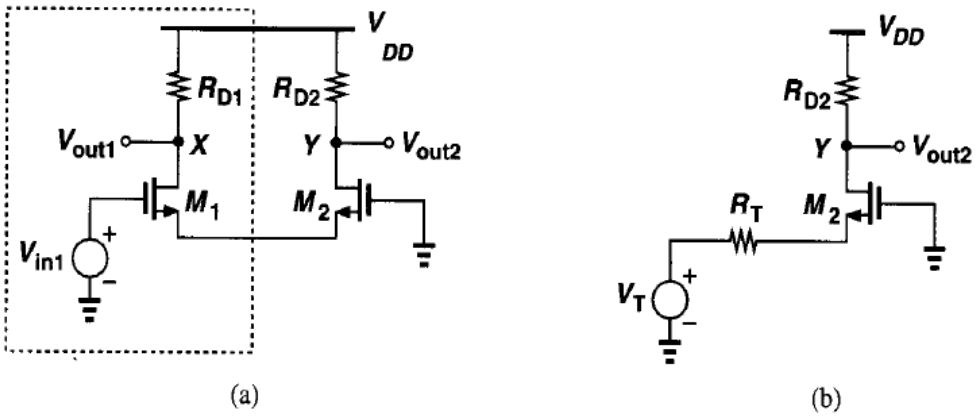


Fig.6.8: Replacing M_1 by thevenin equivalent

Here the thevenin voltage $V_t = V_{in1}$and the resistance $R_t = 1/g_m$, and the transistor M_2 will function as common source stage whose gain is equal to

$$\frac{V_Y}{V_{in1}} = \frac{R_D}{\frac{1}{g_{m1}} + \frac{1}{g_{m2}}}$$

From the above two equations, the overall voltage gain for V_{in1} is

$$(V_X - V_Y) = \frac{-2R_D}{\frac{1}{g_{m1}} + \frac{1}{g_{m2}}} V_{in1}$$

Which, for $g_{m1} = g_{m2} = g_m$ reduces to,

$$(V_X - V_Y) = -g_m R_D V_{in1}$$

Due to symmetry, the effect of the input voltage V_{in2} at nodes X and Y is identical to that of V_{in1} except that there will be change in the polarity.

$$(V_X - V_Y) = g_m R_D V_{in2}$$

For superposition we now add the obtained voltages, we get

$$\frac{(V_X - V_Y)}{V_{in1} - V_{in2}} = -g_m R_D$$

Comparison of the above obtained voltages indicates that the magnitude of differential gain is equal to $g_m R_D$ regardless of the type and how the inputs are applied. If the output is single ended then the overall gain is halved.

6.3.1 Differential pair with MOS loads:

Instead of resistors as load of differential amplifier, we can also employ diode connected and current source loads. For our work we are going to use a differential amplifier with such topology shown in Fig.6.9, so as to generate tan sigmoid function. The differential gain of such circuit can be obtained by the small signal analysis of half circuit.

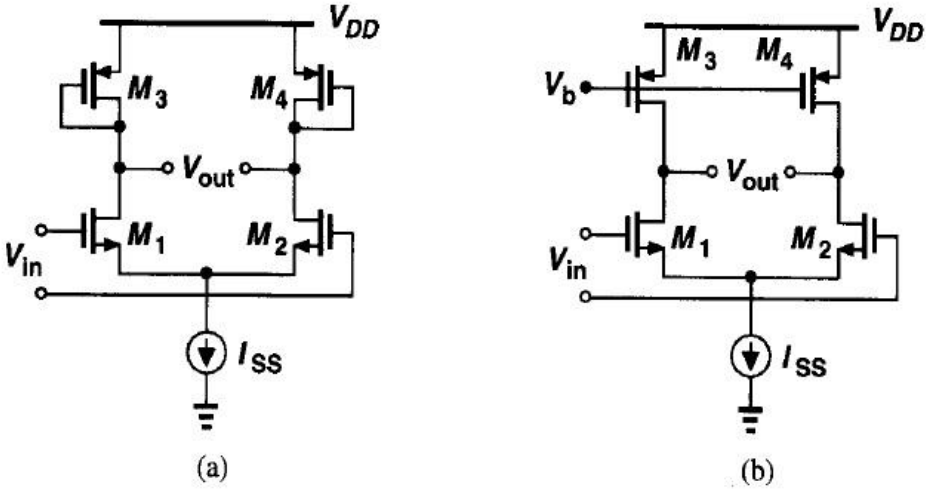


Fig.6.9: differential pair with (a) diode connected and (b) current source loads

For the differential amplifier with a diode connected load, the small signal circuit gain is given by

$$A_v = -g_{mN}(1/g_{mP} \| r_{oN} \| r_{oP}) \approx -\frac{g_{mN}}{g_{mP}}$$

Where the subscripts in the equation represents P and N in the equation represents the PMOS and NMOS transistors respectively. By expressing the transconductances g_{mN} and $-g_{mP}$ in terms of device dimensions, we have

$$A_v = \sqrt{\frac{\mu_n (W/L)_N}{\mu_P (W/L)_P}}$$

For the differential amplifier with a current source load, the small signal circuit gain is given by

$$A_v = -g_{mN}(r_{oN} \| r_{oP})$$

In the circuit shown in Fig.6.10, the diode connected loads of differential pair consumes severe voltage headroom, thus resulting in the tradeoffs between output voltage swings, overall voltage gain, and the input common mode range. For a given input bias current and the device dimensions of the circuits, the PMOS's over drive voltage and the overall gain of the circuit scales together. To achieve a higher gain, the aspect ratios of PMOS transistors have to be reduced and thereby increasing $V_{gsP} - V_{thP}$ and lowering the common mode levels of the input voltages at nodes X and Y.

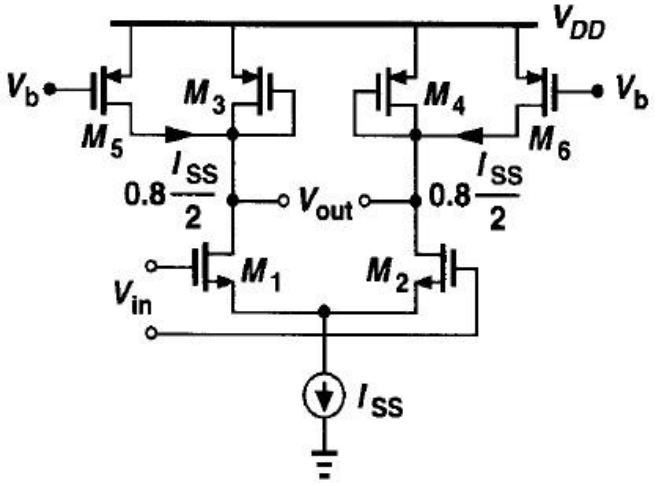


Fig.6.10: addition of extra current sources to hike the voltage gain

In order to overcome the above difficulty, part of the biasing currents of the transistors at the input can be provided by PMOS current sources. It is accomplished from the below figure whose transconductance g_m is reduced instead of aspect ratio of the transistors. If suppose the transistor pair M_5 and M_6 carries 80% of the drain current of the transistor pair M_1 and M_2, and the current through the transistors M_3 and M_4 is reduced by

a factor of 5. For a given $|V_{gsP} - V_{thP}|$, the transconductances of the transistors M_3 and M_4 are reduced as the aspect ratios of the transistors are also reduced by the same factor. Now the differential gain is approximately five times that of the situation with no PMOS current sources.

6.3.2 Modified Circuit:

By following all the above conditions we are using a modified circuit shown in Fig.6.11 instead of a simple differential pair. The external bias voltage bias voltage is used to drive the transistors in the sub threshold region and to improve the output swing of the circuit. This circuit generates tan sigmoid function as its output which can be observed as the circuit's DC characteristics.

This circuit can be further extended to generate the derivative of the tan sigmoid function which has its application while training a neural network. As we are concerned only with the action potential of the neuron, we don't go for its learning process. The expected DC characteristics of the below circuit is also shown below.

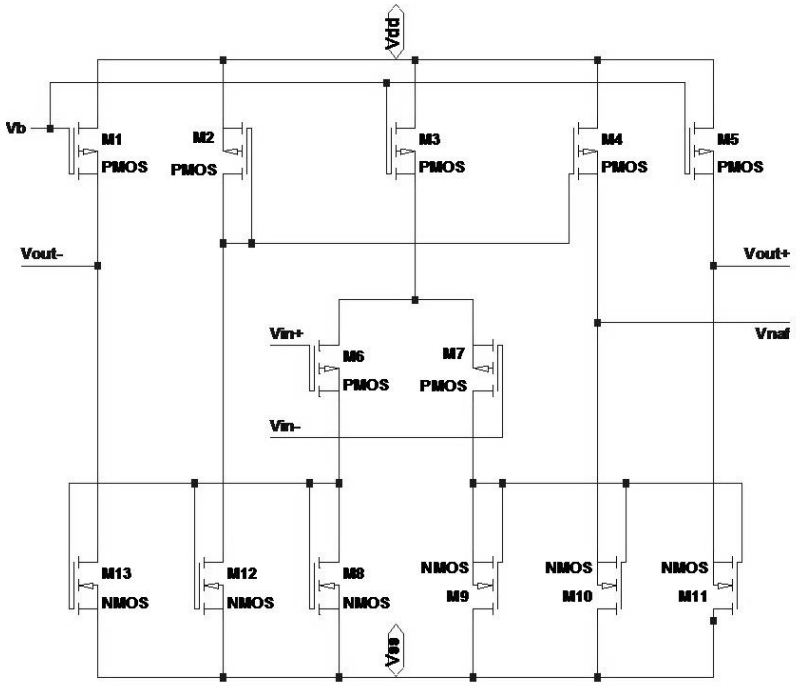


Fig.6.11: modified neural activation function generator

Table 2: W/L ratios of modified NAF generator

Transistors	W/L ratios
$M_1 - M_5$	2µ
M_6, M_7	1µ
$M_8 - M_{13}$	0.92µ

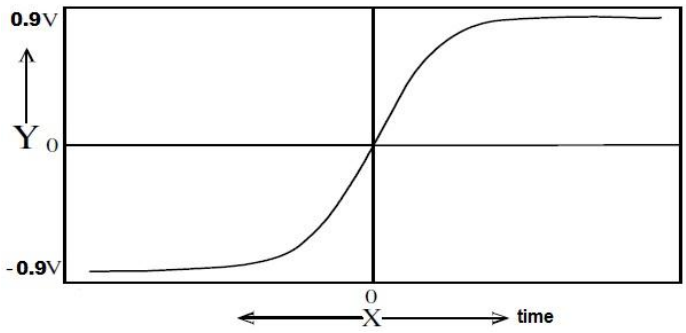


Fig.6.12: DC characteristics of modified circuit

The analog components that have implemented in the previous sections are used to build a neural architecture. The below figure shows the neural network architecture having three layers: input layer, hidden layer and the output layer. The simple neuron is a combination of two gilbert cell multipliers and a neural activation function generator.

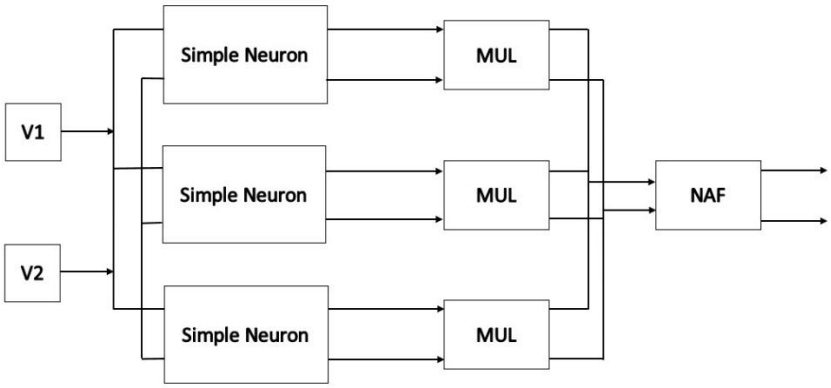


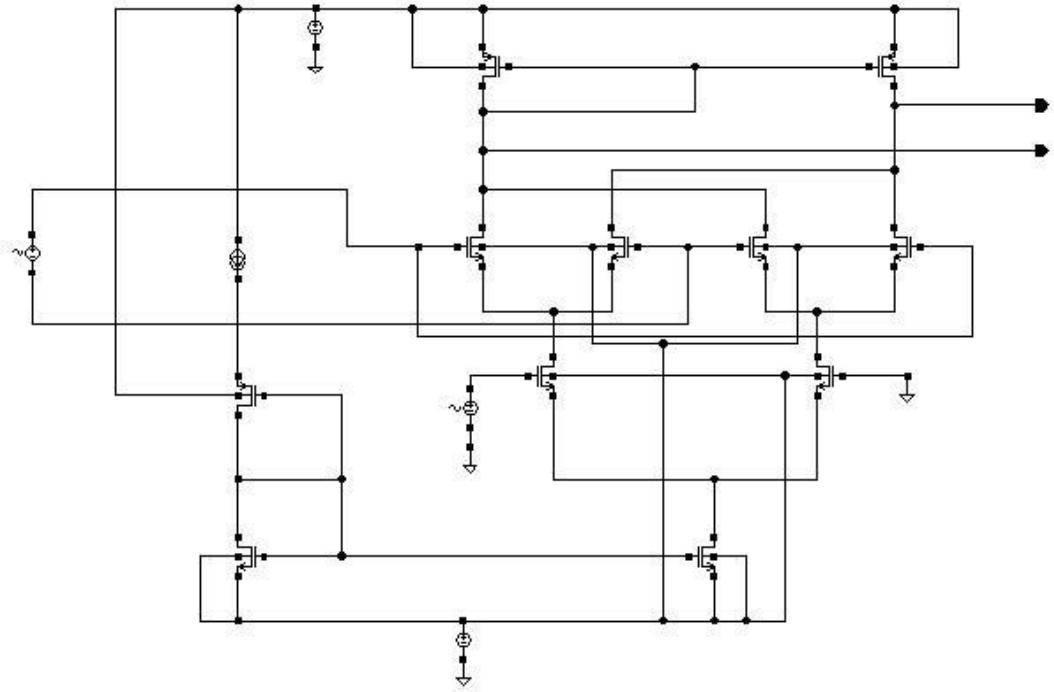
Fig.6.13: Neural architecture

We can consider the above shown neural architecture as a highly simplified model of the structure of biological neural network. The processing units in the architecture whose summing part input values, weights each value, and evaluates a weighted sum of those inputs. Here we have provided an input voltage of 100mV which would be the output of other neurons feeding this neuron architecture. The reason for applying that voltage is that every neuron triggers an action potential whose amplitude varies from -60mV to +40mV, so we approximately take a peak to peak voltage of 100mV as input. This collective functioning of artificial neurons closely resembles the dynamics of real neuron, whose output responses are shown in the next section. The power consumed by a single simple neuron is found to be 169μW and the overall power consumption of neural network architecture is 859.5μW.

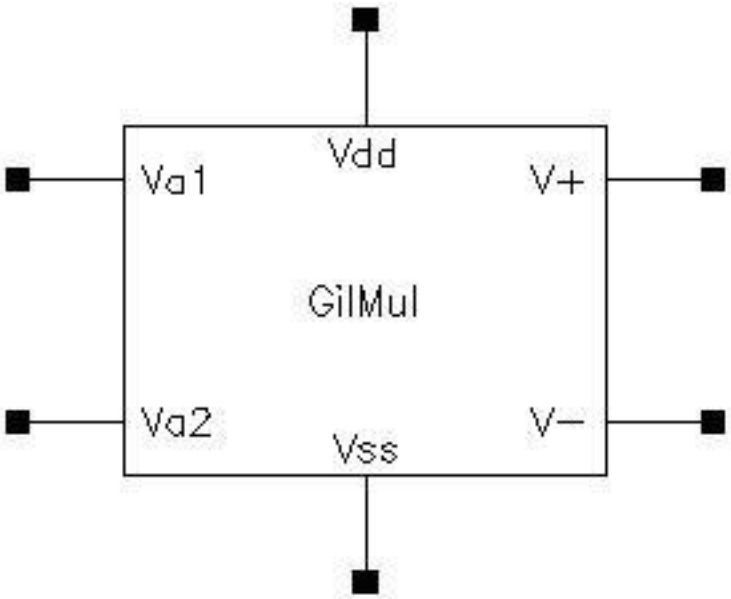
CHAPTER-7

RESULTS AND DISCUSSION

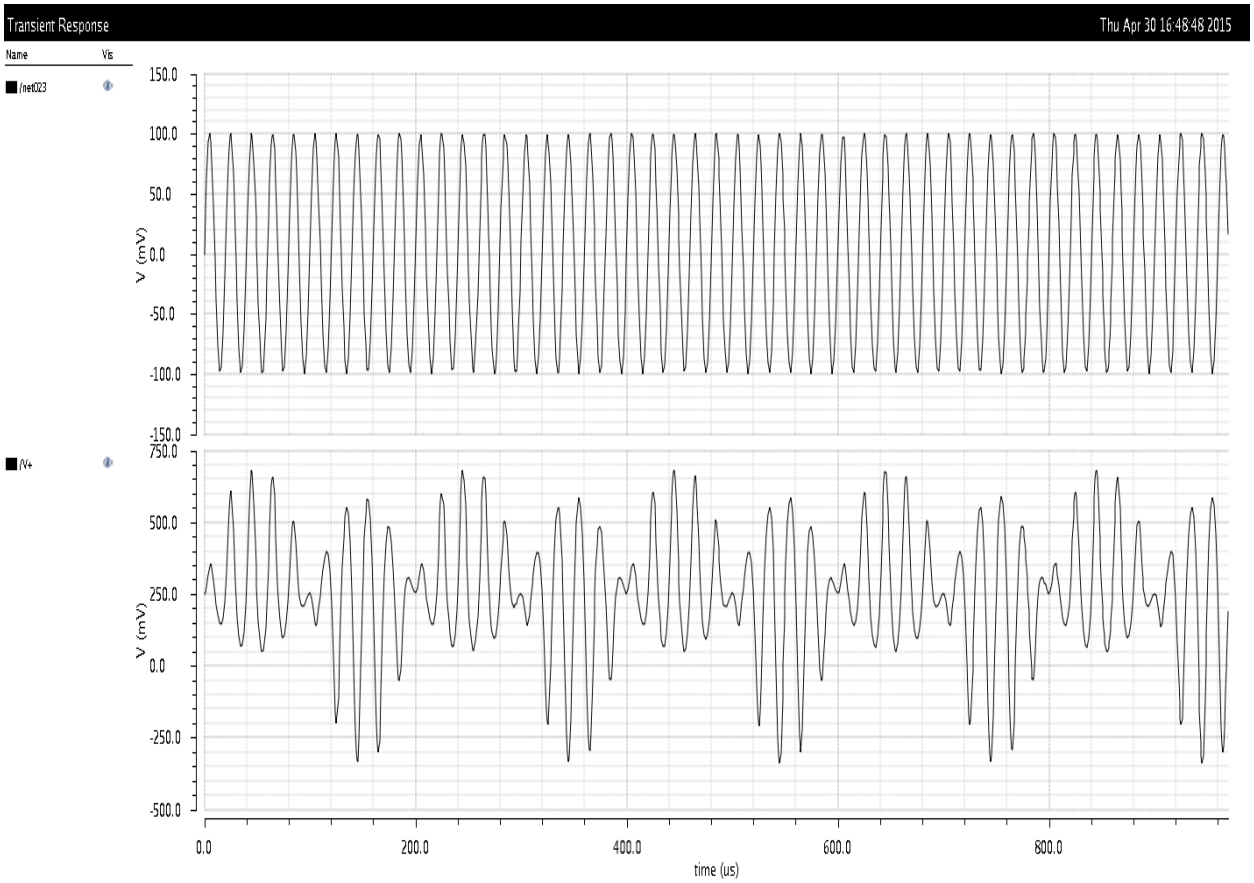
Schematic of gilbert cell multiplier:



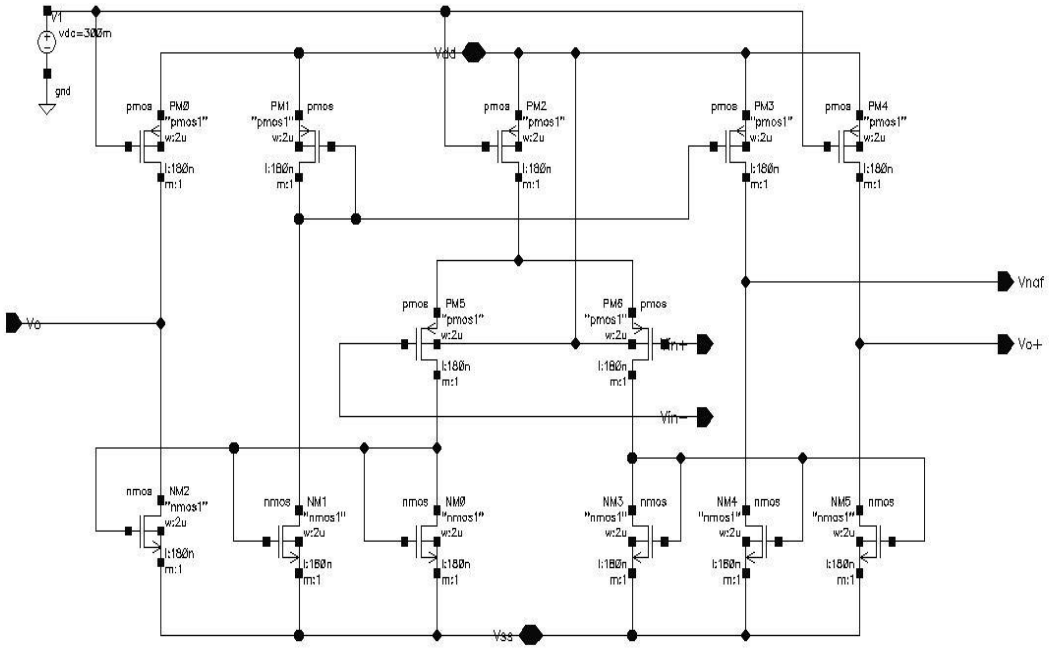
Symbol of gilbert multiplier:



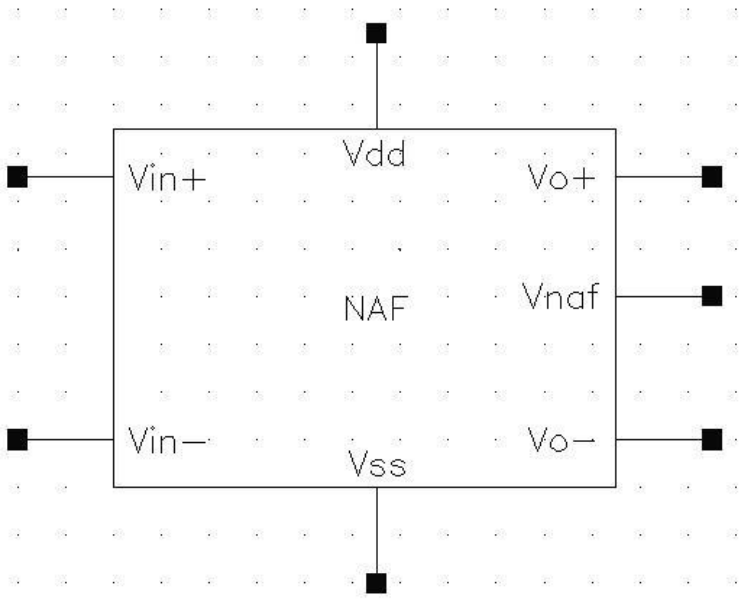
Transient response of multiplier:



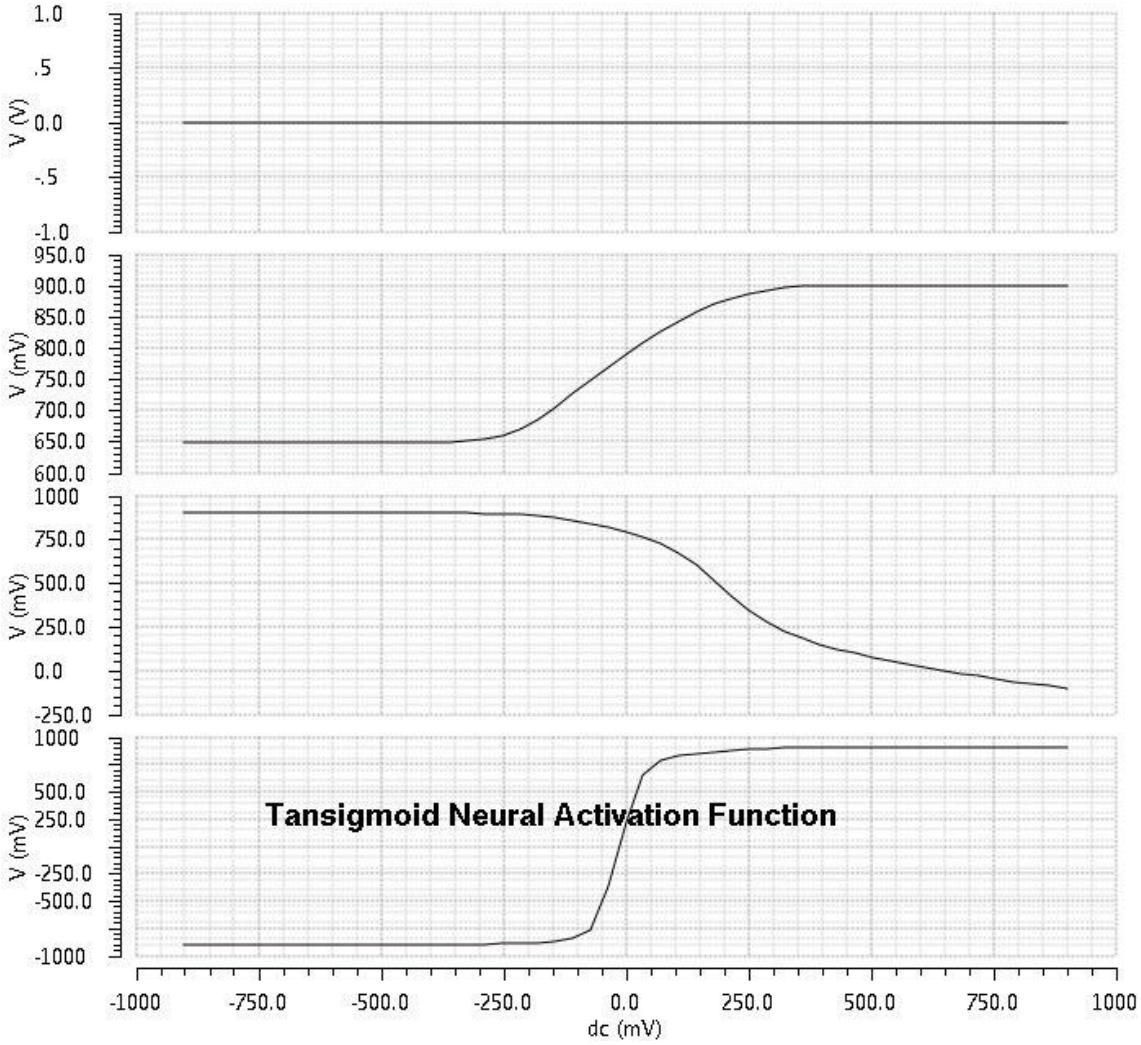
Schematic of modified activation function generator:



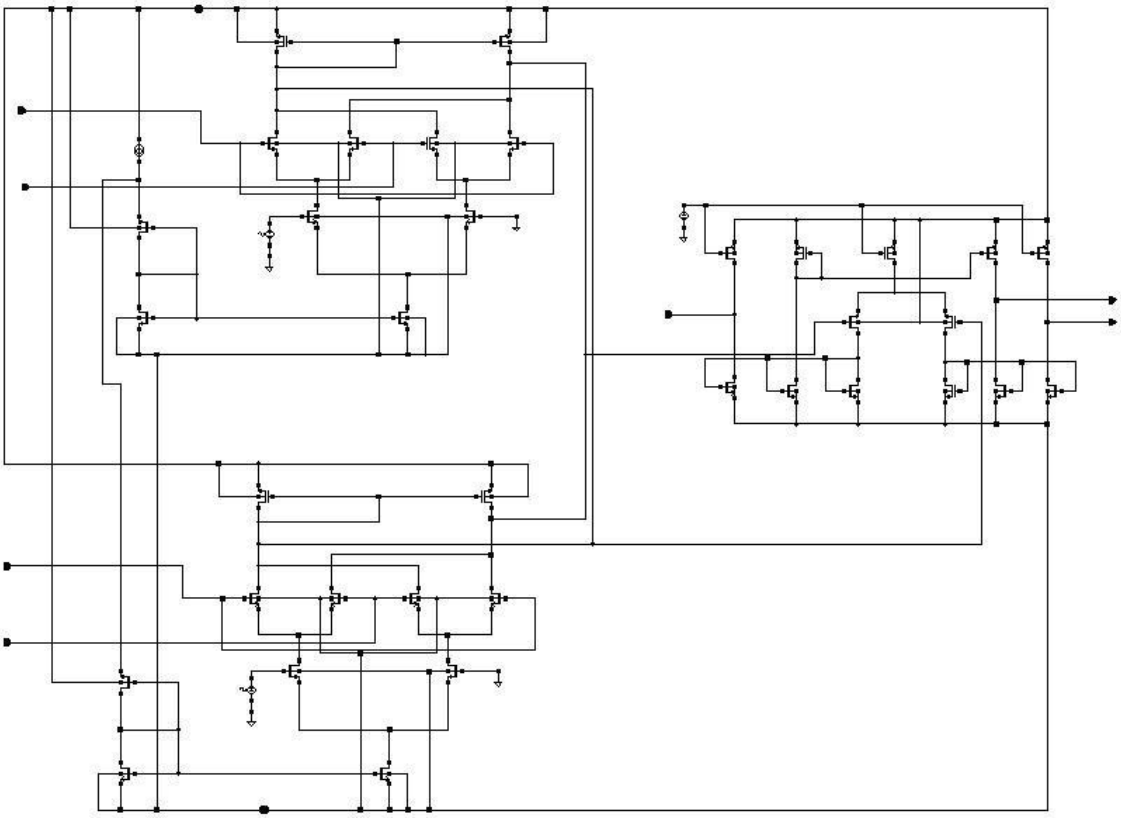
Symbol of NAF:



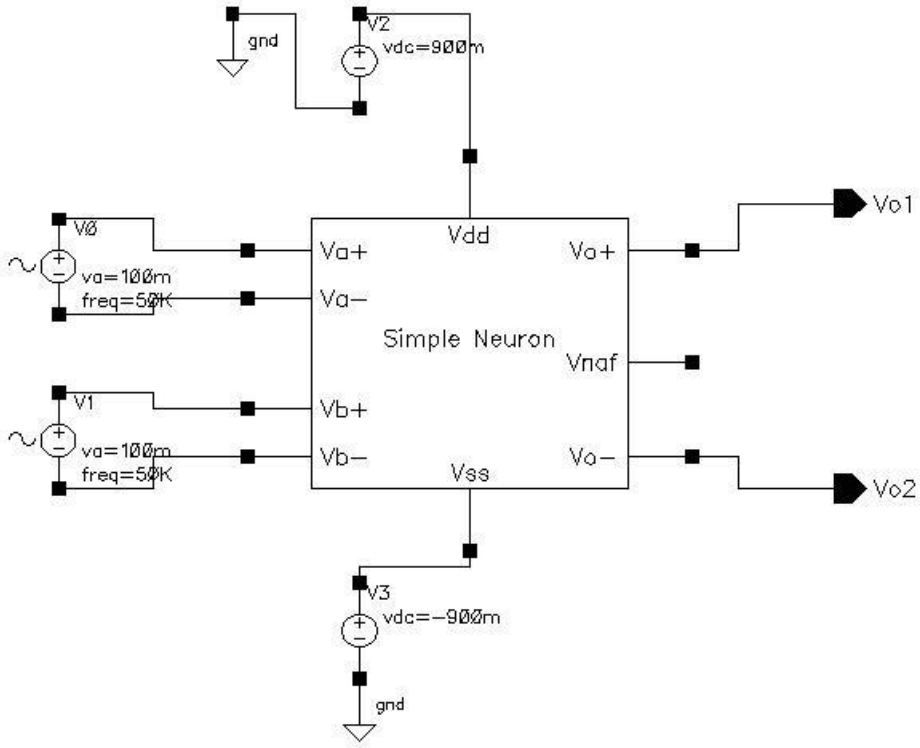
DC characteristics of NAF:



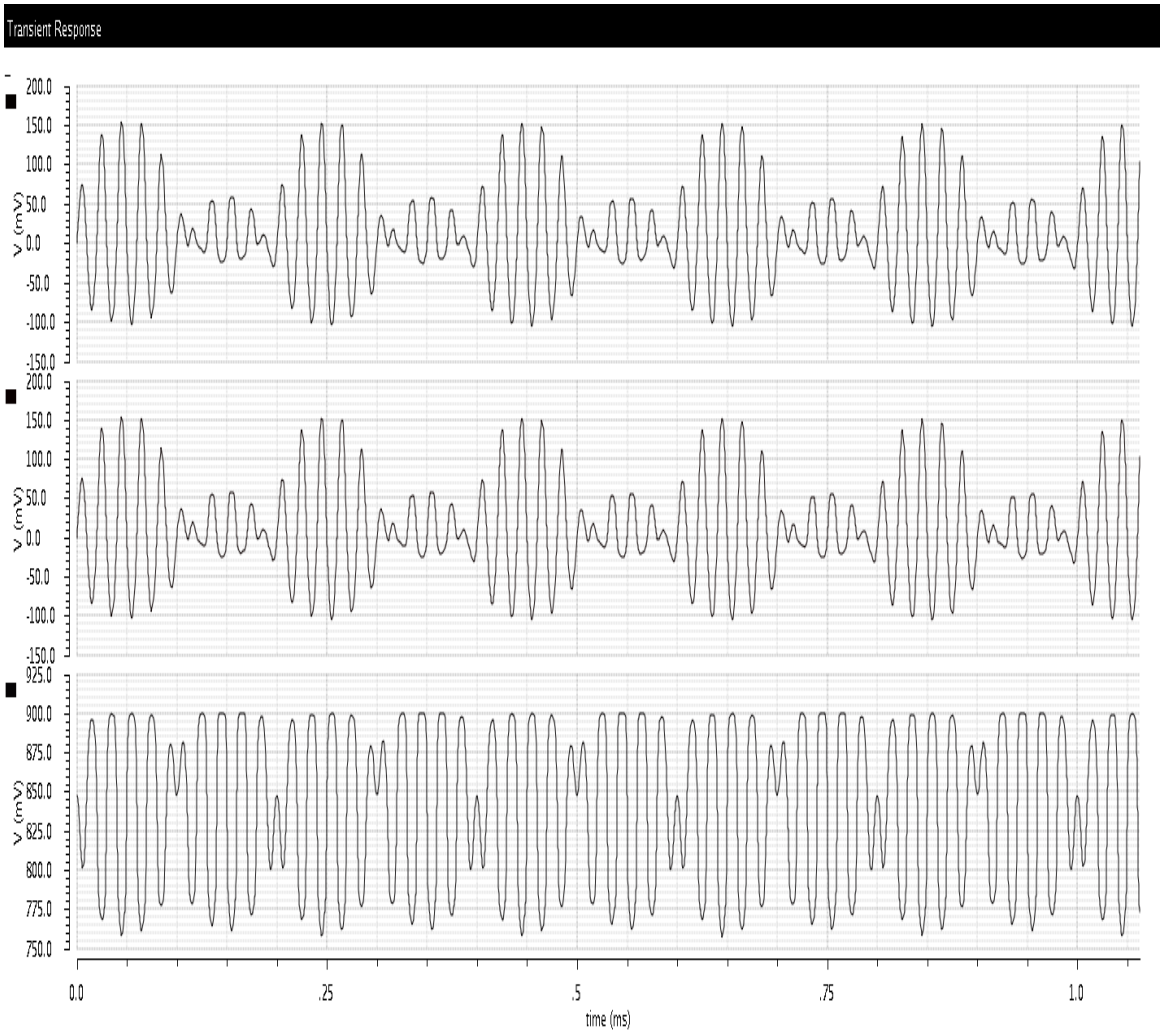
Schematic of simple neuron:

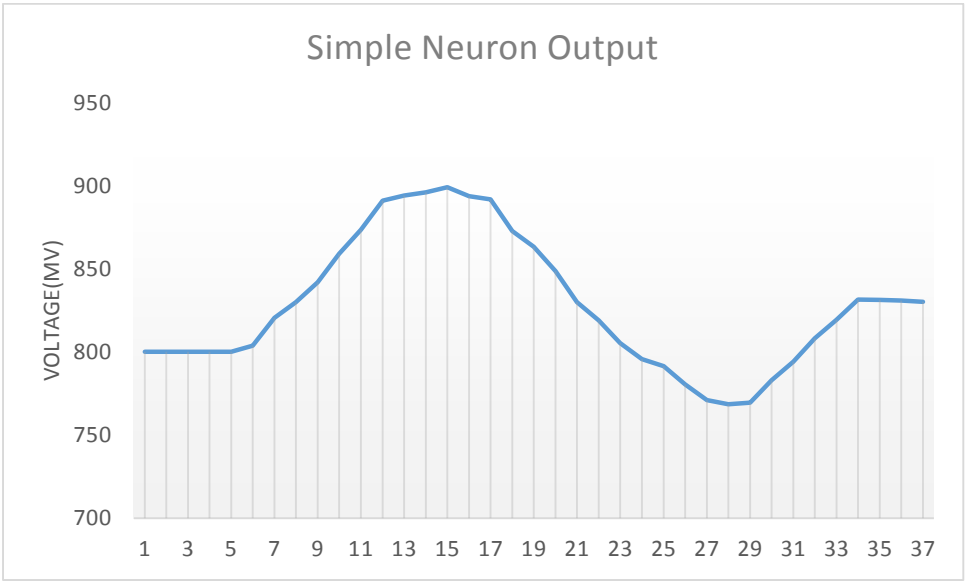


Symbol of simple neuron:

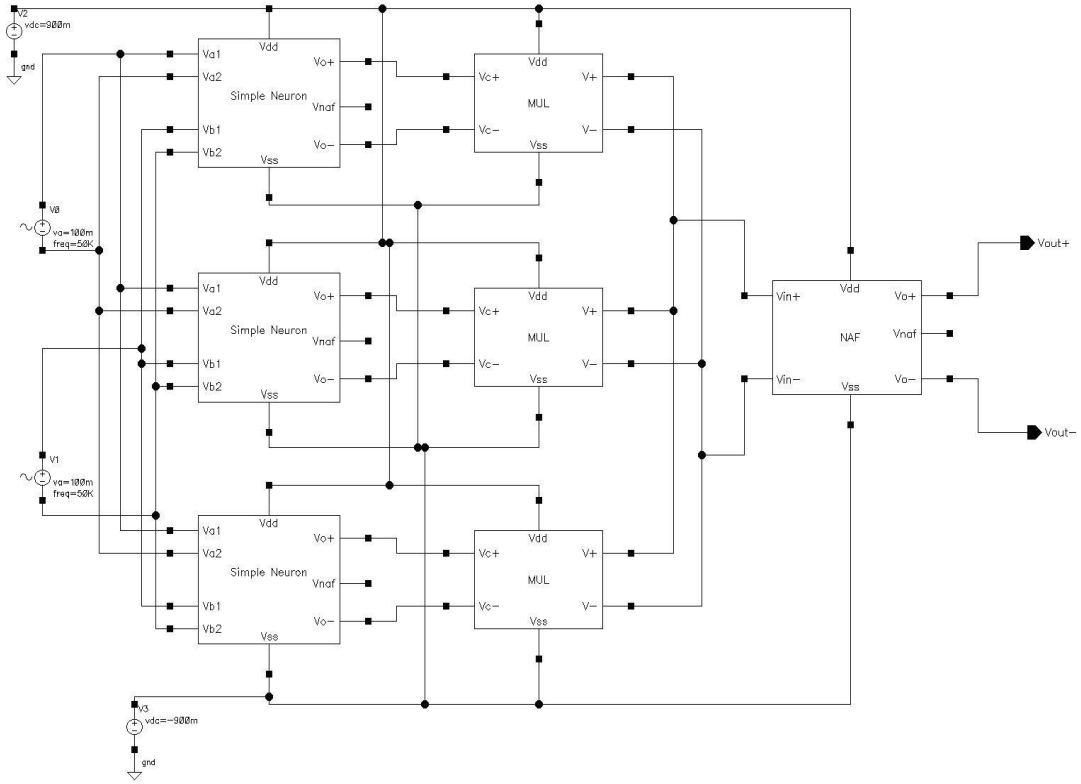


Output response of simple neuron:

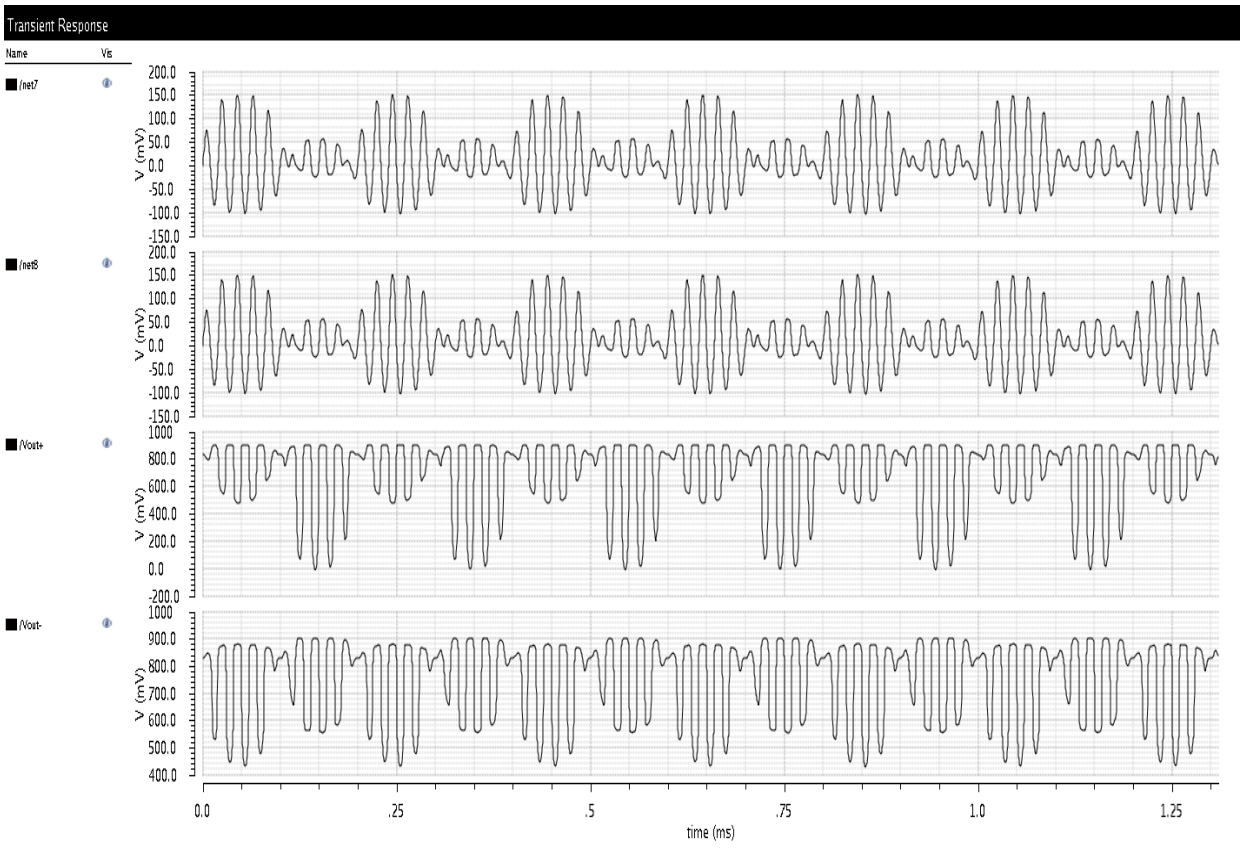




Neural architecture schematic:



Output response of neural architecture:



CHAPTER-8
CONCLUSION AND FUTURE SCOPE

The work demonstrates the VLSI implementation of artificial neural networks in 180nm CMOS technology node. Analog components such as gilbert multipliers, adders and differential amplifiers are employed to realize the biological neural network. The synapses are designed using multipliers along with the modified differential amplifier to generate tan sigmoid activation function. The modified circuit has an external bias which can be varied to enhance the output response of the circuit. The response of the proposed CMOS neuron is approximately equivalent to that of the real neuron. The inter connection of such CMOS neuron models has a wide variety of application in many fields out of which it has a bigger share in biomedical sciences.

The future technology is paving a better outlook in neuron-electron interface with the help of CMOS neural networks, which forms an interface to human brain. Now a days biomedical sciences are very much relying on the CMOS based micro-electronic devices. Further improvements and modifications in this work can lead to the development of generic technology which results in the bidirectional communication between the living neurons and the micro-electronic neurons. We can also refer them as neuromorphic chips which can replace the damaged neural networks of brain. By this we can restore the functionalities of the brain for example regaining of eye sight by connecting the nerve links through such neuromorphic chips.

REFERENCES

[1]. VLSI Implementation of Neural Networks B. M. Wilamowski, J. Binfet, and M. O. Kaynak.

[2]. Bulletin of the polish academy of sciences technical sciences Vol. 54, No. 4, 2006 Trans-resistance CMOS neuron for adaptive neural networks implemented in hardware by R.Wojtyna and T. Tala´Ska.

[3]. IEEE Journal Of Solid-State Circuits, Vol. 26, No. 7, July 1991 A CMOS Implementation Of Fitzhugh–Nagumo Neuron Model by Bernab Linares-Barranco, Student Member, IEEE, Edgar S6nchez-Sinencio, Senior Member, IEEE, Angel Rodriguez-Vazquez, Member, IEEE, And Jos6 L. Huertas, Member, IEEE.

[4]. IEEE Transactions On Neural Networks, Vol. 10, No. 3, May 1999 Analog Implementation Of Pulse-Coupled Neural Networks by Yasuhiro Ota And Bogdan M. Wilamowski, Senior Member, IEEE.

[5]. Analog Integrated Circuits and Signal Processing, 13, 111–121 (1997) VLSI Implementation of a Neural Model Using Spikes F. J. Pelayo, E. Ros1 X. Arreguit,2 A. Prieto.

[6]. A Low Power VLSI Implementation of the Izhikevich Neuron Model by A. Samil Demirkol, Serdar Ozoguz Istanbul Technical University, Department of Electronics and Communications Engineering.

[7]. Ota, Yasuhiro Wilamowski, Bogdan M. "Analog Implementation of Pulse-Coupled Neural Networks." IEEE Transactions on Neural Networks, May 1999 Issue.

[8]. II Song Han. "Biologically plausible VLSI neural network implementation with asynchronous neuron and spike-based synapse", Proceedings 2005 IEEE International Joint Conference on Neural Networks 2005.

[9]. “Analog VLSI Implementation of Neural Network Architecture for Signal Processing”, Neeraj Chasta, Sarita Chauhan and Yogesh Kumar.

[10]. E. Farquhar. "A Bio-Physically Inspired Silicon Neuron", IEEE Transactions on Circuits and Systems I Regular Papers, 3/2005.

[11]. B.M. Wilamowski. "VLSI architecture for analog bidirectional pulse-coupled neural networks", Proceedings of International Conference on Neural Networks (ICNN 97)ICNN-97, 1997

[12]. Raj P., Cyril Prasanna. "Design and Analog VLSI Implementation of Neural Network Architecture for Signal Processing", European Journal of Scientific Research/1450216X, 20090220

[13]. B.M. Wilamowski. "CMOS implementation of a pulse-coded neural network with a current controlled oscillator", 1996 IEEE International Symposium on Circuits and Systems Circuits and Systems Connecting the World ISCAS 96 ISCAS-96, 1996.

[14]. Manuel Carrasco-Robles. "A Novel Minimum- Size Activation Function and Its Derivative", IEEE Transactions on Circuits and Systems II Express Briefs, 2009.

[15]. Novel Circuit Realizations Of Neuron Activation Function And Its Derivative With Continuously Programmable Characteristics And Low Power Consumption V. Suresh Babu, M.R.Biju.

[16]. Developing A VLSI Architecture Simulation For Neural Networks Based Image Compression Prem Kumar, G.R.C.Kaladhara Sarma.

[17]. Realization of the Conscience Mechanism in CMOS Implementation of Winner-Takes-All Self-Organizing Neural Networks Rafał Długosz, Tomasz Talas´ka, Witold Pedrycz, Fellow, IEEE, and Ryszard Wojtyna.

[18]. MLP Neural Network Implementation and Integration in CMOS Technology D. Roviras, H. Abdulkader, H. Tap-Beteille, F. Castanie, M.

[19]. Analysis of Integrate-and-Fire Neurons: Synchronization of Synaptic Input and Spike Output A. N. Burkitt G. M. Clark Bionic Ear Institute, East Melbourne, Victoria 3002, Australia.

[20]. Hardware Implementation of CNN, S.Veni, B.Yamuna, Proceedings of lCISIP – 2005.

[21]. Analog VLSI Design of Neural Network Architecture for Implementation of Forward Only Computation AdoniArun, Harshvardhan Srivastava, SubhaMada, M.B.Srinivas 2012 Asia Pacific Conference on Postgraduate Research in Microelectronics & Electronics (PRIMEASIA).