Design of ANN based Handover Chip Using VHDL

A Thesis

Submitted towards the partial fulfillment of the requirements for the award of the degree of

Master of Engineering
In
Electronics and Communication

by

Chhavi Dhawan
Regn. No. 8024104

Under the guidance of

Mr. R.K Khanna
Assistant Professor
Department of Electronics &
Communication Engineering
T.I.E.T, Patiala

Prof. Shakti Kumar
Additional Director
Haryana Engineering College
Jagadhri

Department of Electronics & Communication Engineering

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
(Deemed University)
PATIALA-147004

June 2004

Certificate

This is to certify that the thesis titled, **" Design of ANN based Handover Chip using VHDL"**, submitted by **Ms. Chhavi Dhawan (Regn. No. 8024104)** to **Thapar Institute of Engineering and Technology (Deemed University), Patiala (Punjab)** towards the partial fulfillment of the requirements for the award of the degree of **Master of Engineering in Electronics and Communication Engineering,** is a bonafide work carried out by her under our guidance and supervision and no part of this thesis has been submitted to any other university or institute for award of any degree.

Mr. R. K. Khanna

Asst. Professor
Department of Electronics &
Communication Engineering
T.I.E.T.
(Deemed University)
Patiala. 147004

Prof. Shakti Kumar

Additional Director

Haryana Engineering College
Jagadhri, Yamunanagar
(Kurukshetra University)

(Countersigned by)

Dr. R.S Kaler
Professor & Head
Department of Electronics &
Communication Engineering
T.I.E.T.
Patiala. 147004

Dr. D.S. Bawa
Dean, Academic Affairs
T.I.E.T.
Patiala. 147004

Acknowledgement

The real sprit of achieving a goal is through the way of excellence and austerous discipline. I would have never succeeded in completing my task without the cooperation, encouragement and help provided to me by various personalities.

Moreover, a thesis is never the work of an individual; it is more a combination of ideas, suggestions, contribution and above all the proper guidance. One of the most important tasks of writing a thesis is to extend my deep sense of respect and sincere appreciation for the kind help, support and guidance extended to me by my supervisors **Mr. Rajesh Khanna,** Assistant Professor, Electronics and Communication Engineering Department, Thapar Institute of Engineering and Technology, Patiala, Punjab and **Prof. Shakti Kumar**, Additional Director, Haryana Engineering College, Jagadhri, Haryana. Without their invaluable suggestions, constructed ideas, much needed diligence and encouragement and their friendly attitude this mission would have never been successful.

With the feeling of high esteem and deep reverence, I wish to express my profound gratitude to **Dr. R.S. Kaler**, Professor and Head, Department of Electronics and Communication Engineering, for extending computer facility, advice and moral support at every step.

I also offer my sincere thanks to all faculties, teaching and non-teaching, especially **Mr. Sanjay Batish** and **Ms. Sushma Jain** of VLSI Department and **Mr. Amit Bhardwaj,** Electronics & Communication Engineering Department, TIET, Patiala; and **Ms. Savita Wadhawan** of Haryana Engineering College, Jagadhri; and the staff of central library, TIET, Patiala for their assistance.

I am also thankful to the **authors** whose works I have consulted and quoted in this work.

Last, but not the least, very special thanks are due to my **Family** and **Friends** for their constant love and support. Their patience, understanding and encouragement without which this study would not have been in its present form, is greatly appreciated. Above all I render my gratitude to the **ALMIGHTY** who bestowed self-confidence, ability and strength in me to complete this work.

Chhavi Dhawan

Place: T.I.E.T. Patiala, INDIA

(CHHAVI DHAWAN)

Abstract

In order to meet the escalating requirements for wireless communication services, smaller cells are deployed to boost the traffic capacity in limited spectrum allocation. But with the decrease in cell size, the number of cell boundary crossings by the user i.e. Mobile Station (MS) also increase. Consequently, the number of handoffs, process that transfers an ongoing call from one cell to another, also increases rapidly. These handoffs should be imperceptible to the users and also the quality of service (QOS) should be maintained throughout. The performance of the cellular systems is heavily influenced by the handover schemes.

Thus, we may observe the complexity associated with the handoff decision. It is indeed, a complicated task to predict accurately the need of handoff and also to execute it efficiently. Hence, there exists a need to develop an intelligent system that would predict the requirement of handoff to a much greater degree of accuracy. Also, this system should be capable to perform the handoff at a faster rate than the existing ones. This intelligence can be given to the system using one of the emerging technologies such as fuzzy logic, neural networks, particle swarm optimization etc. In this work, we have used the Neural Networks, which have a far reaching potential as building blocks in tomorrow's computational world.

In this thesis, the hardware implementation of neural networks based on digital components is done. A generalized Neural Network is designed in a Hardware Description Language i.e. VHDL, which enables us to get the entire design on a chip. Perhaps the most promising approach in emulating neural models digitally is the Field Programmable Gate Array (FPGA). Hence, the design is chosen to be implemented on an FPGA.

The design of handoff circuit can be converted to a parallel processing system that can monitor all the MSs being served in the cell. Being parallel in nature the processing speed goes up tremendously thereby, improving the quality of service (QoS). Also the circuit can be embedded in MS leading to Mobile Controlled Handoff.

Table of Contents

List of Abbreviations

ANN	Artificial Neural Network
ASIC	Application Specific Integrated Circuits
AUC	Authentication Center
BER	Bit Error Rate
BSS	Base Station Subsystem
BSCs	Base Station Controllers
BTSs	Base Transceiver Stations
CAD	Computer Aided Design
CDMA	Code Division Multiple Access
DSSS	Direct Sequence Spread Spectrum
EDA	Electronic Design Automation
EIR	Equipment Identity Register
FDMA	Frequency Division Multiple Access
FPD	Field Programmable Device
FPGA	Field Programmable Gate Array
GSM	Global System for Mobile communication
1G	First Generation
2G	Second Generation
GMSC	Gateway Mobile services Switching Center
GIWU	GSM Inter-Working Unit
HO	Handoff
HHO	Hard Handoff
HLR	Home Location Register
IS	Interim Standard
LOS	Line of Sight
MS	Mobile Station
MTSO	Mobile Telephone Switching Office
MSC	Mobile services Switching Center
MSC	Mobile Switching Center
MSN	Mobile Service Node

Mk RQ	Make Request
Mk HO	Make handoff
NLOS	Non Line of Sight
NO HO	No Handoff
OSS	Operation Support Subsystem
PN	Pseudorandom noise
QOS	Quality of Service
RF	Radio Frequency
RSS	Received Signal Strength
SER	Symbol Error Rate
SHO	Soft Handoff
SIR	Signal-to-Interference Ratio
SS	Switching Subsystem
TDMA	Time Division Multiple Access
VHDL	VHSIC Hardware Description Language
VHSIC	Very High Speed Integrated Circuits
VLR	Visitor Location Register

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, the Global System for Mobile communication (GSM) standard is discussed briefly with the emphasis on the Handoff process associated with it. Further, the objectives and the organization of the thesis are also included.

1.1 Cellular Mobile Communication: The User's Choice

Cellular mobile communication is one of the fastest growing and most demanding telecommunications applications. Currently there are more than 45 million cellular subscribers worldwide. It is expected that cellular systems using a digital technology will become the universal method of telecommunications. By the year 2005, experts predict that there will be more than 100 million cellular subscribers worldwide as can be seen in Figure 1.1 [23].

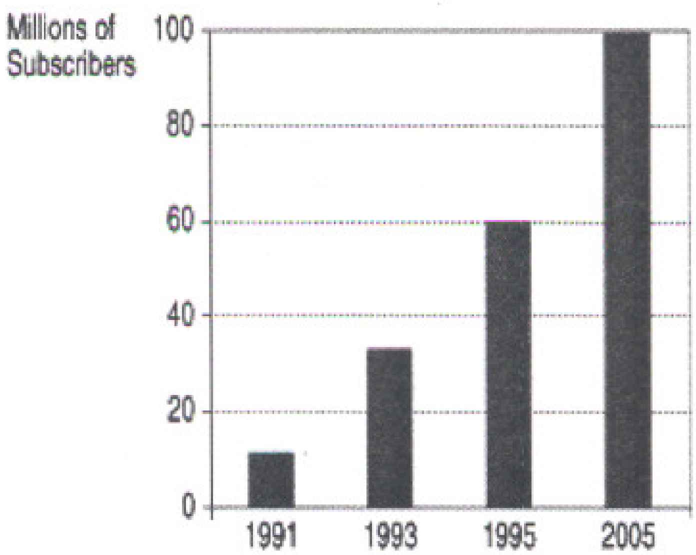


Figure 1.1 Cellular Subscriber Growth Worldwide

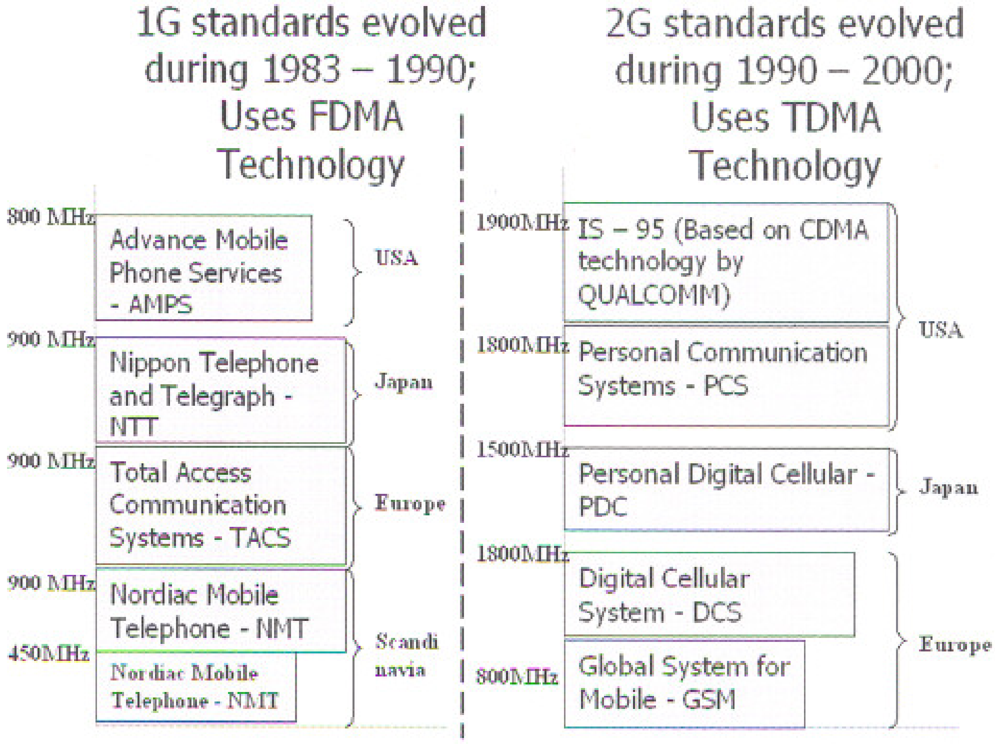
The objective of this chapter is to introduce the architecture of GSM network before the related problem is formulated.

This chapter is divided into 6 sections. Section 2 introduces GSM networks. Section 3 introduces the handoff in GSM networks. Section 4 presents the objectives of this work and chapter 5 highlights the organization of this thesis.

1.2 The GSM Network

Since the evolution of cellular mobile communications, two major standards i.e. first generation (1G) and second generation (2G) have contributed maximum to the present day society. The technology and the operating frequencies of these standards are given in table 1.1.

Table 1.1 Evolutions of 1G and 2G Standards



The GSM standard is one of the most widely used wireless technologies, which enables the users to roam across most of the world. It has some special features like facilitating the users to use the same number everywhere, providing facility to the operators to buy the requisite equipments from different suppliers etc. that makes it a successful standard.

The GSM network is divided into three major systems: the Base Station System (BSS), the Switching System (SS), and the Operation Support System (OSS) [1]. The basic GSM network elements are shown in Figure 1.2 [23] and their brief introduction is given in the following sections.

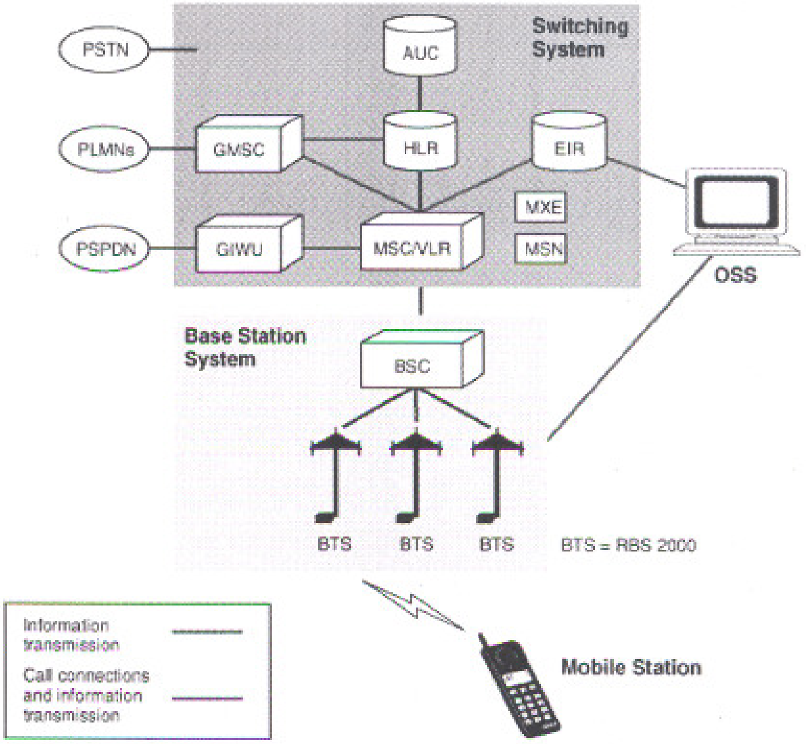


Figure 1.2 Basic GSM Network

1.2.1 Base Station Subsystem (BSS)

All radio-related functions are performed in the BSS, which consists of Base Station Controllers (BSCs) and the Base Transceiver Stations (BTSs). The BTSs and their controlling BSC are often collectively referred to as the BSS. Since the cellular topology of the network is a result of limited radio spectrum so, in order to use the radio spectrum efficiently, the same frequencies are reused in nonadjacent cells. The geographic region is divided up into cells and each cell has a BTS that transmits data via a radio link to MSs within the cell. A group of BTSs are connected to a BSC that are further grouped together and connected to a Mobile Switching Center (MSC) via microwave links or telephone lines. The MSC connects to the public switched telephone network, which switches calls to other mobile stations or land-based telephones [2]. Thus, we can say that:

- **BSC:** The BSC provides all the control functions and physical links between the MSC and BTS. It is a high-capacity switch that provides functions such as handover, cell configuration data, and control of Radio Frequency (RF) power levels in base transceiver stations. A number of BSCs are served by an MSC.
- **BTS:** The BTS handles the radio interface to the mobile station. The BTS is the radio equipment (transceivers and antennas) needed to service each cell in the network. A group of BTSs are controlled by a BSC.

However, the Figure 1.3 shows how these components of BSS are linked together and also describes its linking criteria with the MSCs and PSTNs.

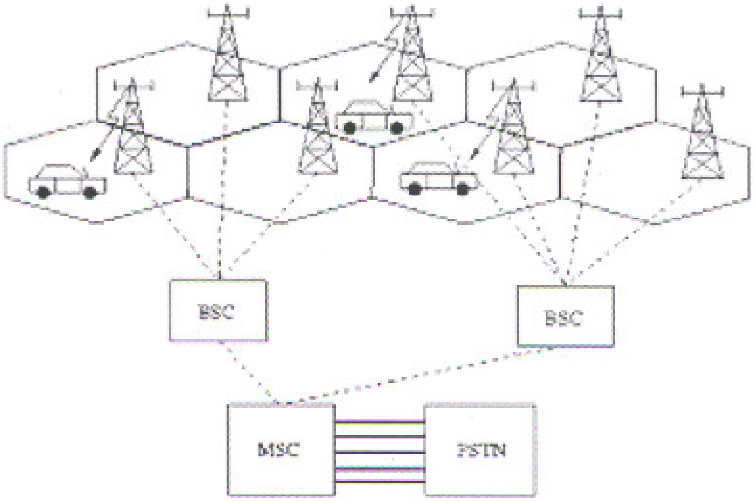


Figure 1.3 Detail Cellular System

1.2.2 Switching Subsystem (SS)

The switching system (SS) is responsible for performing call processing and subscriber-related functions [1]. The switching system includes the following functional units.

- **Home Location Register (HLR):** The HLR is a database used for storage and management of subscriptions. The HLR is considered the most important database, as it stores permanent data about subscribers, including a subscriber's service profile, location information, and activity status. When an individual buys a subscription from one of the PCS operators, he or she is registered in the HLR of that operator.
- **Mobile services Switching Center (MSC):** The MSC performs the telephony switching functions of the system. It controls calls to and from other telephone and data systems. It also performs such functions as toll ticketing, network interfacing, common channel signaling, and others.
- **Visitor Location Register (VLR):** The VLR is a database that contains temporary information about subscribers that is needed by the MSC in order to service visiting subscribers. The VLR is always integrated with the MSC. When a mobile station roams into a new MSC area, the VLR connected to that MSC would request data about the mobile station from the HLR. Later, if the mobile

station makes a call, the VLR will have the information needed for call setup without having to interrogate the HLR each time.

- **Authentication Center (AUC):** A unit called the AUC provides authentication and encryption parameters that verify the user's identity and ensure the confidentiality of each call. The AUC protects network operators from different types of fraud found in today's cellular world.
- **Equipment Identity Register (EIR):** The EIR is a database that contains information about the identity of mobile equipment that prevents calls from stolen, unauthorized, or defective mobile stations. The AUC and EIR are implemented as stand-alone nodes or as a combined AUC/EIR node.

The figure 1.4 illustrates switching subsystem.

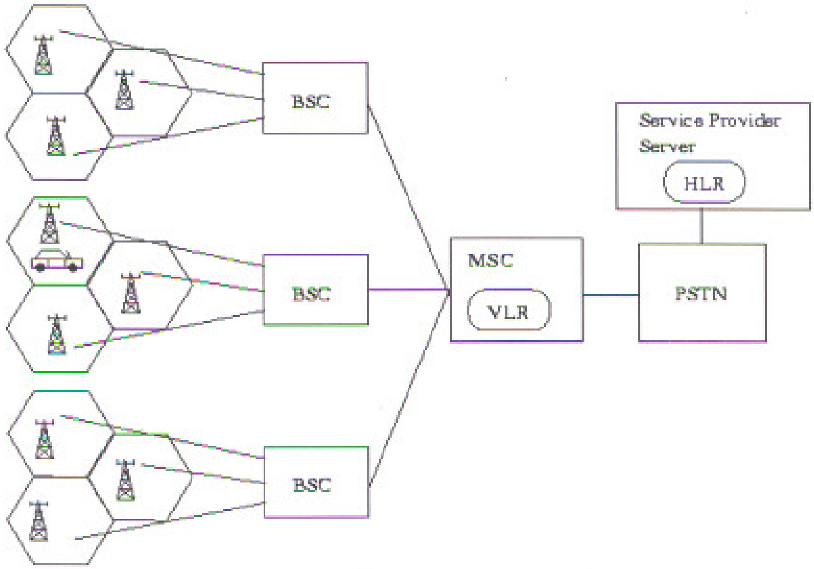


Figure 1.4 Detail Switching Subsystem

1.2.3 Operation and Support Subsystem (OSS)

The operations and maintenance center (OMC) is connected to all equipment in the switching system and to the BSC. The implementation of OMC is called the operation and support subsystem (OSS). The OSS is the functional entity from which the network operator monitors and controls the system. The purpose of OSS is to offer the customer cost-effective support for centralized, regional, and local operational and maintenance activities that are required for a GSM network. An important function of OSS is to provide a network overview and support the maintenance activities of different operation and maintenance organizations.

1.2.4 Additional Functional Elements

Other functional elements shown *in* Figure 1.2 are as follows:

- **Message Center (MXE):** The MXE is a node that provides integrated voice, fax, and data messaging. Specifically, the MXE handles short message service, cell broadcast, voice mail, and fax mail, e-mail, and notification.
- **Mobile Service Node (MSN):** The MSN is the node that handles the mobile intelligent network (IN) services.
- **Gateway Mobile services Switching Center (GMSC):** A gateway is a node used to interconnect two networks. The gateway is often implemented in an MSC. The MSC is then referred to as the GMSC.
- **GSM Inter-Working Unit (GIWU):** The GIWU consists of both hardware and software that provides an interface to various networks for data communications. Through the GIWU, users can alternate between speech and data during the same call. The GIWU hardware equipment is physically located at the MSC/VLR.

1.2.5 GSM Specifications

Specifications for different personal communication services (PCS) systems vary among the different PCS networks [3]. Listed below is a description of the specifications and characteristics for GSM.

- **Frequency Band:** The frequency range specified for GSM is 1,850 to 1,990 MHz (mobile station to base station).
- **Duplex Distance:** The duplex distance is 80 MHz. Duplex distance is the distance between the uplink and downlink frequencies. A channel has two frequencies, 80 MHz apart.
- **Channel Separation:** The separation between adjacent carrier frequencies. In GSM, this is 200 kHz.
- **Modulation:** Modulation is the process of sending a signal by changing the characteristics of a carrier frequency. This is done in GSM via Gaussian minimum shift keying (GMSK).

- **Transmission Rate:** GSM is a digital system with an over-the-air bit rate of 270 kbps.
- **Access Method:** GSM utilizes the Time Division Multiple Access (TDMA) concept. TDMA is a technique in which several different calls may share the same carrier. Each call is assigned a particular time slot.

Table 1.2, summarizes the different standards i.e. Global System for Mobile Communications (GSM), Interim Standard –136 (IS-136) and Interim Standard –95 (IS-95).

Table 1.2 GSM and IS standards

	GSM	**IS-136**	**IS-95**
Year of Introduction	1990	1991	1993
Access Method	TDMA	TDMA	CDMA
BS Txn Band (MHz)	935-960	869-894	869-894
MS Txn Band (MHz)	890-915	824-849	824-849
Spacing between FWD & REV Channel ((MHz)	45	45	45
Channel BW (KHz)	200	200	200
No. of Duplex Channels	125	832	20
MS Max. Power	20	3	0.2
Users per Channel	8	3	35
Modulation	GMSK	Π/4 DQPSK	QPSK

As the basic network scheme of GSM is discussed in 1.2, it is found that one of the main functions of BSCs is the Handover. The introduction of this Handover concept is discussed in the section 1.3

1.3 Handoff in GSM

One of the most essential elements of cellular communication is the Handoff (HO). It is defined as the process of changing the current radio channel to a new radio channel which mainly takes place because of the movement of mobile unit and unfavorable radio conditions (degradation of the received signal strength) inside the individual cell or between numbers of adjacent cells [3]. Thus, Handover is the process that transfers an ongoing call from one cell to another as MS moves through the coverage area of a cellular system. Figure 1.5 shows the handover of a call from one cell supported with base station BS0 to another with base station BS1.

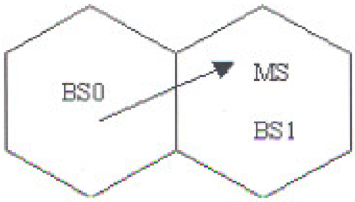


Figure 1.5 Handover from Cell_1 to Cell_2

In GSM, the concept of Hard Handoff (HHO) is used which is essentially a "break before make" connection. Thus in this, the link to the prior Base Station (BS) is terminated before or as the user is transferred to the new cell's base station such that the mobile is linked to at most only one base station at a given time. This hard handoff is primarily used in Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) in which the different frequency ranges are used in adjacent channels so as to reduce the adjacent channel. This HHO can be seen in figure 1.6.

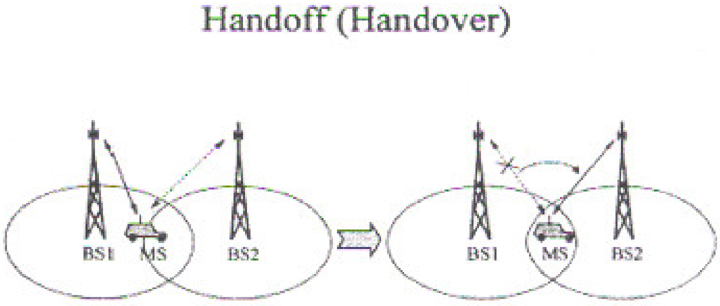


Figure 1.6 Hard handoff between the two base stations

The Initiation of the handoff (HO) may begin when the signal strength at the mobile received from the target base station is greater than that of serving base station. But when the mobile is between the two base stations, it causes the mobile to wildly switch links with either base station [4]. MS is switched back and forth between adjacent BSs near the cell boundary. This is known as ping- pong effect. Consequently, the number of handoffs also increases rapidly. Thus, these handoffs should be performed in such a manner that these should be imperceptible to the users and also the Quality of Service (QOS) should be maintained throughout. Moreover, the performance of the cellular systems is heavily influenced by the handover schemes.

1.4 Problem Specification

In order to deal with such complexities associated with the handover, there exists a need for some new systems, which are intelligent enough to predict the timely need for the handover and are capable of executing the same swiftly once initiated. This intelligence can be imparted to the system using one of the emerging technologies such as fuzzy logic, neural networks, particle swarm optimization etc. In this work, we have used the artificial neural network, which have a far reaching potential as building blocks in tomorrow's computational world.

1.4.1 Objectives of Thesis

- To carry out the literature survey to find the current trends in handover techniques in 2G cellular networks.
- To design a feedforward backpropagation based Artificial Neural Network (ANNs) for handover in GSM.
- To implement the same using VHDL on Field Programmable Gate Array (FPGA).
- To design a fuzzy inference system to evaluate the value of TDROP in order to update ACTIVE set of mobile stations in soft handoff region in CDMA.
- To design a feedforward backpropagation based ANN for above objective.
- To design a general-purpose VHDL code that can implement any two hidden layer based ANN with minor modifications of weight and bias values.

1.5 Organization of the thesis

In chapter 2, the basics of Cellular Handover process is discussed. This includes the different phases, desirable features, performance metrics and criteria for Handoff. Further, the different handover algorithms based on the conventional techniques and the emerging techniques are also discussed briefly.

Chapter 3 presents fundamentals of ANNs required for the design of ANNs. The chapter focuses on feedforward backpropagation learning algorithm of the Neural Networks. We also discuss the basics of fuzzy logic based systems.

Chapter 4 discusses the VHDL implementation of the Neural Networks. An introduction to the FPGAs is also included in this chapter so as to give the hardware concept to the designed system.

Chapter 5 discusses the implementation of the HO design for GSM. The algorithm is discussed and its conceptual design including the Fuzzy logic, ANNs, and the hardware implementation on FPGA is discussed.

Chapter 6 presents the basics of CDMA network. An algorithm for variation of threshold levels in CDMA is discussed. Finally, its conceptual design including the fuzzy logic, ANNs, and the hardware implementation on FPGA is also discussed.

Chapter 7 enlists the conclusions and future scope of the work.

Chapter 2

Cellular Handover Process: Current Trends

In this chapter, one of the most essential elements of the cellular phones i.e. handoff is discussed covering the basic phases, desirable features, performance metrics etc. The handover criteria and the types of algorithms based on the conventional and emerging techniques are also discussed briefly.

2.1 Basic Definition

During the last decade, there has been tremendous interest and progress in the field of wireless communication. The reason for this fast growth is the mobility of a tetherless terminal, which leads to growing use of wireless communication services. To meet this demand, cellular architecture is used, in which smaller cells are deployed to increase the traffic capacity in a limited spectrum allocation [1,2]. But smaller cells, in turn, need more cell boundary crossings as mobile station (MS) moves through the service area. When a call is in progress, the network is required to maintain the call across the cell boundaries. Handover is the process that transfers an ongoing call from one cell to another as MS moves through the coverage area of a cellular system.

A simple handoff scenario is shown in Figure 2.1[6] in which a Mobile Station (MS) travels from BS A to BS B. Initially, the MS is connected to BS A. The overlap between the two cells is the handoff region in which the mobile may be connected to either BS A or BS B. At a certain time during the travel, the mobile is handed off from BS A to BS B. When the MS is close to BS B, it remains connected to BS B.

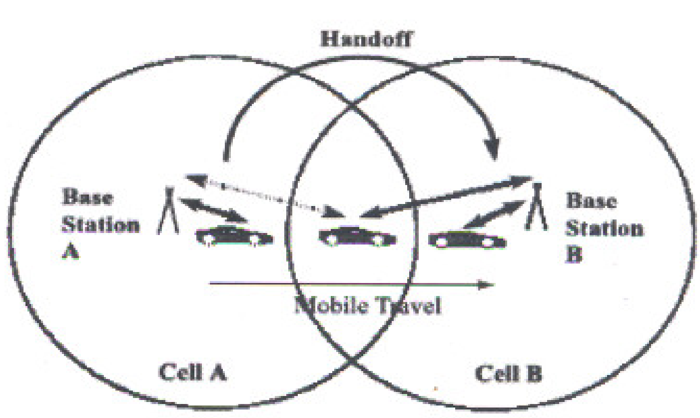


Figure 2.1 Handoff Scenario in cellular Systems

2.2 Factors Effecting Handoffs

Handoff may be caused by the factors [5] related to:

- Radio link,
- Network management, or
- Service options.

I. **Radio Link Related Causes**. Radio link related causes reflect the quality perceived by users. Some of the major variables affecting the service quality are received signal strength (RSS), signal-to-interference ratio (SIR), and system related constraints. Insufficient RSS and SIR reduce the service quality. Moreover, if certain system constraints are not met, service quality is adversely affected.

Handoff is required in the following situations due to reduced RSS:

(i) when the MS approaches the cell boundary (the RSS drops below a threshold).

(ii) when the MS is inside the signal strength holes in a cell (the signal is too weak to be detected easily).

II. **Network Management Related Causes**. The network may handoff a call so as to avoid the congestion in a cell. For a macroscopic diversity call, the handoff of calls in progress may be required since the same channel must be obtained in a number of BSs. If the network identifies that the path used for information transfer is malfunctioning or is not the shortest one, it may handoff the call.

III. **Service Options Related Causes**. When an MS asks for a service that is not provided at the current BS, the network may initiate a handoff so that the desired service can be offered. A handover may also be initiated by the MS to connect to a service provider with a lower tariff.

Network management and service related handoffs are usually infrequent and easy to tackle [5]. However, radio link related handoffs are most commonly encountered and most difficult to handle. A handoff made within the currently serving cell (e.g., by changing the frequency) is called an **intracell handoff**. A handoff made from one cell to another is referred to as an **intercell handoff**. Handoff may be hard or soft. **Hard handoff (HHO)** is 'break before make', meaning that the connection to the old BS is broken before a connection to the candidate BS is made. HHO occurs when handoff is made between disjointed radio systems, different frequency assignments, or different air interface characteristics or technologies. For e.g. the GSM based systems uses the HHO. **Soft handoff (SHO)** is 'make before break', meaning that the connection to the old BS is not broken until a connection to the new BS is made. In fact, more than one BSs are normally connected simultaneously to the MS as can be seen in the emerging CDMA based systems. These concepts are discussed in detail in the later chapters.

2.2.1 Handoff Decision

The decision making process of handoff may be centralized or decentralized (i.e. the handoff decision may be made at the mobile station or network) [5]. From decision process point of view one can find at least three different kind of handoff:

- In a network controlled handoff protocol, the network makes a handoff decision based on measurements of the mobile stations at a number of base stations. In general, the handoff process (including data transmission, channel switching and network switching) takes 100-200 ms. Information about the signal quality for all users is available at a single point in the network. This information facilitates the resource allocation. Network controlled handoff is used in first generation analog systems.
- In a mobile assisted handoff process the mobile station makes measurements, and the network makes decision. In circuit switched GSM, base station controller is in charge of the radio interface management. This means mainly the allocation and release of radio channels and handoff management. The handoff time (time between handoff decision and execution) in circuit switched GSM is approximately 1s.
- In mobile controlled handoff the mobile station is completely in control of the handoff process. This type of handoff has a short reaction time (on the order of 0.1s). The mobile station measures the signal strengths from surrounding base stations and interference levels on all channels. A handoff can be initiated if the signal strength of the serving base station is lower than that of another base station by certain threshold.

2.2.2 Phases of Handoff

Handover process consists of the following three phases [4]:

1. **Handover initiation**. Either the MS or network identifies the need for a handover and the handover is initiated.

2. **Allocation of resources**. After the HO is initiated, necessary resources either network or radio are reserved to support the handover.

3. **Handover Execution**. In this phase, the actual handover of the call without violating the quality of service (QoS) requirements of the user takes place.

2.3 Desirable Features of Handoff

An efficient handoff algorithm can achieve many desirable features by trading different operating characteristics [1,2,5]. A handoff algorithm should have the following desirable features.

- **Speed:** Handoff should be fast so that the user does not experience service degradation or interruption. Service degradation may be due to a continuous reduction in signal strength or an increase in Co-Channel Interference (CCI). Service interruption may be due to a "break before make" approach of HHO. Note that the delay in the execution of a handoff algorithm adds to the network delay at the Mobile Switching Center (MSC) or Mobile Telephone Switching Office (MTSO). Fast handoff also reduces CCI since it prevents the MS from going too far into the new cell.
- **Reliability:** Handoff should be reliable. This means that the call should have good quality after handoff. SIR and RSS help determine the potential service quality of the candidate BS.
- **Success:** Handoff should be successful that is a free channel should be available at the candidate BS. Efficient channel allocation algorithms and some traffic balancing can maximize the probability of a successful handoff.
- **QoS:** The effect of handoff on the Quality of Service (QoS) should be minimal. The quality of service may be poor just before handoff due to a continuous reduction in RSS, SIR, etc.
- **Congestion:** Handoff should maintain the planned cellular borders to avoid congestion, high interference, and use of assigned channels inside the new cell. Each BS can carry only its planned traffic load. Moreover, there is a possibility of

increased interference if the MS goes far into another cell site while still being connected to a distant BS because cochannel distance is reduced and the distant BS tends to use a high transmit power to serve the MS.

- **Number:** The number of handoffs should be minimized. Excessive handoffs lead to heavy handoff processing loads and poor communication quality, which may be due to the following:
  I. The more the attempts for handoff, the more chances that a call will be denied access to a channel, resulting in a higher handoff call dropping probability,
  II. A lot of handoff attempts cause more delay in the MSC processing of handoff requests, which will cause signal strength to decrease over a longer time period to a level of unacceptable quality. Also, the call may be dropped if sufficient SIR is not achieved. Handoff requires network resources to connect the call to a new BS. Thus, minimizing the number of handoffs reduces the switching load. Unnecessary handoffs should be prevented; the current BS might be able to provide the desired service quality without interfering with other MSs and BSs.
  III. The target cell should be chosen correctly since there may be more than one candidate BS for handoff. Identification of a correct cell prevents unnecessary and frequent handoffs.
- **Drop-Outs:** The handoff procedure should minimize the number of continuing call drop-outs by providing a desired QoS (e.g., by ensuring a certain SIR).
- **Call Blocking:** Handoff should have minimal effect on new call blocking. For example, if some channels (called guard channels) are reserved exclusively for handoff, new call blocking probability will increase due to the reduction in the number of channels available for a new call.
- **Balance:** The handoff procedure should balance traffic in adjacent cells, eliminating the need for channel borrowing, simplifying cell planning and operation, and reducing the probability of new call blocking.
- **Interference:** The global interference level should be minimized by the handoff procedure. Transmission of bare minimum power and maintenance of planned cellular borders can help achieve this goal.

Figure 2.2 summarizes the major desirable features of handoff algorithms.

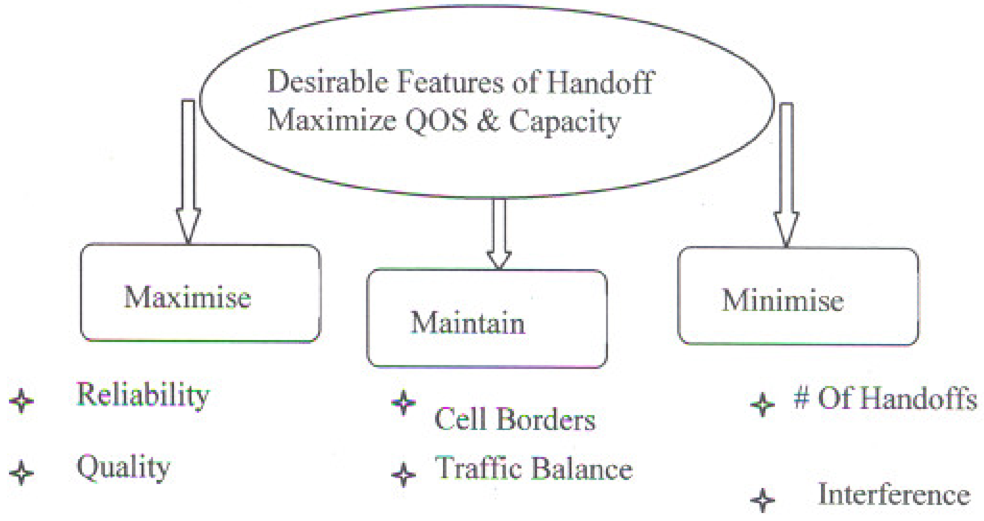


Figure 2.2 Desirable Features of Handoff

2.4 Performance Metrics

This subsection discusses basic performance aspects of hand-over [3]. The performance metrics used to evaluate handover algorithms are:

- **Call blocking probability:** the probability that a new call attempt is blocked.
- **Handover blocking probability:** the probability that a handover attempt is blocked.
- **Handover probability:** the probability that, while communicating with a particular cell, an ongoing call requires a handover before the call terminates. This metric translates into the average number of handovers per call.
- **Call dropping probability:** the probability that a call terminates due to a handover failure. This metric can be derived directly from the handover blocking probability and the handover probability.
- **Probability of an unnecessary handover:** the probability that a handover is stimulated by a particular handover algorithm when the existing radio link is still adequate.
- **Rate of handover:** the number of handovers per unit time. Combined with the average call duration, it is possible to determine the average number of handovers per call, and thus the handover probability.

- **Duration of interruption:** the length of time during a handover for which the mobile terminal is in communication with neither base station. This metric is heavily dependent on the particular network topology and the scope of the handover
- **Delay:** the distance the mobile moves from the point at which the handover should occur to the point at which it does.

2.5 Criteria for Handoff

Several variables have been proposed and used as inputs, or basic criteria, to handoff algorithms such as signal strength, SIR, distance, traffic, call and handoff statistics, and velocity [3,4,7].

- **Received Signal Strength (RSS):** This criterion is simple, direct, and widely used. Many systems are interference limited, meaning that signal strength adequately indicates the signal quality, and this is the motivation behind signal strength based decision. Moreover, there is a close relation between the RSS and the distance between the BS and the MS. The lack of consideration of CCI is a disadvantage of this criterion. CCI is more important in microcells (with a cell radius less than 1 km) than in macrocells since microcellular systems are interference limited and macrocells are noise-limited, (with cell radius exceeding 35 km in rural areas). Moreover, several factors (e.g., topographical changes, shadowing due to buildings, and multipath fading) can cause the actual coverage area to be quite different from the intended coverage area. The RSS criterion can also lead to excessive number of handoffs.

- **Signal-to-Interference Ratio (SIR):** An advantage of using SIR as a criterion is that SIR is a parameter common to voice quality, system capacity and dropped call rate. BER is often used to estimate SIR. When actual Channel to Interference Ratio (C/I) is lower than the designed (C/I), voice quality becomes poor, and the rate of dropped calls increases. SIR also determines the reuse distance. Unfortunately, CIR may oscillate due to propagation conditions and may cause the ping-pong effect (in which the MS repeatedly switches between the adjacent BSs). Another disadvantage is that even though BER is a good indicator of link quality, bad link quality may be experienced near the serving BS, and handoff may not be desirable in such situations. In an interference-limited environment, deterioration in BER does not necessarily imply the need for an intercell handoff; an intracell handoff may be sufficient. In [7], two methods for estimating raw channel BER over a Rayleigh fading channel are presented.

I. The first method uses a preselected Pseudo-Noise (PN) sequence as a frame synchronization pattern. This sequence is transmitted over a Rayleigh fading channel. A decision variable is calculated by finding the autocorrelation of the channel-corrupted PN sequence with the original PN sequence. The autocorrelation value is related to the channel Signal-to-Noise Ratio (SNR) to compute the raw BER estimate.

II. The second method assumes that bounded distance decoders are used in a Reed-Solomon based system to estimate the number of errors in a received word prior to decoding. First, the channel Symbol Error Rate (SER) is estimated by assuming that the number of errors in a received word are less than the error correction capability of the code. Then, using the channel SER, raw channel Bit Error Rate (BER) and post-decoder BER are derived. Reference [8] proposes BER as a handoff criterion for an integrated system that consists of a cellular system and a terrestrial system.

The level crossing rate (LCR) is an important quantity that characterizes the rate at which the envelope of Rayleigh faded signals crosses a specified signal level in the positive slope [1]. The LCR is used to determine average signal strength. From the average signal strength, the BER is determined based on the modulation scheme used for data transmission.

- **Distance:** This criterion can help preserve planned cell boundary. The distance can be estimated based on signal strength measurements, delay between the signals received from different BSs, etc. Distance measurement can improve the handoff performance. If handoff occurs at the midway between two BSs, it distributes the channel utilization evenly. The distance measurement requires some special technical equipment and is possible in systems with only a common clock. Synchronization between the BSs is also required. Since future systems will require the location information of the MS, distance measurement will be available for use as a handoff criterion. The distance criterion may be useful for a macro cellular system, but it is prohibitive in a microcellular system since the precision of the distance measurement decreases with smaller cell sizes [4,9]. Theoretical analysis does not consider distance criterion better than others. The determination of cell boundaries can avoid unnecessary handoffs.

- **Transmit Power:** Transmit power can be used as a handoff criterion to reduce the power requirement, reduce interference, and increase battery life.

- **Traffic:** Traffic level as a handoff criterion can balance traffic in adjacent cells. Statistics of dwell times are important for teletraffic performance evaluation.

- **Call and Handoff Statistics:** Statistics such as total time spent in the cell by a call and arrival time of a call in a cell can also be used as handoff criteria [1]. Elapsed time since last handoff is also a useful criterion since it can reduce the number of handoffs.

- **Velocity:** Velocity is an important handoff criterion, especially for the velocity adaptive algorithms. Several algorithms use an estimate of velocity to modify handoff parameters. In [5] a method to adaptively change the averaging interval in a handoff algorithm for both small and large cells is presented. The method is based on the estimation of mobile velocity through maximum Doppler frequency fd. Some of the methods for estimating fd use the following criteria: mobile velocity, level crossings, spectrum or autocorrelation analysis, and squared deviations of logarithmically compressed envelope (amplitude) measurements. Reference [5] describes the last approach. The effects of averaging interval, CCI, and Additive White Gaussian Noise (AWGN) on fd estimation are analyzed. The fd estimates can then be used to modify weighting factors of exponential averaging in a handoff algorithm.

2.6 Handoff Algorithms

Handoff algorithms are distinguished from one another in two ways, handoff criteria and processing of handoff criteria [7]. However, these algorithms can be broadly classified into two main categories i.e. Conventional and Emerging Handoff Algorithms. The figure 2.3 summarizes these various algorithms.

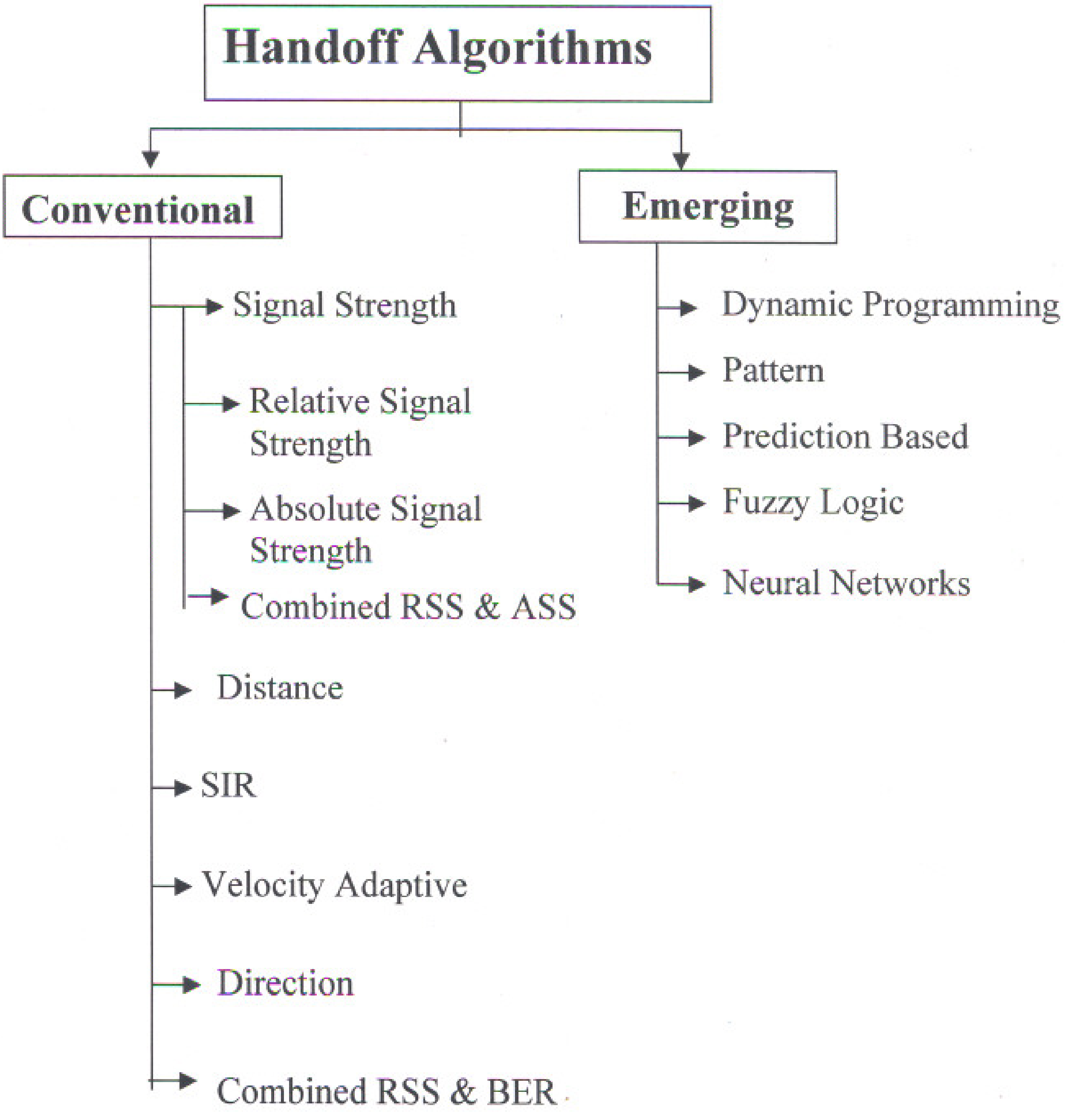


Figure 2.3 Handoff Algorithms at a Glance

2.6.1 Conventional Handoff Algorithms

Conventional handoff algorithms are the ones based on the parameters discussed in section 2.5. Some of the main algorithms are discussed below.

2.6.1.1 Signal Strength Based Algorithms

There are several variations of signal strength based algorithms, including relative signal strength algorithms, absolute signal strength algorithms, and combined absolute and relative signal strength algorithms [4]. These algorithms are briefly discussed next.

- **Relative Signal Strength Algorithms**

The averaged signal strength of base station 1 decreases as the mobile moves away from it. Similarly, the averaged signal strength of base station 2 increases as the mobile approaches it. Using this figure, the following discussion explains various approaches [4,5]:

I. **Relative signal strength:** Relative signal strength chooses the strongest received base station at all times. The decision is based on an averaged measurement of the received signal. In Fig. 2.4, the handover will occur at position A. This method is shown to stimulate too many unnecessary handovers when the current base station signal is still adequate [9].

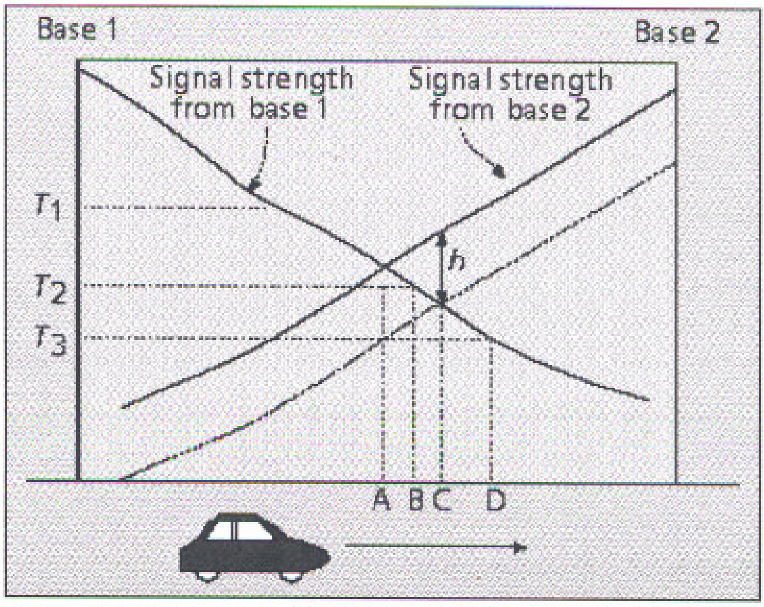


Figure 2.4. Variation in Signal Strength of MS

II. **Relative signal strength with threshold**: Relative signal strength with threshold allows a user to handover only if the current signal is sufficiently weak (less than a threshold) and the other is the stronger of the two. [7] The effect of the threshold depends on its value compared to the signal strengths of the two base stations at the point at which they are equal. If the threshold is higher than this value, say T_1 in Fig. 2.41, this scheme performs exactly like the relative signal strength scheme, so the handover occurs at position A. If the threshold is lower than this value, say T_2 in Fig. 2.41, the mobile will delay handover until the current signal level crosses the threshold at position B. In the case of T_3, the delay may be so long that the mobile drifts far into the new cell. This reduces the quality of the communication link and may result in a dropped call. In addition, this causes additional interference to co-channel users. Thus, this scheme may create overlapping cell coverage areas [9]. A threshold is not used alone in practice because its effectiveness depends on prior knowledge of the crossover signal strength between the current and candidate base stations.

III. **Relative signal strength with hystersis:** Relative signal strength with hysteresis allows a user to handover only if the new base station is sufficiently stronger (by a hysteresis margin, h, in Fig. 2.41) than the current one. In this case the handover will occur at point C. This technique prevents the so-called ping-pong effect, the repeated handover between two base stations caused by rapid fluctuations in the received signal strengths from both base stations [4,9,13]. The first handover, however, may be unnecessary if the serving base is sufficiently strong.

IV. **Relative signal strength with hystersis and threshold:** Relative signal strength with hysteresis and threshold hands a user over to a new base only if the current signal level drops below a threshold and the target base station is stronger than the current one by a given hysteresis margin [7]. In Fig. 2.4, the handover will occur at point C if the threshold is either T_1 or T_2, and will occur at point D if the threshold is T_3.

- **Absolute Signal Strength Algorithms**

For this algorithm, when the RSS drops below a threshold level, handoff is requested. Typical threshold values are -100 dBm for a noise-limited system and -95 dBm for an interference-limited system [2]. Better handoff initiation can be obtained by varying the threshold [2]. The threshold level should be varied according to the path loss slope L of the RSS and the level crossing rate (LCR) of the RSS.

- If the slope L is high or LCR is high, the MS is quickly moving away from the BS, and, hence, handoff should be made fast (i.e., the handoff initiation threshold should be made higher in magnitude).
- If the slope L or LCR is low, the MS is moving slowly. So, handoff can be slow; the handoff initiation threshold can be made comparatively smaller.

Thus, the mobile velocity and path-loss slope L can be used to determine the handoff initiation threshold dynamically such that the number of unnecessary handoffs is minimized and necessary handoffs are completed successfully.

This algorithm has a serious disadvantage. When a threshold level is set based on the RSS, the following situations pose a problem [7]:

I. when RSS is high due to high interference, the handoff will not take place, although, ideally, handoff is desirable to avoid interference;

II. when RSS is low, handoff takes place even if voice quality is good, although, ideally, such a handoff is not required, and some systems use supervisory audio tone (SAT) information with the RSS to avoid handoff.

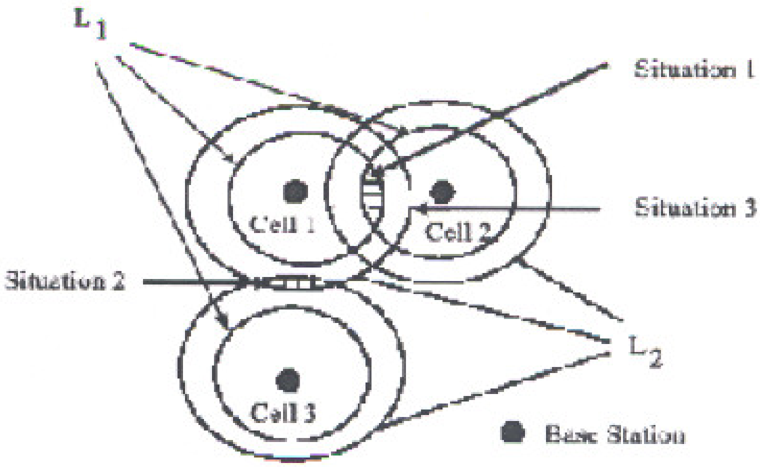


Fig.2.5. A Two Level Handoff Algorithm

A variation of the basic threshold algorithm is a two-level algorithm, which provides more opportunity for a successful handoff. Two handoff thresholds, L1 and L2, are defined with L1 higher than L2 as shown in Figure 2.5. When the RSS drops below L1, handoff request is initiated. If the MS is in a signal strength hole in the current cell or the candidate cell is busy, the possibility of handoff must be assessed. In this case, handoff is requested periodically.

The handoff request is entertained only if the new RSS is stronger (Situation 1 in Figure 2.5). However, if the RSS reaches L2, handoff will be made regardless of the relative signal strength of the candidate BS (Situation 2 in Figure 2.5). Due to the two-level algorithm, the MS may come out of the hole, or the candidate BS may have a free channel for handoff between L1 and L2.

If a single threshold L2 were used, the L2 boundary might have been close to the candidate BS, causing interference (Situation 3 in Figure 2.5). However, in a two-level algorithm, L1 boundaries of the BSs will have allowed handoff to be made earlier, avoiding high interference levels.

- **Combined Absolute and Relative Signal Strength Algorithms**

According to this algorithm, handoff takes place if the following two conditions are satisfied [5]:

I. The average signal strength of the serving BS falls below an absolute threshold T (dB), and

II. The average signal strength of the candidate BS exceeds the average signal strength of the current BS by an amount of h (hysteresis) dB.

The first condition prevents the occurrence of handoff when the current BS can provide sufficient signal quality. An optimum threshold (T) achieves the narrowed handoff area (and hence reduced interference) and a low expected number of handoffs.

Basic variables for this handoff algorithm are the length and shape of the averaging window, the threshold level, and the hysteresis margin [4]. Some of the findings for this algorithm are:

I. The probability of not finding a handoff candidate channel decreases as the overlap region increases;

II. The probability of not finding a handoff candidate increases as the handoff threshold increases;

III. The probability of a late handoff decreases as handoff threshold increases;

IV. The probability of unnecessary handoffs (i.e., the ping-pong effect) increases as handoff threshold increases;

V. The probability of unnecessary handoff decreases as hysteresis increases.

2.6.1.2 Distance Based Algorithms

This algorithm connects the MS to the nearest BS. The relative distance measurement is obtained by comparing propagation delay times. This criterion allows handoff at the planned cell boundaries, giving better spectrum efficiency compared to the signal strength criterion.

However, it is difficult to plan cell boundaries in a microcellular system due to complex propagation characteristics. Thus, the advantage of distance criterion over signal strength criterion begins to disappear for smaller cells due to inaccuracies in distance measurements [5]. A relative signal strength based algorithm gives less interference probability compared to a relative distance based algorithm. In particular, when an LOS path exists between the current and distant BS and the MS, the current BS gives stronger signal strength compared to the nearer NLOS BS. In such cases, the relative signal strength criterion can avoid interference; relative distance criterion experiences more interference.

2.6.1.3 SIR Based Algorithms

For good quality voice, SIR at the cell boundary should be relatively high (e.g., 18 dB for AMPS and 12 dB for GSM). However, a lower SIR may be used for capacity reasons since co channel distance and cluster size (i.e., the number of cells per cluster) are small for lower SIR and channels can be reused more frequently in a given geographical region [2].

SIR is a measure of communication quality. This algorithm makes a handoff when the current BS's SIR drops below a threshold and another BS can provide sufficient SIR. Hysteresis can be incorporated in the algorithm. The lower SIR may be due to high interference or low carrier power. In either case, handoff is desirable when SIR is low. However, SIR-based handoff algorithms prevent handoffs near nominal cell boundaries and cause cell dragging and high transmit power requirements [1].

In analog systems, measuring SIR during a call is difficult. Hence, sometimes interference power is measured before a call is connected, and combined signal and interference power is measured during the call.

2.6.1.4 Velocity Adaptive Algorithms

Handoff requests from fast moving vehicles must be processed quickly. A handoff algorithm with short temporal averaging windows can be used to tackle fast users. However, the concept of a "short" averaging window is relative to the mobile speed. Thus, optimal handoff performance will be obtained only at one speed if the length of the averaging window is kept constant [10].

A velocity adaptive handoff algorithm provides good performance for MSs with different velocities by adjusting the effective length of the averaging window. A velocity adaptive handoff algorithm can serve as an alternative to the umbrella cell approach to tackle high speed users if low network delay can be achieved, which can lead to savings in the infrastructure.

One of the velocity estimation techniques uses level crossing rate (LCR) of the RSS in which the threshold level should be set as the average value of the Rayleigh distribution of the RSS [1], requiring special equipment to detect the propagation dependent average receiver power.

2.6.1.5 Direction Biased Algorithms

In an NLOS handoff, the MS experiences the corner effect. Hence, if the MS moves fast and is not handed off quickly enough to another BS, the call will be dropped. Connecting the fast moving vehicles to an umbrella cell is one solution, and using better handoff algorithms is another solution [1,3].

A direction biased handoff algorithm represents such an alternative solution [5]. Direction biasing improves cell membership properties and handoff performance in LOS and NLOS scenarios in a multi-cell environment. A handoff algorithm is said to possess good cell membership properties if the probability that the MS is assigned to the closest BS is close to one throughout the call duration [8]. Improvement in cell memberships leads to fewer handoffs and reduced interference.

The basic idea behind this algorithm is that handoffs to the BSs towards which the MS is moving are encouraged, while handoffs to the BSs from which the MS is receding are discouraged. This algorithm reduces the probability of dropped calls for hard handoffs (e.g., for TDMA systems). The algorithm also reduces the time a user needs to be connected to more than one base station for soft handoffs (e.g., for CDMA systems), allowing more potential users per cell.

A variation of the basic direction-biased algorithm is the pre-selection direction biased algorithm. If the best BS is a receding one and has a quality only slightly better than the second best BS that is being approached, the handoff should be made to the second best BS because it is more likely to improve its chances of being selected. This provides a fast handoff algorithm with good cell membership properties without the undesirable effects associated with large hysteresis.

2.6.1.6 RSS and BER Based Algorithms

An algorithm based on both RSS and BER is described in [4,5]. For RSS, a threshold is used for the current BS, and a hysteresis window is used for the target BS. For BER, a separate threshold is defined. The target BS can be either included or excluded from the handoff decision process. The latter scheme is used in GSM in which the mobile does not know the signal quality of the target BS. In principle, it is possible to measure BER of the control channel of the target BS.

Three parameters considered in [7] are RSS threshold, BER threshold, and RSS hysteresis window size. The effects of these parameters on handoff probability are discussed. In general, a low threshold value reduces the handoff request probability. The best threshold value is the average signal level at the mid-point between two BSs. However, due to the propagation environment, this threshold must be estimated for each base site. An RSS hysteresis delays handoff significantly. The higher the BER threshold, the earlier the handoff request. Moreover, if the BER threshold of target BS is used, handoff request is delayed. The handoff request probability differs significantly with location (or BS sites), showing that propagation characteristics are highly dependent on local terrain features and environment.

From experimental results, it was found that the signal level and BER profiles varied significantly. RSS gives a direct indication of the received energy at the MS, while BER gives an indication of CCI and transmission quality. The effect of threshold level of RSS on handoff is opposite to that of BER since the gradient of BER level is opposite to that of RSS level. If the gradient of the signal level is steep, the handoff region is less sensitive to a small variation in the threshold. Hysteresis is useful in preventing premature handoff requests if signal profiles are fluctuating. This is very useful in small site cells. In large site cells, hysteresis should be relatively small since it may introduce a delay in handoff initiation. Actual data (i.e. measured RSS and BER) is used in a software simulator that implements handoff schemes.

2.6.2 Emerging Handoff Algorithms

The Algorithms based on the emerging techniques discussed in figure 2.3 are known as the Emerging Algorithms. These algorithms are the recent algorithms and offer a wide range of study. However, much of the research was not carried out in this field earlier but now the scenario is changed and a lot of work is based on these new techniques[7]. Moreover, the performance of the systems based on these techniques is much better than the conventional systems. These algorithms are discussed briefly in the section below.

2.6.2.1 Dynamic Programming Based Handoff Algorithms

Dynamic programming allows a systematic approach to optimization. However, it is usually model dependent (particularly the propagation model) and requires the estimation of some parameters and handoff criteria, such as signal strengths. So far, dynamic programming has been applied to very simplified handoff scenarios only [5].

Handoff is viewed as a reward/cost optimization problem. RSS samples at the MS are modeled as stochastic processes. The reward is a function of several characteristics (e.g., signal strength, CIR, channel fading, shadowing, propagation loss, power control strategies, traffic distribution, cell loading profiles, and channel assignment). Handoffs are modeled as switching penalties that are based on resources needed for a successful handoff. Dynamic programming is used to derive properties of optimal policies for handoff. Simulation results show this algorithm to be better than relative signal strength based algorithm.

Reference [4,5] views signal strength based handoff as an optimization problem to obtain a tradeoff between the expected number of handoffs and number of service failures, events that occur when the signal strength drops below a level required for an acceptable service to the user. An optimal solution is derived based on dynamic programming and is used for comparison with other solutions. The handoff problem is defined as a finite horizon dynamic programming problem, and an optimal solution is obtained through a set of recursive equations. This optimal solution is complex and requires a priori knowledge of the mobile trajectory. A locally optimal (or greedy) algorithm has been derived that uses the threshold level and gives a reasonable number of handoffs. A technique for adapting the algorithm is also suggested.

2.6.2.2 Pattern Recognition Based Handoff Algorithms

Pattern recognition (PR) identifies meaningful regularities in noisy or complex environments. These techniques are based on the idea that the points that are close to each other in a mathematically defined feature space represent the same class of objects or variables. Explicit PR techniques use discriminant functions that define (n-1) hyper surfaces in an n-dimensional feature space. The input pattern is classified according to their location on the hyper surfaces. Implicit PR techniques measure the distance of the input pattern to the predefined representative patterns in each class. The sensitivity of the distance measurement to different representative patterns can be adjusted using weights [7]. The clustering algorithms and fuzzy classifiers are examples of implicit methods. The environment in the region near cell boundaries is unstable, and many unnecessary handoffs are likely to occur. The PR techniques can help reduce this uncertainty by efficiently processing the RSS measurements.

2.6.2.3 Prediction-based Handoff Algorithms

Prediction-based handoff algorithms use the estimates of future values of handoff criteria, such as RSS. In [7] proposes this technique and shows it to be better than the relative signal strength algorithm and the combined absolute and relative signal strength algorithm via simulations. An adaptive prediction based algorithm has been proposed to obtain a tradeoff between the number of handoffs and the overall signal quality [12]. Signal strength based handoff algorithms can use path loss and shadow fading to make a handoff decision. The path loss depends on distance and is determinate. The shadow fading variations are correlated and hence can be predicted. The correlation factor is a function of the distance between the two locations and the nature of the surrounding environment [7]. The proposed prediction based algorithm exploits this correlation property to avoid unnecessary handoffs. The future RSS is estimated based on previously measured RSSs using an adaptive FIR filter. The FIR filter coefficients are continuously updated by minimizing the prediction error. Depending upon the current value of the RSS (RSSc) and the predicted future value of the RSS (RSSp), handoff decision is given a certain priority. Based on the combination of RSSc and RSSp, hysteresis may be added if it will not affect the handoff performance adversely. The final handoff decision is made based on the calculated handoff priority.

2.6.2.4 Neural Handoff Algorithms

Most of the proposed neural techniques have shown only preliminary simulation results or have proposed methodologies without the simulation results. These techniques have used simplified simulation models. Learning capabilities of several paradigms of neural networks have not been utilized effectively in conjunction with handoff algorithms to date. Reference [10] presents a signal strength based handoff initiation algorithm using a binary hypothesis test implemented as a neural network. However, simulation results are not presented.

2.6.2.5 Fuzzy Handoff Algorithms

A fuzzy based multicriteria handoff algorithm is proposed in [11]. The fuzzy handoff algorithm has been shown to possess enhanced stability (i.e., less frequent handoffs). The algorithm so developed is based on the received signal strength threshold, received signal strength with the hystersis margin and the time interval for reduced signal level condition. The effect of threshold value and hystersis margin on the performance of HO algorithm has been studied in a descriptive manner. The only hystersis value used in conventional handoff algorithm may not be enough for heavy fadings, while fuzzy logic has inherent fuzziness that can model the overlap region between the adjacent cells. In [10], a fuzzy classifier is used to process the signal strength measurements to select a BS to serve a call. The performance of this algorithm in a microcellular environment is evaluated. It also explains the concept of cell membership degree to handoff. It is shown that the change of RSS threshold as a means of introducing a bias is an effective way to balance traffic while allowing few or no additional handoffs. It is suggested that a combination of range and RSS modified by traffic weighting might give good performance. Different fuzzy composition methods to combine the cell membership degrees of different criteria methods are investigated. Effects of changes in the cell membership degrees on handoff performance have been evaluated.

2.7 Conclusion

Handoff in cellular systems is a complex problem. Handoff is to be initiated and executed at an appropriate time so as to ensure the quality of service (QoS). The network designer has to ensure that while a handoff is initiated to ensure QoS, he has to keep the rate of handoff to minimum, thereby reducing the burden on the network while still maintaining the QoS. An early HO can cause ping pong effect whereas a late one can cause QoS to degrade or even cause the call to drop. Under such conflicting requirements, an intelligent HO system is the need that reduces the HO rate to minimum required while maintaining QoS. Artificial neural network is the one of the intelligent control strategies, which is used for such conflicting control requirements.

Since, it has been have decided to switch over to ANN based HO, this chapter is used to build the foundation on which our HO design is super structured.

Chapter 3

Neural Networks and Fuzzy Logic

In this chapter, the basic concept of the Fuzzy Logic and Neural Networks is discussed briefly. The introduction of the different types of neurons with their basic requisites is given. Further, the Feedforward Backpropagation algorithm is also included.

3.1 Artificial Neural Networks

Definition

According to Eberharty and Dobbins "A Neural Network is an analysis tool that is modeled after the massively parallel structure of the brain: It simulates a highly interconnected, parallel computational structure with many relatively simple individual processing elements, or neurons."[14]

It is rather an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. Artificial neural networks are collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. The key element of the ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to neurons and are tied together with weighted connections that are analogous to synapses.

3.1.1 Biological Model

The most basic component of neural networks i.e. neuron is modeled after the structure of the brain. Therefore, neural networks have a strong similarity to the biological brain. The most basic element of the human brain is a specific type of cell, which provides us with the abilities to remember, think, and apply previous experiences to our every action. These cells are known as neurons; each of these neurons can connect with up to 200000 other neurons. The power of the brain comes from the numbers of these basic components and the multiple connections between them.

All natural neurons have four basic components, which are dendrites, soma, axon, and synapses. The signal flow goes from left to right, from the dendrites, through the cell body, and out through the axon. The signal from the neuron is passed on to another by means of a connection between the axon of the first and a dendrite of the second. This connection is called a synapse. Large number of neurons collectively forms a neural network and this network continues to change throughout life. These later changes tend to consist mainly of strengthening or weakening of synaptic junctions. Hence it is believed that new memories are formed by modification of these synaptic strengths.

Basically, a biological neuron receives inputs from other sources, combines them in some way, performs a generally nonlinear operation on the result, and then output the final result. The Figure 3.1 [23] shows a simplified biological neuron and the relationship of its components.

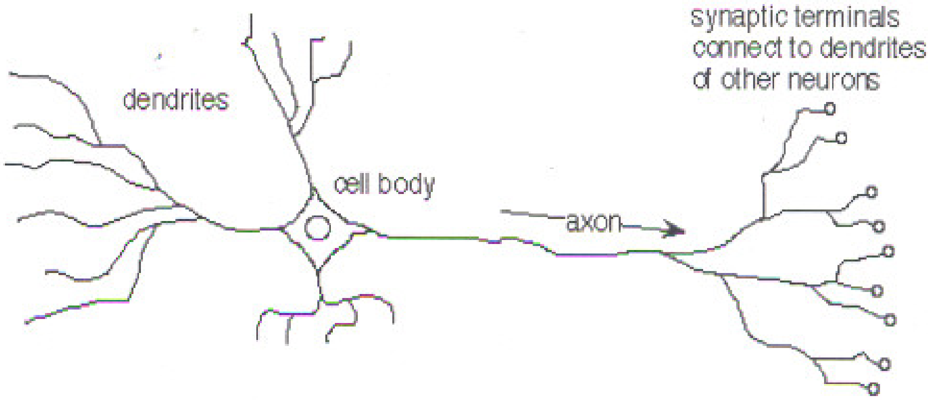


Figure3.1 Biological model of Brain

3.1.2 Artificial Neuron

Corresponding to the biological neuron, discussed above, ANN has an artificial neuron as the basic unit of neural networks that simulates the four basic functions of natural neurons. Artificial neurons are much simpler than the biological neuron [15].

3.1.3 Modeling of Artificial Neuron

Multiple Input Neuron

Typically, a neuron has more than one input. A neuron with R inputs is shown in the Figure 3.2. The individual inputs p_1, p_2, p_3... p_R are each weighted by corresponding elements $w_{1,1}$, $w_{1,2}$... $w_{1,R}$ of the weight matrix W.

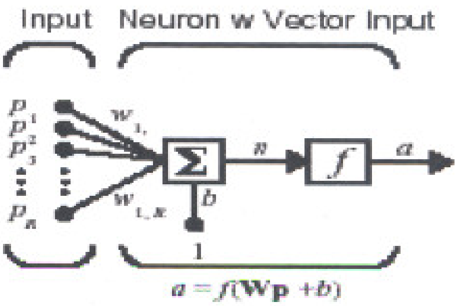


Figure 3.2 Multiple Input Neuron

The neuron has a bias b, which summed with the weighted inputs to form the net input n:

$$N = w_{1,1}p_1 + w_{1,2}p_2 + \dots\dots w_{1,R}p_R + b \qquad \text{(i)}$$

where R = number of elements in input layer

This expression can be written in the matrix form:

$$N = Wp + b \qquad \text{(ii)}$$

where the matrix W for the single neuron case has only one row.

Now the neuron output can be written as

$$A = f(Wp + b) \qquad \text{(iii)}$$

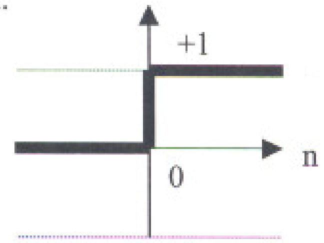
In the Eq. (iii), the actual output depends upon the particular transfer function, discussed in the next section. The bias is much like a weight, except that it has a constant input of 1. However, if one doesn't want to have a bias in a particular neuron, it can be omitted.

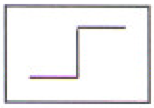
A particular convention in assigning the indices of the elements of the weighted matrix is adopted. The first index indicates the particular neuron destination for that weight. The second index indicates the source of the signal fed to the neuron. Thus, the indices in $w_{1,2}$ indicates that this weight represents the connection to the first neuron from the second source.

3.1.4 Types of Transfer Function:

Three types of Transfer functions are most commonly used and are given below:

Hard limit Transfer Function:





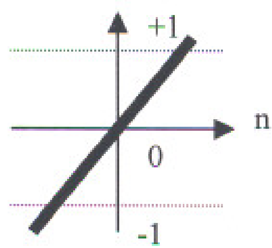
a=hardlim(n)

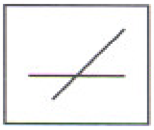
The output of neuron is set to be 0 if the summer output that is, n, is less than 0 and output is 1 if n is greater than or equal to 0. That is,

$$a = 0 \text{ if } \quad n<0$$

$$a = 1 \text{ if } \quad n>=0$$

Linear Transfer Function:



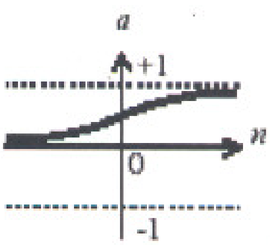


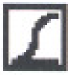
a=purelin(n)

The output of a linear transfer function is equal to its input, that is,

$$a = n$$

Logsigmoid Transfer Function:





a = logsig(n)

This transfer function takes the input and squashes the output into range 0 to 1, according to the expression:

$$a = \frac{1}{1+e^{-n}}$$

3.1.5 Network Architecture

Biologically, neural networks are constructed in a three dimensional way from microscopic components. These neurons seem capable of nearly unrestricted interconnections. Artificial neural networks are the simple clustering of the primitive artificial neurons. This clustering occurs by creating layers, which are then connected to one another. How these layers connect may also vary. Basically, all artificial neural networks have a similar structure of topology. Some of the neurons interface the real world to receive its inputs and other neurons provide the real world with the network's outputs. All the rest of the neurons are hidden from the view.

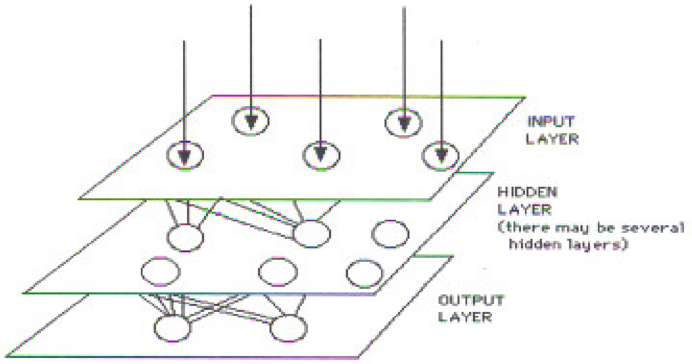


Figure 3.3 Layers of Neural Network

As the Figure3.3 [23] shows, the neurons are grouped into layers. The input layer consists of neurons that receive input form the external environment. The output layer consists of neurons that communicate the output of the system to the user or external environment. There are usually a number of hidden layers between these two layers; the figure above shows a simple structure with only one hidden layer. When the input layer receives the input its neurons produce output, which becomes input to the other layers of the system. The process continues until a certain condition is satisfied or until the output layer is invoked and fires their output to the external environment.

Thus we can say that, " An ANN is a massively parallel-distributed processor that stores experimental knowledge; this knowledge is acquired by a learning process and is stored in the form of parameters of the ANN." The basic characteristics of ANNs are massively parallel distributed architecture, ability to learn and generalize, fault tolerance, nonlinearity, and adaptivity.

3.1.6 Layered Classification of Neural Networks

3.1.6.1 Single Layer Neural Network

A single layer neuron network of S neurons is shown in the Figure.3.4. Here each of the R input is connected to each of the neurons and the weight matrix has S rows.

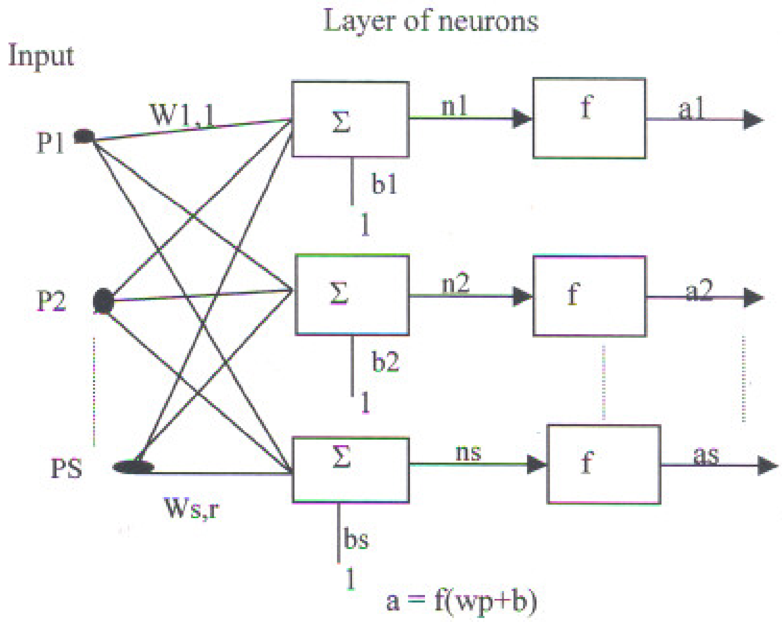


Figure 3.4 Single layer Neurons

The layer includes the weight matrix, the summers, the bias vector b, the transfer function and the output vector *a*. Each element of the input vector p is connected to each neuron through weight matrix W. However the number of inputs to the layer may be different from the numbers of neurons in the layer [15].

3.1.6.2 Multiple layers of neural Network

Now consider a network with several layers. Each layer has its own weight matrix W, bias vector b, net input vector n and an output vector a. Some additional notations are used to distinguish between these layers. Specifically, the number of the layer is appended as a superscript to the names for each of these variables. Thus the weight matrix of the first layer is written as W^1 and the weight matrix for the second layer as W^2. This notation is used in the three-layer network as shown in Figure 3.5.

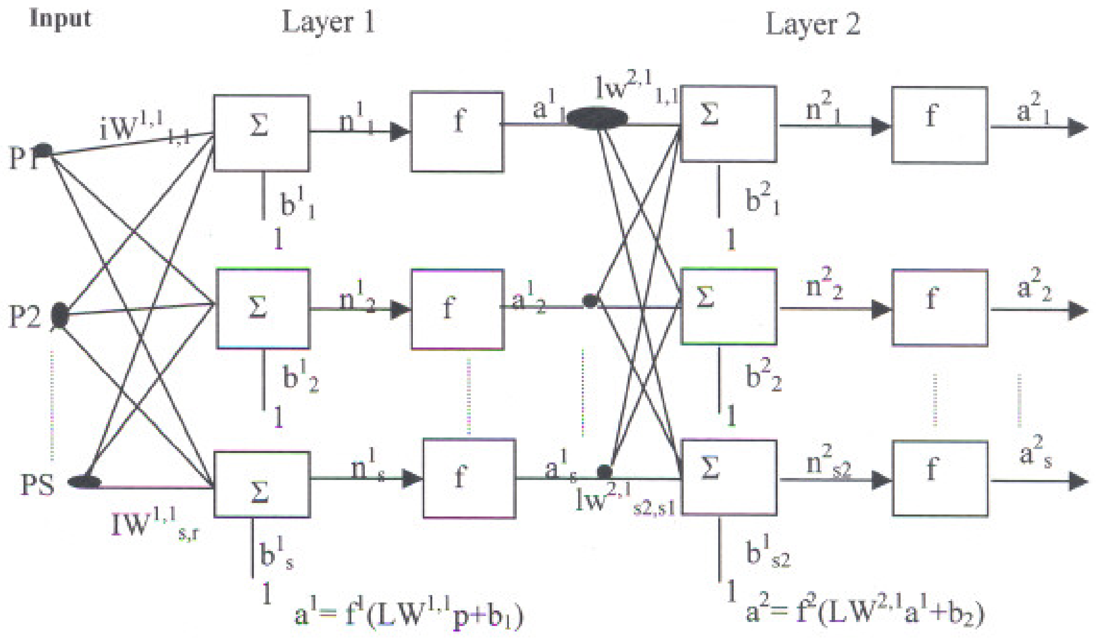


Figure 3.5 Multiple Layer Neurons

As shown there are R inputs, S^1 neurons in the first layer, S^2 neurons in the second layer, etc. Different layers may have different number of neurons. The outputs of layers one and two are the inputs for the layers two and three, respectively.

Thus layer two can be viewed as a one layer network with $R = S^1$inputs, $S=S^2$ neurons, and an $S^1 \times S^2$ weight matrix W^2. The input to layer two is a^1, and output is a^2. A layer whose output is the network output is called output layer while the other layers are called hidden layers.

3.1.7 Type of Networks as per Interconnection Scheme

According to the interconnection scheme, there are two types of networks i.e. Feedforward and Feedback/Recurrent networks [16].

3.1.7.1 Feedforward networks

In this type of network, the information flow is from the input layer towards the output layer i.e. in a single direction as shown in Figure 3.6. Each output is a constant and a function of network input only. This type of network is used in applications where the propagation and combination of activation is performed successively in forward direction from the input layer to the output layer characterizing the inference behavior.

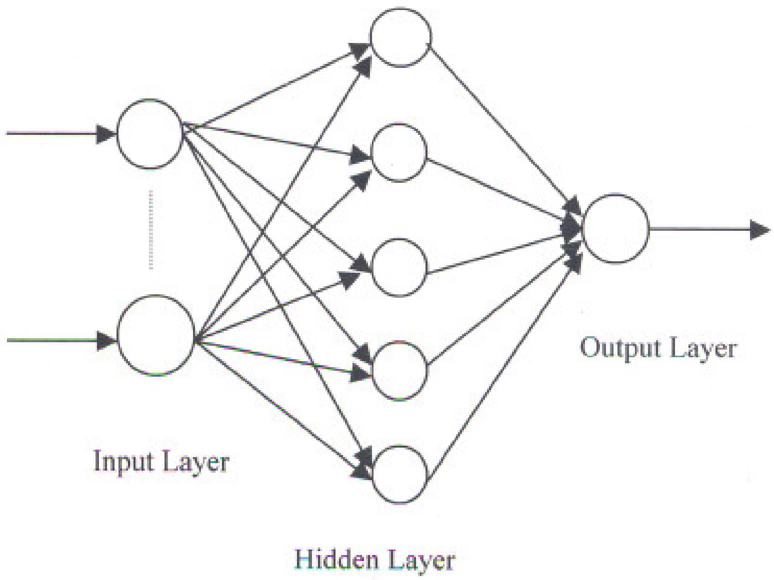


Figure 3.6 Feedforward network

3.1.7.2 Feedback/Recurrent Networks:

A feedback/Recurrent distinguishes itself from a feedforward neural network in that it has at least one feedback loop. The presence of feedback loops has a profound impact on the learning capability of the network and on its performance. These networks are potentially more powerful than the feedforward network, since they are able to recognize and recall the temporal as well as spatial pattern. Output of these networks is a function of time.

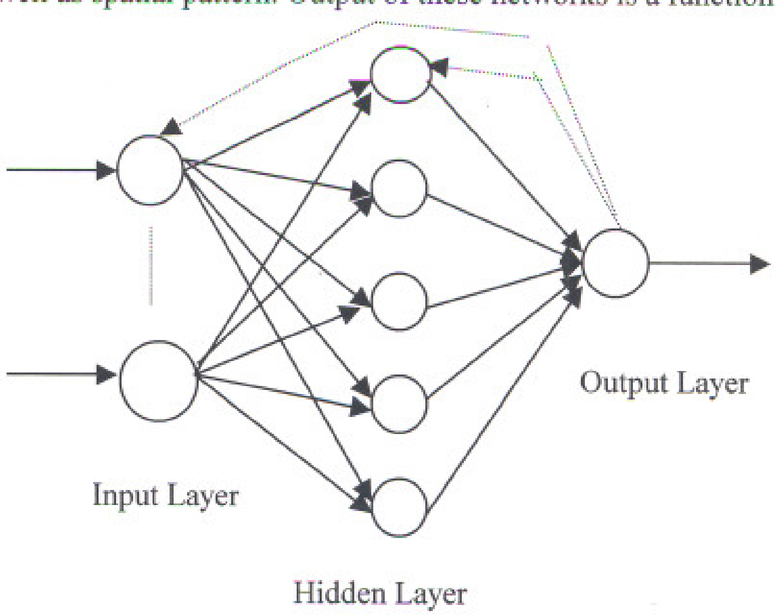


Figure 3.7 Feedback/Recurrent network

3.1.8 Learning in Artificial Neural Networks

The property that is of primary significance for a neural network is the ability of the network to learn from its environment, and to improve its performance through learning [14].In ANN, learning is the process by which the neural network adapts itself to a stimulus, and eventually it produces a desired response. During the learning process, the network adjusts its parameters, the synaptic weights, in response to an input stimulus so that its actual output response converges to the desired output response. When the actual output response is the same as the desired one, the network has completed the learning phase; i.e. it has "acquired knowledge."

As different learning methodologies suit different people, so do different learning techniques suit different artificial neural networks. Thus, the neural networks are commonly categorized in terms of their corresponding training algorithms [16]:

- **Supervised Learning**

Supervised learning networks have been the mainstream of neural model development. In this, the learning rule is provided with a set of examples (the training set) of proper network behavior. As the input stimulus is applied to the network, the network outputs are compared to the target response. If the actual response differs from the target response, the neural network generates an error signal, which is then used to calculate the adjustment that should be made to the network's synaptic weights so that the actual output matches the target output, thereby, minimizing the error. Thus, the supervised learning can be considered as the learning under guidance.

- **Unsupervised Learning**

In an unsupervised learning rule, the training set consists of input training patterns only and the synaptic weights are modified in response to network inputs. During the training session, the neural net receives at its input many different excitations and it arbitrarily organizes the patterns into categories. When a stimulus is later applied, the neural net provides an output response indicating the class to which the stimulus belongs. Most of these algorithms perform some kind of clustering operation. Thus, the unsupervised learning can be considered as the learning without guidance.

➢ **Reinforced Learning**

Reinforcement Learning is similar to the supervised learning, except that, instead of indicating the closeness of actual output to the desired output, it indicates whether the actual output is the same as the target output or not. During the learning phase an input stimulus is applied and an output response is obtained in the form of a grade. This grade is the measure of the network performance over some sequence of outputs. This type of learning appears to be most suited to control system applications.

➢ **Competitive Learning**

Competitive Learning is another form of supervised learning that is distinctive because of its characteristic operation and architecture. In this, there are several neurons at the output layer. When an input stimulus is applied, each output neuron competes with the others to produce the closest output signal to the target. This output becomes the dominant and the other outputs cease producing an output signal for that stimulus. For another stimulus, another output neuron becomes the dominant one and so on. Thus each output neuron is trained to respond to a different input stimulus.

3.1.9 Important ANN Parameters

The performance of the neural network is described by the figure of merit, which expresses the number of recalled patterns when input patterns are applied that are complete, partially complete, or noisy. A 100% performance in recalled patterns means that for every trained stimulus signal, the network always produces the desired, or target, output pattern [16]. When designing an artificial neural network, one should be concerned with the following:

- Network Topology
- Number of Layers in the Network
- Number of neurons per layer
- Learning algorithm to be adopted (supervised learning)
- Number of iterations per pattern during training
- Number of calculations per iteration
- Speed to recall a pattern
- Network Performance
- Degree of adaptability of the ANN

- Bias terms
- Boundaries of the synaptic weights

3.1.10 Neural Network Paradigms

The characteristics of Artificial Neural Networks, inspired from the biological world, are known as Paradigms [17]. The search for the best representative paradigms- the paradigm that truly emulates the biological neural network- is still underway. However the true power and advantage of neural networks lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modeled.

A typical paradigm is structured in layers of neurons. Some have one Layer, Single layer neural network (SLNN), while others may have multiple layers (MLNN). MLP is discussed briefly here.

Multilayer Perceptron (MLP)

The most common neural network model is the multilayer perceptron (MLP). This type of neural network is known as a supervised network because it requires a desired output for learning. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown. Figure.3.8 [23] shows the representation of MLP.

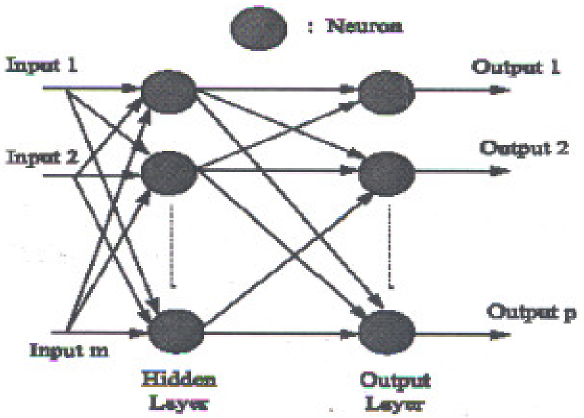


Figure 3.8 Multilayer Perceptron

The inputs are fed into the input layer and get multiplied by interconnection weights as they are passed from the input layer to the first hidden layer. Within the first hidden layer, they get summed then processed by a nonlinear function (usually the hyperbolic tangent). As the processed data leaves the first hidden layer, again it gets multiplied by interconnection weights, then summed and processed by the second hidden layer. Finally the data is multiplied by interconnection weights then processed one last time within the output layer to produce the neural network output.

3.1.11 Training of Network

The training basically involves feeding training samples as input vectors through a neural network, calculating the error of the output layer, and then adjusting the weights of the network to minimize the error. Each "training epoch" involves one exposure of the network to a training sample from the training set, and adjustment of each of the weights of the network once layer by layer. Selection of training samples from the training set may be random, or selection can simply involve going through each training sample in order [17].

Backpropagation Algorithm

The MLP and many other neural networks learn using an algorithm called backpropagation. With backpropagation, the input data is repeatedly presented to the neural network and with each presentation the output of the neural network is compared to the desired output and an error is computed. This error is then fed back (backpropagated) to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output.

Back propagation works in two passes

- **Forward pass:** In this pass, weights are fixed, input signals propagated through network and outputs are calculated. Output o_i is compared with the desired output d_i and an error signal $e_i = o_i - d_i$ is calculated.
- **Backward pass:** Starts with the output layer and recursively computes the local gradient δ_i for each node. Then the weights are updated and the process of forward pass is repeated again.

Algorithm

Let Z be an input matrix and D be an output matrix.

P be the training pairs i.e. { z_1 d_1 ,z_2 d_2 , ...z_p d_p}.

z_i (I*1) be an input vector and d_i (K*1) be an output vector , where $i=1,2,\ldots\ldots\ldots p$.

The i^{th} component of each z_i is -1,and

J is the number of neurons in the hidden layer.

Step1. Weights are initialized to a small random number.

Let W(K*J) be the weight matrix for output layer and V(J*I) be the weight matrix for hidden layer.

Step 2. Choose maximum acceptable error e_{max} and learning rate $\eta>0$.

Initialize e=0.

Step 3. Output for hidden layer is calculated as:

$$Y_j \leftarrow f(y_j z), \qquad \text{for } j=1,2,\ldots\ldots\ldots J \tag{iv}$$

Output for output layer is calculated as:

$$O_j \leftarrow f(w_j y) \qquad \text{for } k=1,2,\ldots\ldots\ldots\ldots K. \tag{v}$$

where f is a sigmoidal function.

Step 4. Error signal is computed.

$$e \leftarrow \frac{1}{2}(d_k - o_k)^2 + e \quad , \text{for } k=1,2,\ldots\ldots\ldots K. \tag{vi}$$

Step 5. Local gradient for output later is calculated as:

$$\delta_o = (1/2)(d_j - o_j)(1 - o_k), \quad \text{for } k=1,2,\ldots\ldots\ldots K. \tag{vii}$$

Local gradient for hidden layer is calculated as:

$$\delta_Y = \frac{1}{2}\left(1 - y_j^{\,2}\right)\sum_{k=1}^{k} \delta_{ok} w_{kj}, \qquad \text{for } j=1,2,\ldots\ldots\ldots J. \tag{viii}$$

Step 6. Output layer weights are adjusted.

$$w_{kj} \leftarrow w_{kj} + \eta\delta_{ok} y_j \qquad \text{for } k=1,2,\ldots\ldots\ldots K. \tag{ix}$$

$$\text{for } j=1,2,\ldots\ldots\ldots J.$$

Step 7. Hidden layer weights are adjusted.

$$v_{ji} \leftarrow v_{ji} + \eta\delta_{yj} z_i \qquad \text{for } j=1,2,\ldots\ldots\ldots J. \tag{x}$$

$$\text{for } i=1,2,\ldots\ldots\ldots J.$$

Step 8. If $e<e_{max}$ then terminate training process.

Final values of the weights w, v and e, are recorded.

If $e>e_{max}$, then e=0 and initiate a new training cycle by going to step 3.

Assumptions

- The values of the parameters is not set to zero because the origin of parameter space tends to be a saddle point for the performance space.
- The values of the initial parameters is also not set too large because the performance surface tends to have very flat region as one moves far away from the optimum point.

Hence, small random values must be chosen so that one may stay away from a possible saddle point at the origin without moving out to the very flat regions of the performance surface.

Step Size

Let η be the step size or the learning rate. When η is too small then algorithm become unstable and if η is too large then algorithm either diverges or oscillates. So maximum stable learning rate is limited by largest eign value (second derivative) of the Hessian matrix.

$$\eta < 2/\lambda_{max}$$, where λ_{max} is the largest eign value.

Generally $\eta < 1/2$ for convergence.

Termination of process

- After a certain number of epochs through all the training data (a run through all the training data is called an epoch), the process is terminated, or
- When the total sum squared error reaches at some low level, then the process is terminated.

Advantages

- Guaranteed to converge to hypothesis with local minimum error.
- Training data is not required to be necessarily linearly separable.
- Convergence to the single global minimum.

Disadvantages

- Slow convergence.
- There is no guarantee of converging to global minimum when there are multiple local minima's.

- Need for teacher: Supervised learning.

Improvement in Gradient Descent

After the back propagation algorithm was first introduce in its pure form, it was found that it is quite slow. This is of course no surprise, since the steepest descent algorithm is known for its low speed, especially close to minimum (since gradient is disappearing). One of the reasons is that it uses the fixed step size. In order, to take into account the changing curvature of the error surface, many optimization algorithms use steps that vary with each iteration.

Another reasonable idea, also used in optimization, is to keep the minimization process going in the same general direction not allowing too much "wagging" as the steepest descent algorithm often does and jumps out of a local minimum. All this involves adding a term to weight adjustment that is proportional to the amount of previous weight change. That is, the previous adjustment is "remembered" and used for the modification of the next change. Such term is called a "momentum" term.

$$\Delta w_s(N+1) = -\eta \frac{\partial E}{\partial w_s} + \mu \Delta w_s(N) \qquad \text{(xi)}$$

μ is momentum coefficient (its value exist between 0.1 to 0.9)

The benefit of this term is that it tends to make the trajectory continue in the same direction and speeds up the convergence.

3.2 Introduction to Fuzzy Logic

Many decision-making and problem solving tasks are too complex to be understood quantitatively, however people succeed by using knowledge that is imprecise rather than precise. Fuzzy set theory, originally introduced by Zadeh in the 1960's, resembles human reasoning in its use of approximate information and uncertainty to generate decisions. It was specifically designed to mathematically represent uncertainty and vagueness and provide formalized tools for dealing with the imprecision intrinsic to many problems. Since knowledge can be expressed in a more natural way by using fuzzy sets, many engineering and decision problems can be greatly simplified [18].

Fuzzy logic is a powerful problem solving methodology with a myriad of applications in embedded control and information processing. Fuzzy provides a remarkably simple way to draw definite conclusions from vague, ambiguous or imprecise information. In a sense, fuzzy logic resembles human decision making with its ability to work from approximate data and find near precise solutions.

Unlike classical logic, which requires a deep understanding of a system, exact equations, and precise numeric values, Fuzzy logic incorporates an alternative way of thinking, which allows modeling complex systems using a higher level of abstraction origination from our knowledge and experience. Fuzzy logic allows expressing this knowledge with subjective concepts such as very hot, bright red, and a long time, which are mapped into exact numeric ranges.

3.2.1 Basic Fuzzy Logic System

The general fuzzy system consists of a fuzzification, fuzzy rule base, fuzzy inference engine and defuzzification modules. It operates by repeating a cycle of the following steps.

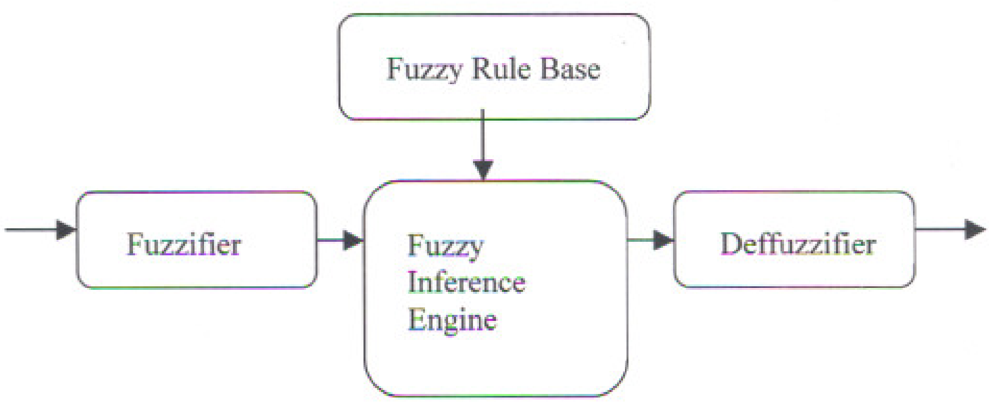


Figure3.9 Block Diagram of FLS

Fuzzification: The fuzzification interface involves the following functions:

- Measures the value of input variables.
- Perform a scale mapping that transfer the range of values of input variables into corresponding universes of discourse.
- Perform the function of fuzzification that converts input data into suitable linguistic values, which may be viewed as labels of fuzzy sets.

Knowledge base: A fuzzy system is characterized by a set of linguistic statements based on expert knowledge. The expert knowledge is usually in the form of "if –then " rules which are easily implemented by fuzzy conditional statements in fuzzy logic. The collection of fuzzy control rules that are expressed as fuzzy conditional statements forms the rule base or the rule set of an FLC.

It comprises knowledge of application domain and the attendant Control goals. It consists of a "Data Base" and a "Linguistic Fuzzy Control Rule Base".

- The database provides the necessary definitions, which are used to define linguistic control rules and fuzzy data manipulations in an FLC.
- The rule base characterizes the control goals & control policy of the domain experts by means of a set of linguistic control rules.

Defuzzification: It performs the following functions:

- Scale mapping, which converts the range of values of output variables into corresponding universes of discourse,
- Defuzzification, which yields a non-fuzzy control action from an inferred fuzzy control action.

There are three types of FLS usually used. These are Sugeno, Mamdani and Larsen types of systems. We have used Mamdani type systems for our design. Mamdani type FLS are different from Sugeno type in a way that in Mamdani type systems consequents are fuzzy sets whereas in Sugeno type systems the consequents are Fuzzy Singletons.

3.3 Neural Encoded Fuzzy Logic Algorithm

In this, a new class of handoff algorithms that combines attractive features of several existing algorithms and adapts the parameters of a handoff algorithm using a neural encoded fuzzy logic system is proposed. The system parameters can be used to design a fuzzy logic system (FLS). Thereafter, the neural encoding of the FLS is done i.e. a neural network learns how the FLS works.

Since neural networks can represent information compactly, good savings in storage and computational requirements can be obtained if a neural network replaces the fuzzy logic rule base. The MLP model is trained to learn the relationship among the inputs and the outputs of the fuzzy logic rule base. The trained neural networks are thus used to adapt the parameters of a handoff algorithm.

VHDL Implementation of Neural Networks

In this chapter the basic requirement and the design procedure of the hardware description language i.e. VHDL is discussed briefly along with the language overview. Some introduction about the FPGAs is also included. Finally, the procedure for implementing the ANNs into VHDL is discussed.

4.1 Historical Development

The development of VHDL was initiated in 1981 by the United States Department of Defense to address the hardware life cycle crisis. The cost of reproducing electronic hardware as technologies became obsolete was reaching crisis point, as the function of the parts was not adequately documented, and the various components making up a system were individually verified using a wide range of different and incompatible simulation languages and tools [19]. This led to the requirement of a language with a wide range of descriptive capability that would work the same on any simulator and was independent of technology or design methodology. Figure 4.1 [23] shows the basic layout of the chip.

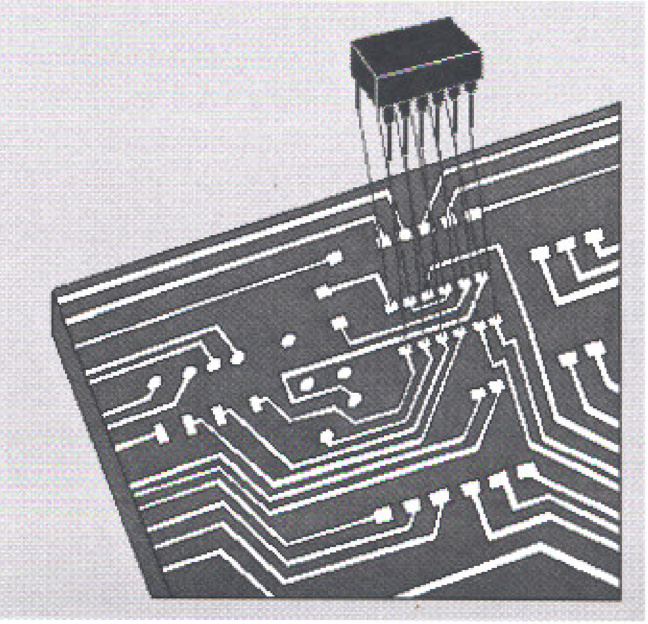


Figure 4.1 Layout of a Chip

4.2 Introduction

VHDL is an acronym for VHSIC Hardware Description Language where VHSIC is yet another acronym which stands for Very High Speed Integrated Circuits. It can describe the behavior and structure of electronic systems, but is particularly suited as a language to describe the structure and behavior of digital electronic hardware designs, such as ASICs and FPGAs as well as conventional digital circuits.

4.3 Language Overview

VHDL has many features appropriate for describing the behavior of electronic components ranging from simple logic gates to complete microprocessors and custom chips. Features of VHDL allow electrical aspects of circuit behavior (such as rise and fall times of signals, delays through gates, and functional operation) to be precisely described. The resulting VHDL simulation models can then be used as building blocks in larger circuits (using schematics, block diagrams or system-level VHDL descriptions) for the purpose of simulation [19].

VHDL is also a general-purpose programming language: just as high-level programming languages allow complex design concepts to be expressed as computer programs, VHDL allows the behavior of complex electronic circuits to be captured into a design system for automatic circuit synthesis or for system simulation. VHDL syntax is like other well-known languages like Pascal, C and C++, but its behavior is completely different. These high-level languages are adapted to a CPU, i.e. a serial machine that performs one instruction at a time, while VHDL is adapted to general structures in hardware. The structure of hardware is largely parallel and hence, we may say that practically performance is always several degrees better with VHDL than that with "conventional" programming languages for CPUs. Thus, VHDL is used mainly as an implementation language and not as software [20].

Moreover, Like Pascal, C and C++, VHDL includes features useful for structured design techniques, and offers a rich set of control and data representation features. Unlike these other programming languages, VHDL provides features allowing concurrent events to be described. This is important because the hardware described using VHDL is inherently concurrent in its operation.

One of the most important applications of VHDL is to capture the performance specification for a circuit, in the form of what is commonly referred to as a test bench.

Test benches are VHDL descriptions of circuit stimuli and corresponding expected outputs that verify the behavior of a circuit over time. Test benches should be an integral part of any VHDL project and should be created in tandem with other descriptions of the circuit.

4.4 Standardization

During the early 80s a number of U.S. companies were involved in designing VHSIC chips for the Department of Defense (DOD), However, at that time, most of the companies were using different hardware description languages (HDL) to describe and develop their integrated circuits. As a result, different vendors could not effectively exchange designs with one another. Also, different vendors provided DOD with descriptions of their chips in different HDL. Thus, a need for a standardized HDL for the design, documentation, and verification of digital systems was generated. Consequently, VHDL was transferred to the IEEE for standardization and thus, became the standard language [21]. This language has also been recognized as an American National Standards Institute (ANSI) Standard. Using a standard language also means that the users can also take the advantage of the most up-to-date design tools and have access to a knowledge base of thousands of other engineers.

4.5 Features

The usage of VHDL for different projects definitely increases the speed and efficiency of the system. VHDL (like a structured software design language) is most beneficial when used in a structured, top-down approach. Moreover, a library of reusable VHDL components can be accumulated and can then be used to increase the productivity. This increase in productivity can also occur when VHDL is used to enhance the communication between the team members and the more powerful tools available for simulation and design verification. In addition, VHDL allows the user to design at a more abstract level. Instead of focusing on a gate-level implementation, one can address the behavioral function of the design.

Moreover, with the rapid pace of development in Electronic Design Automation (EDA) tools and in target technologies, a standard language such as VHDL can greatly improve the chances of moving into more advanced tools (for example, from a basic low-cost simulator to a more advanced one) without having to re-enter the circuit descriptions.

Ones ability to retarget circuits to new types of device targets (for example, ASICs, FPGAs, and complex PLDs) is also improved by using a standard design entry method.

4.6 Design process

The Figure 4.2 [23] shows a very simplified view of the electronic system design process incorporating VHDL. The central portion of the diagram shows the parts of the design process which are most impacted by VHDL.

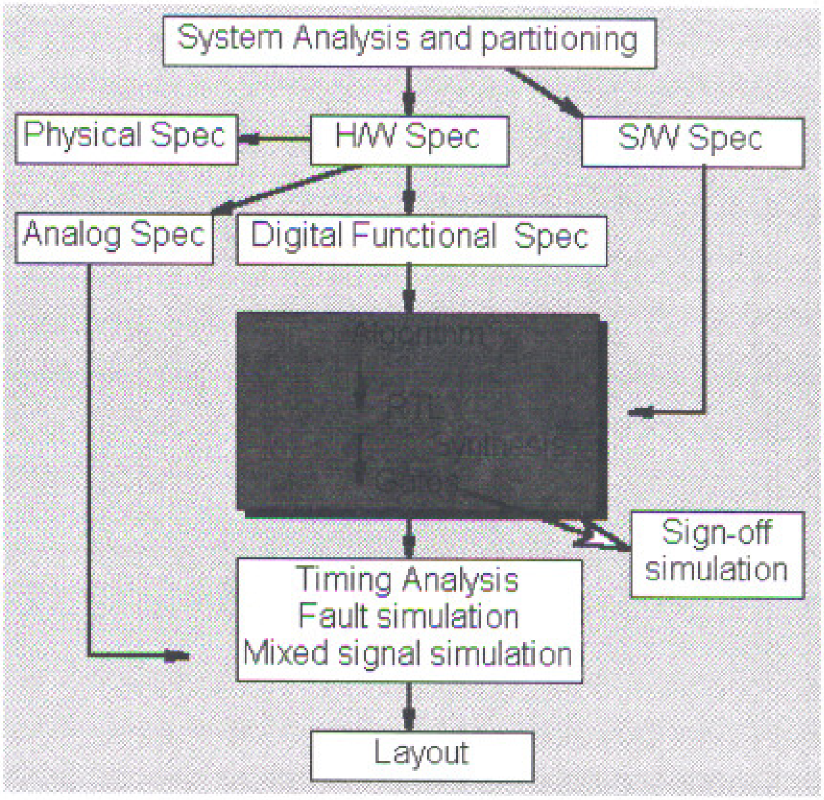


Figure 4.2 View of the electronic system design process

Thus, VHDL is used to describe a model for a digital hardware device. This model specifies the external view of the device and one or more internal views. The internal view specifies the functionality or structure, while the external view specifies the interface of the device through which it communicates with the other models in its environment. An entity declaration is one that defines the inputs and outputs-the ports-of this circuit; and an architecture declaration is the one that defines what the circuit actually does, using a single concurrent assignment.

Thus, every VHDL design description consists of at least one entity/architecture pair. This combination of an entity and its corresponding architecture is also referred to as a design entity and has five different types of primary constructs, called design units [21].

4.6.1 Entity Declaration

An entity declaration describes the circuit as it appears from the "outside" - from the perspective of its input and output interfaces. It provides the complete interface for a circuit. Using the information provided in an entity declaration (the names, data types and direction of each port), one have all the information needed to connect that portion of a circuit into other, higher-level circuits, or to develop input stimuli (in the form of a test bench) for verification purposes. The actual operation of the circuit, however, is not included in the entity declaration.

4.6.2 Architecture declaration and body

The second part of a minimal VHDL design description is the architecture declaration. Before simulation or synthesis can proceed, every referenced entity in a VHDL design description must be bound with a corresponding architecture. The architecture describes the actual function-or contents-of the entity to which it is bound.

4.6.3 Configuration Declaration

A configuration statement is used to bind a component instance to an entity-architecture pair. It can be considered like a parts list for a design.

4.6.4 Package Declaration and Body

Package declaration encapsulates a set of related declarations, such as type declarations, subtype declarations, and subprogram declarations, which can be shared across two or more design units. A package body contains the definitions of subprograms declared in a package declaration.

Finally, when the design entity is ready, it is implemented on ASIC or FPGA, depending upon the size of the circuit, so as to get it on the chip.

4.7 Implementation of VHDL on FPGAs

Perhaps the most promising approach to emulating neural models digitally is the Field Programmable Gate Array, FPGA. Briefly, an FPGA is a device with a large number of generic logic blocks and generic interconnection between those blocks. The functions the logic blocks implement and how these blocks are connected to one another is determined by configuration bits that are loaded into the chip as one would load a program into a computer's memory. It is now possible to buy FPGAs that are capable of emulating millions of logic gates at frequencies approaching several hundred megahertz. There are very sophisticated design tools like Xilinx, Synplify, and Synopsis etc. that allows logic to be expressed in a high-level hardware description language and then be converted to FPGA configuration bits by an automated synthesis process [24].

FPGAs comprise an array of uncommitted circuit elements, called logic blocks, and interconnect resources. The FPGA configuration is performed through the programming by the end user. An illustration of a typical FPGA architecture appears in Figure 4.3 [24]. As the only type of Field Programmable Device (FPD) that supports very high logic capacity, FPGAs have been responsible for a major shift in the way digital circuits are designed.

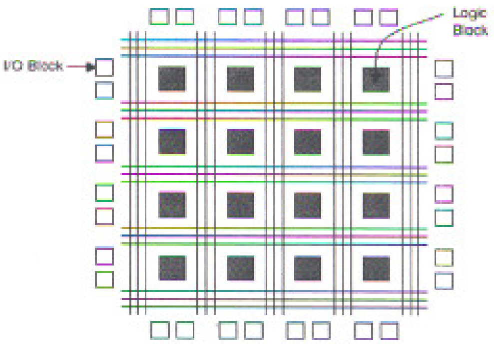


Figure 4.3 Structure of an FPGA

4.7.1 Computer Aided Design (CAD) Flow for FPDs

When designing circuits for implementation in FPDs, it is essential to employ Computer-Aided Design (CAD) programs. A typical CAD system includes the software for the following tasks: initial design entry, logic optimization, device fitting, simulation, and configuration. Design entry may be done either by creating a schematic diagram with a graphical CAD tool, by using a text based system to describe a design in a simple hardware description language, or with a mixture of design entry methods. Since initial logic entry is not usually in an optimized form, algorithms are employed to optimize the circuits, after which additional algorithms analyses the resulting logic equations and "fit" them into the FPDs. Simulation is used to verify correct operation, and the user would return to the design entry step to fix errors. When a design simulates correctly it can be loaded into a programming unit and used to configure an FPD. While the original design entry step is performed manually by the designer, all other steps are carried out automatically by most CAD systems.

Now days, more sophisticated tools are employed for implementing circuits in FPDs. Since the devices like FPGAs are a bit more complex and can accommodate large designs, it is more common to use a mixture of design entry methods for different modules of a complete circuit. These modules are described via a full-featured hardware description language such as VHDL. In the "device fitter" step, FPGAs require at least three steps: a technology mapper to map from basic logic gates into the FPGA's logic blocks, placement to choose which specific logic blocks to use in the FPGA, and a router to allocate the wire segments in the FPGA to interconnect the logic blocks. The CAD tools however may require a long period of time (often more than an hour or even several hours) to complete their tasks.

4.8 Implementation of ANN in VHDL

The inherent parallelism in each FPGA, allows the parallel implementation of the various structures within the neuron. This further requires the implementation of neural networks in VHDL. All the different types of neurons i.e. linear, sigmoid, hard limit etc. are designed in VHDL. The implementation of the linear and hard limit functions is simple and straightforward, but to implement the sigmoid transfer function is a difficult task due to the various limitations posed by the digital circuits. For this, a lookup table is prepared which gives the desired response.

In the work carried out, a generalized model of the neural networks is designed in VHDL Beginning with the structure of a single neuron, we move towards the entire network. The structure of a single neuron is implemented by using the following basic steps:

- Make an entity of the neuron describing its input and output signals.
- Define the architecture of the entity i.e. working of the neuron.
- The product of the inputs to the neuron and the weights is computed and the desired bias value is added to it.

- Finally, this sum is passed through the desired activation function.

Then, all these neurons are connected in a proper manner so as to form the desired network. The network so developed is a general model and can be used to implement any structure irrespective of the type of the neuron and the algorithm used for training and calculating the weights and bias values.

Chapter 5

Implementation of Handoff in GSM

In this chapter a multicriteria handoff algorithm proposed in [11] is discussed. The algorithm takes into account absolute signal strength, relative signal strength with hystersis margin and time delay. Since all the three factors are considered simultaneously, the authors claimed that the algorithm is highly efficient and adaptive. Further we note that GSM standards allow only 150ms interruption in the service due to handover process [11]. Giving the due consideration to the above requirements and further realizing that GSM networks are popular for data transfer. We have further reduced the time delay to a maximum of 1 second. As against maximum of 5 seconds proposed in [11]. This reduction in time delay makes our algorithm more suitable for the high speed MS.

5.1 The Algorithm

A simple 2-cell model is considered. An MS initially served by BS0 is moving towards BS1 as shown in Figure 5.1. The handover decision is based on the aggregation of the following three criteria:

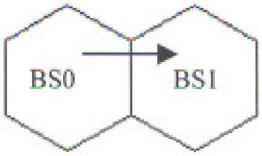


Figure 5.1 MS moving from BS0 to BS1

I. Absolute value of averaged RSS from the current BS drops below a threshold value.

Two thresholds are defined:

Minimum threshold (T_{min}). This is the trigger level for initiating handover process and is used for detecting the need for handover.

Signal Threshold (T_{drop}). It is the minimum acceptable signal level to maintain the call and should match the receiver sensitivity. At this level, the handover must be commanded; otherwise, the call will terminate prematurely. T_{min} is a few dBs above the T_{drop}.

The need for the handover is detected when

RSS0 < T_{min} , where RSS0 is the absolute value of the averaged RSS from serving BS. The Figure 5.6 shows the variation of RSS as MS moves from BS0 to BS1.

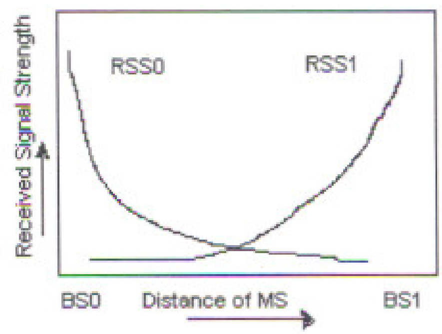


Figure 5.2 Variation of RSS as MS moves from BS0 to BS1

II. Averaged relative signal strength from the target BS exceeds that from the current BS by a hysteresis margin h.

RSS1 – RSS0 >h, where RSS1 is the averaged RSS from the target BS.

III. 3. The first two conditions persists for at least Γ seconds where Γ >= 1 sec.

In order to ensure that our design works well when implemented in ANNs, we simulated the design in MATLAB using Fuzzy Logic Toolbox. Though, the concept was working yet in order to arrive at the minute details of the system, the "design feasibility" study was carried out. We further used this design to generate the training data for our ANN to be designed.

5.2 Conceptual Design

5.2.1 Fuzzy Logic Approach

Using the FIS Editor, the three inputs and a single output FIS is framed.

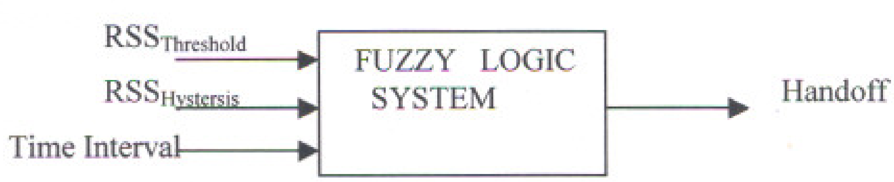


Figure 5.3. Fuzzy Logic Based Algorithm

The above discussed three parameters i.e. RSSthreshold, RSShysteresis and time interval are taken as the inputs to the FLS. The inputs are divided into four, three and two ranges (fuzzy sets) respectively as shown in figures 5.4, 5.5, 5.6,5.7. The output of the system is the decision regarding the need of Handoff. It is divided into three fuzzy sets, namely, no handoff (NO HO), make request (Mk RQ) and make handoff (Mk HO). The membership function with ranges is shown in Figure 5.8.

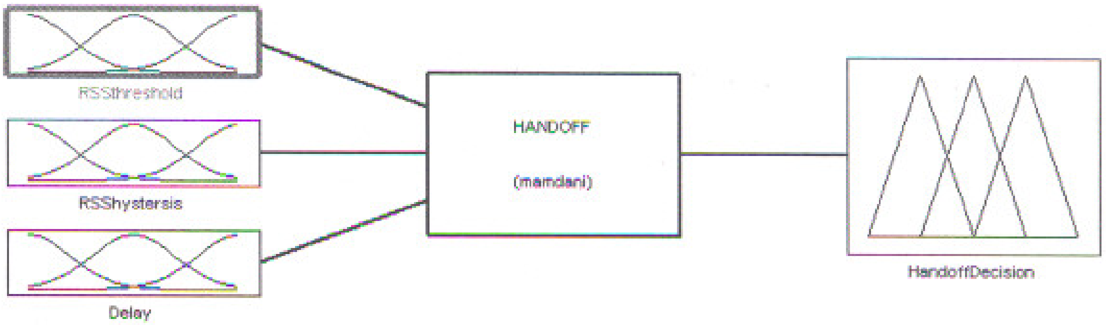


Figure 5.4 Basic FLS for HHO

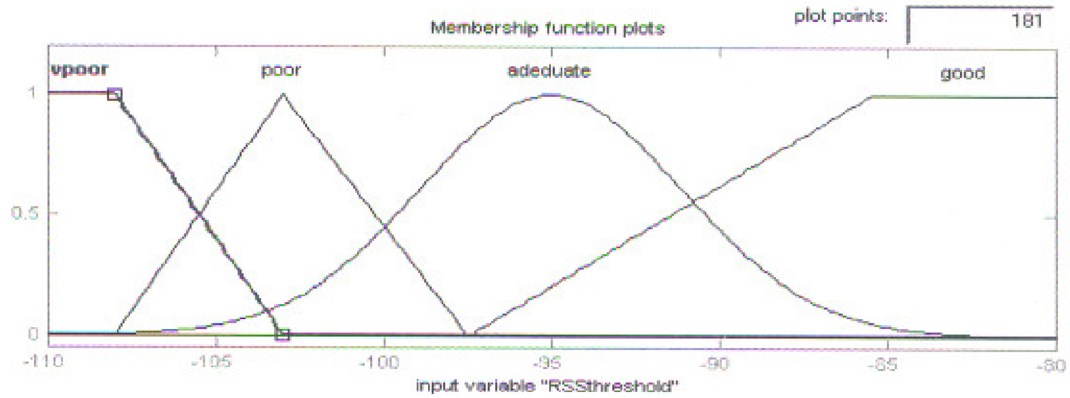


Figure 5.5 Membership function for $RSS_{Threshold}$

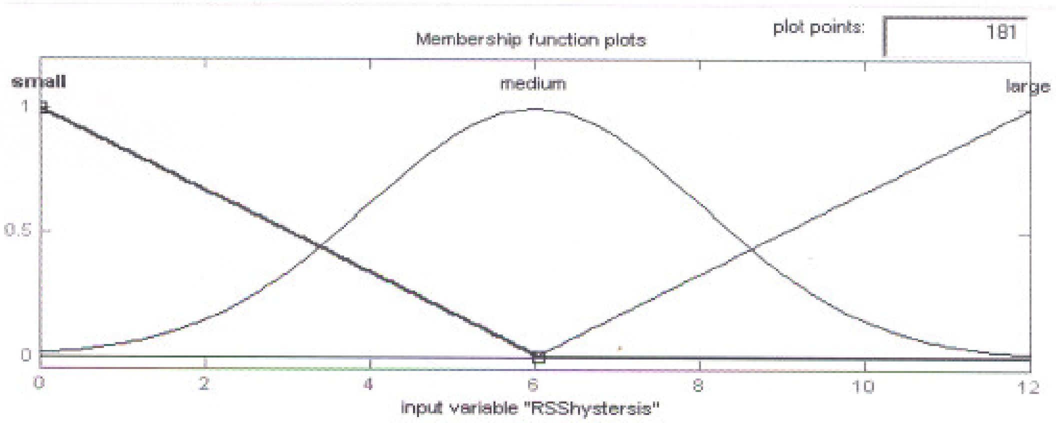


Figure 5.6 Membership function for $RSS_{Hystersis}$

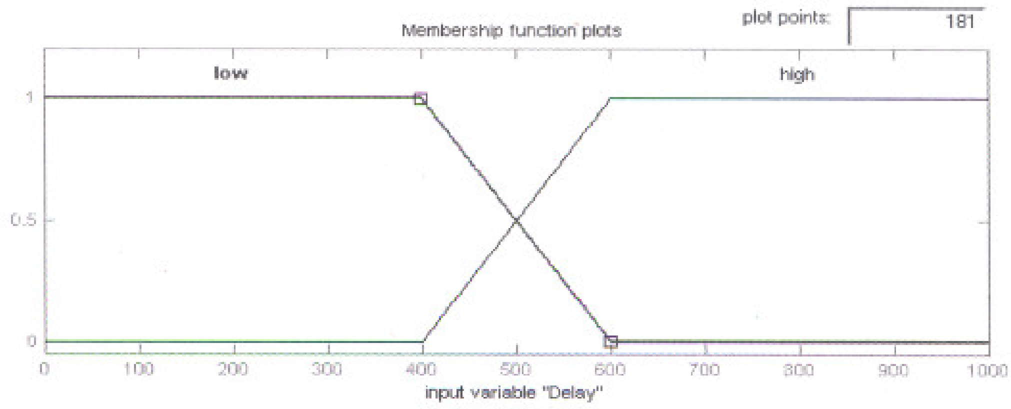


Figure 5.7 Membership function for Delay

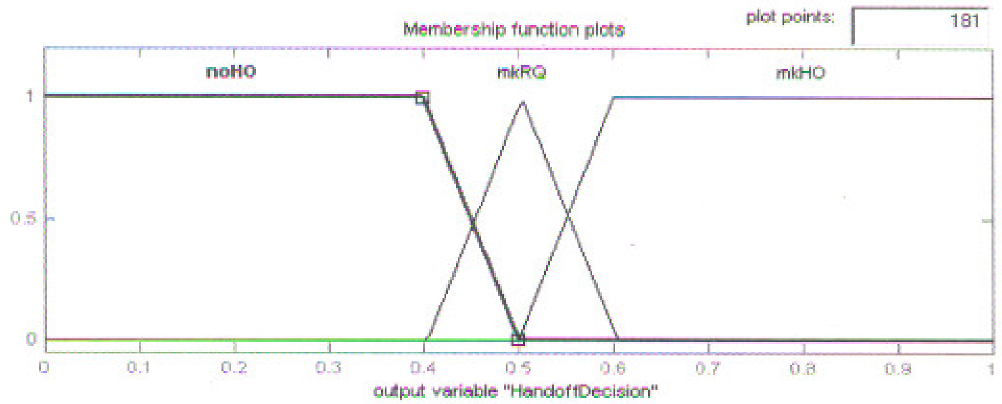


Figure 5.8 Membership function for Handoff Decision

For the simulation, we use the following operators.

- Rule Composition Operator: MIN
- Implication Operator: MIN (Mamdani Model)
- Defuzzification Technique: Weighted Average

The rule base for the designed system consists of 25 rules. The system makes the handover decision depending upon the values of the three inputs and the HO is initiated when the handover performance coefficient exceeds 0.5 on a 0 - 1 scale.

Table 5.1 System Rule Base

$RSS_{Threshold}$	$RSS_{Hysteresis}$	**Delay**	**Handover Decision**
V Poor	Small	Low	No HO
V Poor	Small	High	No HO
V Poor	Medium	Low	Mk RQ
V Poor	Medium	High	Mk HO
V Poor	Large	Low	Mk RQ
V Poor	Large	High	Mk HO
Poor	Small	Low	No HO
Poor	Small	High	No HO
Poor	Medium	Low	Mk RQ
Poor	Medium	High	Mk RQ
Poor	Large	Low	Mk HO
Poor	Large	High	Mk HO
Adequate	Small	Low	No HO
Adequate	Small	High	No HO
Adequate	Medium	Low	Mk RQ
Adequate	Medium	High	No HO
Adequate	Large	Low	Mk RQ
Adequate	Large	High	No HO
Good	Small	Low	No HO
Good	Small	High	No HO
Good	Medium	Low	No HO
Good	Medium	High	No HO
Good	Large	Low	No HO
Good	Large	High	No HO
None	Small	None	No HO

After the system structure and Rule base is specified, we simulate the system response. The system performance can be obtained as a single crisp value for a given set of crisp inputs as specified in figure 5.9(rule view). Here the 3 inputs are {-95, 6 & 500ms.} for which the output crisp value is 0.287, which gives an indication that the handoff should not be commanded for this particular set of inputs. Moreover, the system response can also be obtained as 3D graph as given in Figure 5.10.

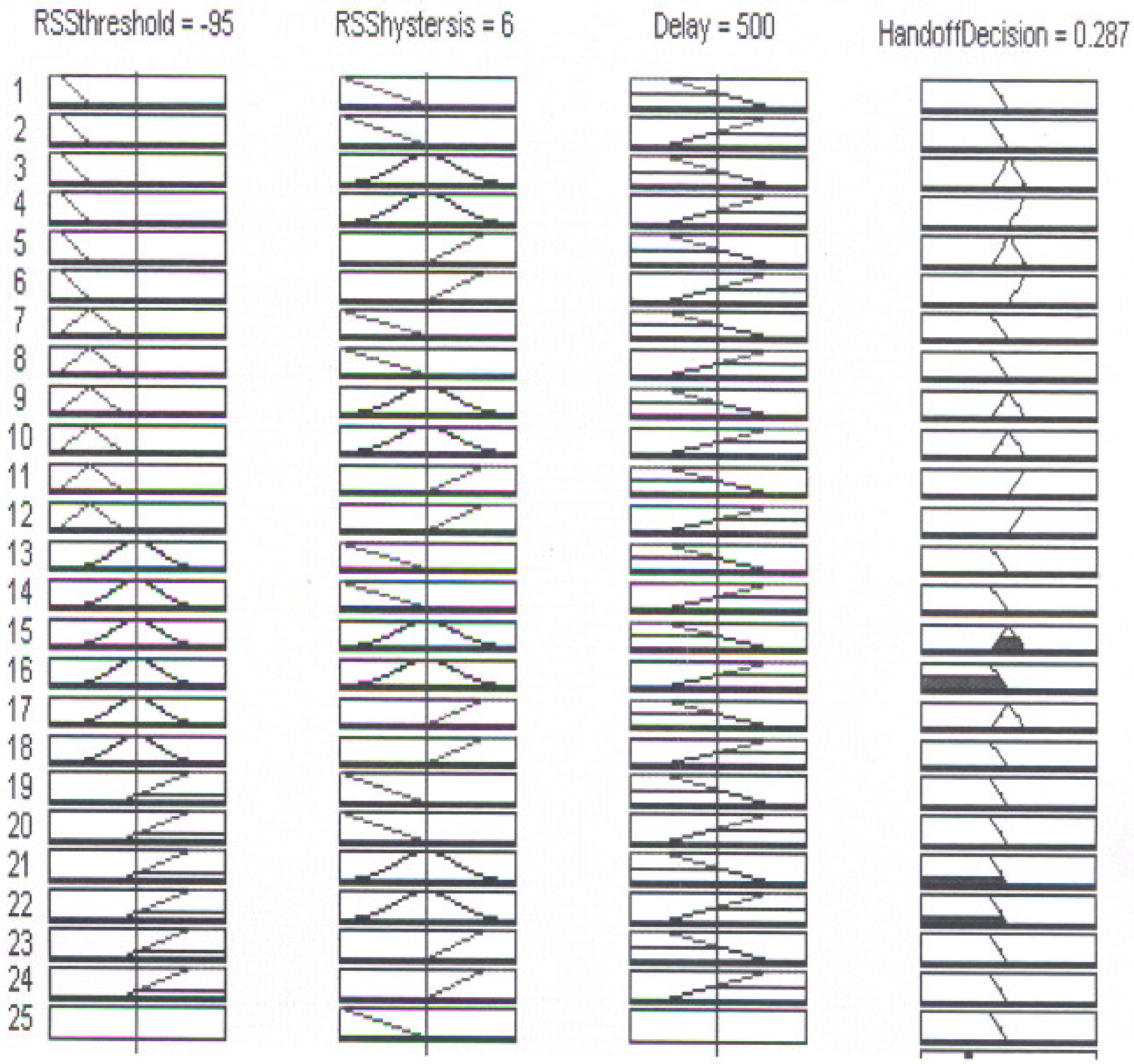


Figure 5.9 Rule view for the Handoff

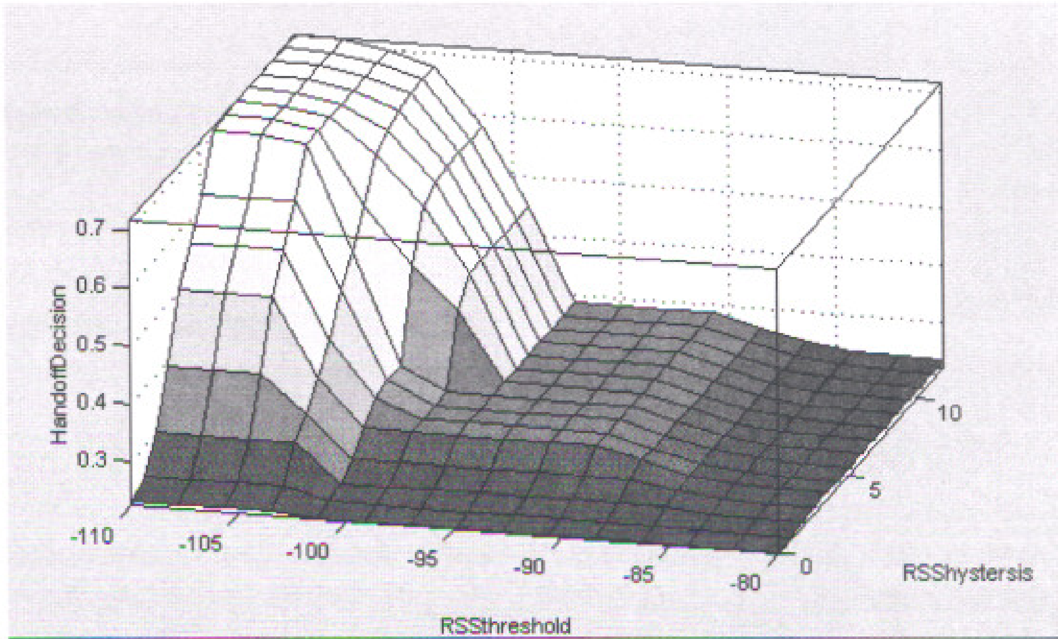


Figure 5.10 Surface view of FLS for 500ms.

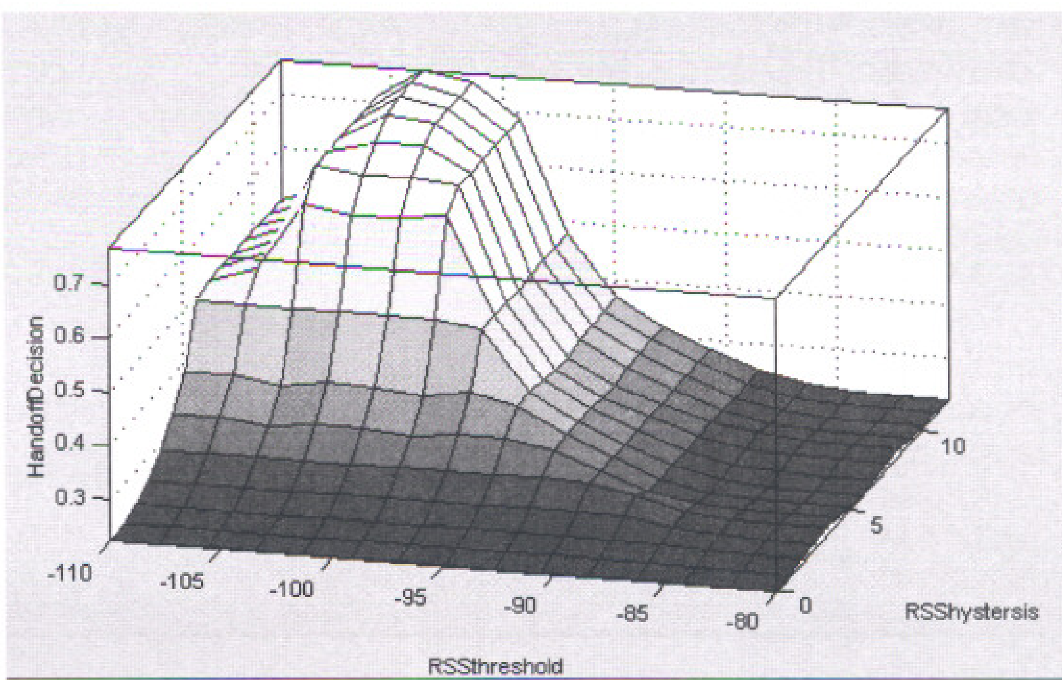


Figure 5.11Surface view of FLS for 200ms.

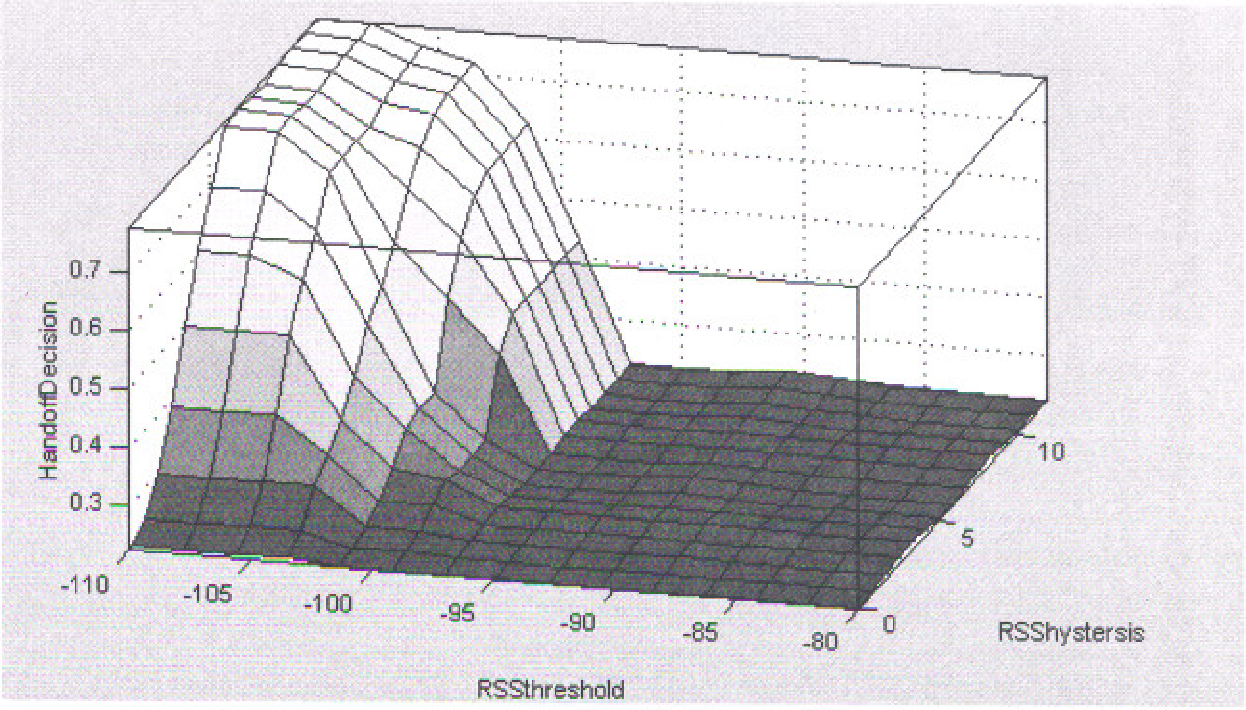


Figure 5.12 Surface view of FLS for 800ms.

Observations

Certain interesting observations can be inferred:

- For a delay of about 500ms, the handoff is not desired in the conditions where the threshold is greater than –95dB and hystersis margin can take any value. Also, if the hystersis margin is in the small range (< 2) and threshold takes any value from –110 dB to –90 dB, the HO is not initiated. But, if the threshold is very poor and large hystersis margin (>8) exists, then the handover must be initiated. Further, for the poor value of threshold and medium hystersis margin, the HO may be requested.

- For the delay of 800 ms., the results are found to be quite similar, except that the region for making HO request is bit increased and the region of NO HO is more flattened.

- For the low delay values like 200 ms., some more variations are observed. There is no need of HO if the hystersis is small for any value of threshold and if the threshold value is good enough despite of the hystersis margin. Further, if very poor value for threshold exists and the hysteresis is large enough, then the request for HO is generated. Moreover, for the poor and adequate values of threshold with sufficient hysteresis margin, the HO must be initiated.

The inherent parallelism in FLSs allows an efficient implementation of the fuzzy logic based algorithm. However, the algorithm is still much more complex than conventional algorithms that consist of only a few binary IF-THEN rules. Moreover, as the number of inputs to the FLS increases or as the universes of discourse for the fuzzy variables are divided into more fuzzy sets, the complexity of the FLS increases even further. A simple handoff algorithm with fewer computations and less storage requirements is desirable that can be executed quickly and does not consume a lot of the available resources. Thus, we use the Neural encoding of the Fuzzy Data.

5.2.2 ANN Approach

The basics of ANNs are discussed in chapter 3. An ANN can be trained to learn complex relationships among the inputs and the outputs of a system. After the ANN is trained, the parameters of the ANN can be used to estimate the outputs for given inputs. The training mechanism of the ANN in a supervised learning mode is used. A generic training procedure with its application to the FLS mapping is thus explained next.

- Get the data set that contains the system inputs and desired outputs. Different possible combinations of the inputs are applied to the FLSs, and the corresponding outputs of the FLS are calculated. The FLS inputs and outputs constitute the data set.
- Determine the structure of the neural network and the associated learning algorithm. Here the Feedforward backpropagation-learning algorithm is used.
- Determine the input set and the target output set for the neural network. Since the mapping between the FLS inputs and outputs is static, input and output data sets collected in Step 1 can serve as the input set and the target output set with proper scaling.
- Select the training parameters (such as the learning rate and number of neurons) and train the neural network using an appropriate algorithm. Once the network has been trained, the mapping between the FLS inputs and the corresponding outputs is stored in the parameters of the ANN.

The basic structure of the desired network is obtained from the NNTOOL of the MATLAB by training the FFBP net with the proper training data. The values of the weights and biases for each neuron are also taken from the same for the minimum error. Figure.5.13 shows the structure obtained with the best performance.

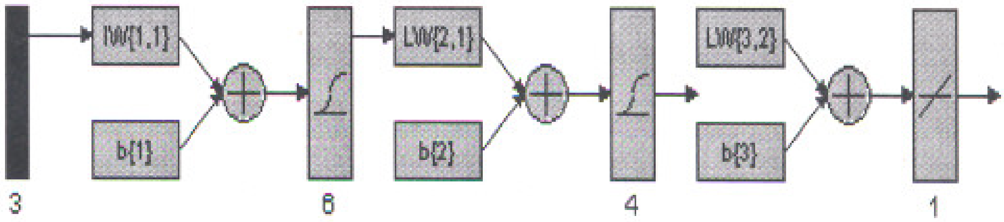


Figure5.13 Structure of desired network

5.2.3 Hardware Implementation

Once the structure of the Neural Network is obtained, it can be implemented in VHDL. This concept is also discussed in chapter 4. The basic parameters used in the neural network developed in VHDL includes the number of inputs to the system, the number of hidden layers used for the network, the number and type of neurons used in the hidden layers and the number of outputs of the system which constitutes the number of neurons in the output layer. For this system, discussed above, these parameters have values 3, 3, 6-logsigmoid, 4-logsigmoid and 1 respectively. The figure 5.14 shows the basic structure of the designed system.

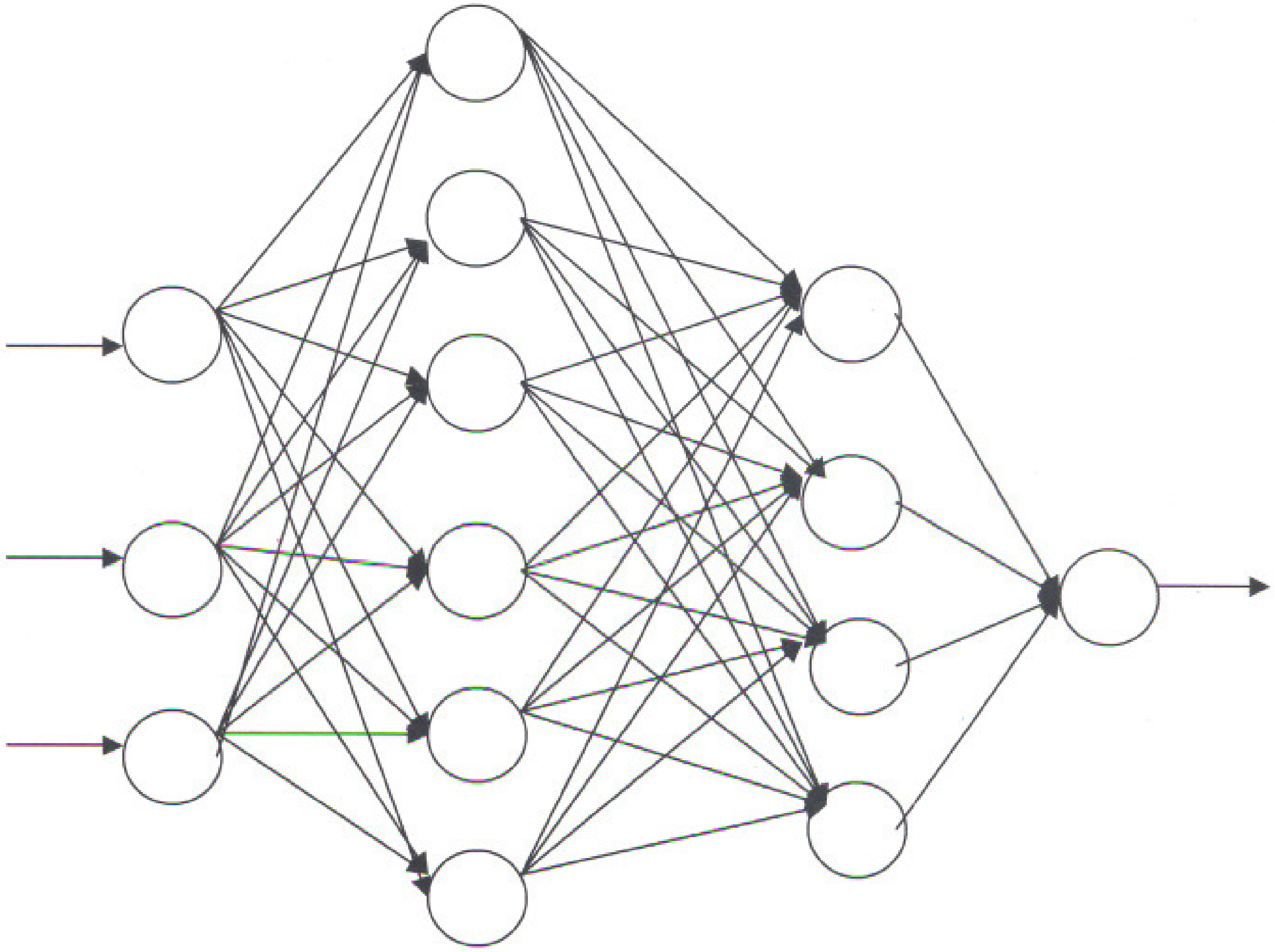


Figure 5.14 Detail structure of designed system

Simulation Results

The simulation results of the system so developed are shown in figure 5.15. The signals The signals 'P_in1<1>', 'P_in1<2>', 'P_in1<3>' are taken as the inputs for the network and the signal 'fin_res' is the final result, deciding whether the Handoff is desired for a particular set of input values or not. The value '1' of the output 'fin_res' corresponds to the need of handoff while '0' corresponds to the condition that handoff is not desired.

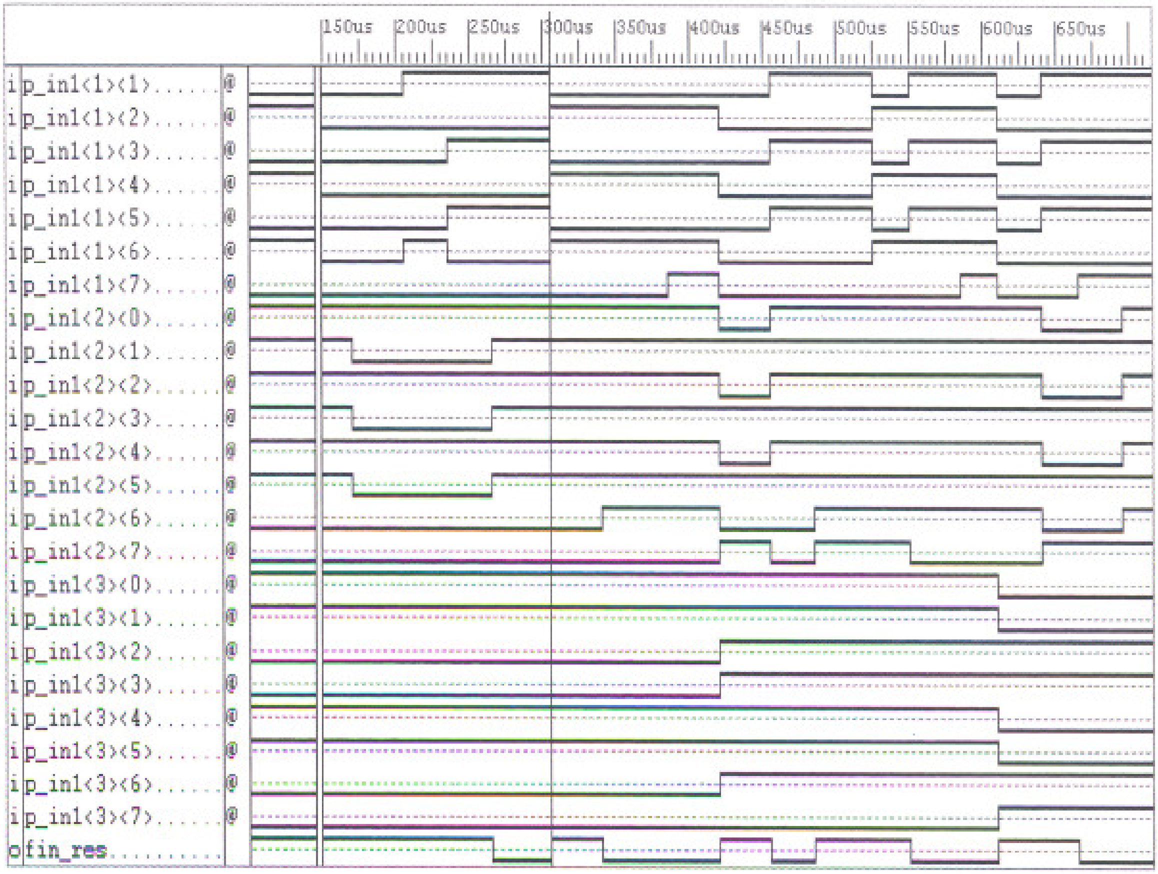


Figure 5.15 Simulation results of network in XILINX

The simulation results are further plotted in MATLAB. The plots so obtained verify the result with respect to the surface view in fuzzy logic.

It is observed that for the delay of 500 ms, there should not be HO if the threshold is good for any value of hystersis margin and even for the cases when the hystersis is small for all values of the threshold. Further the HO should be initiated if the threshold is very poor or poor and sufficient hystersis margin is available.

Similarly, the HO decision can be taken for the different values of the delay. The plots for the 800ms and 200ms delay are shown in the Figure 5.16 and 5.17.

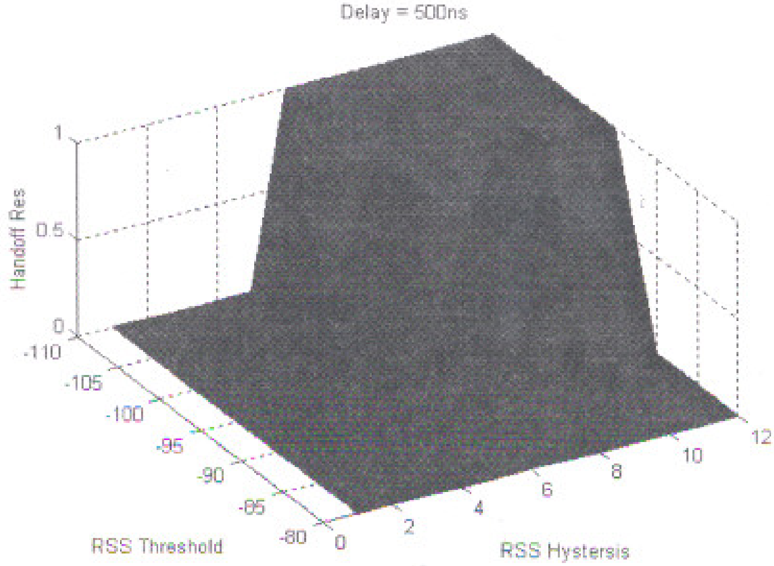


Figure 5.16 Surface plot of network parameters for 500 ns Delay

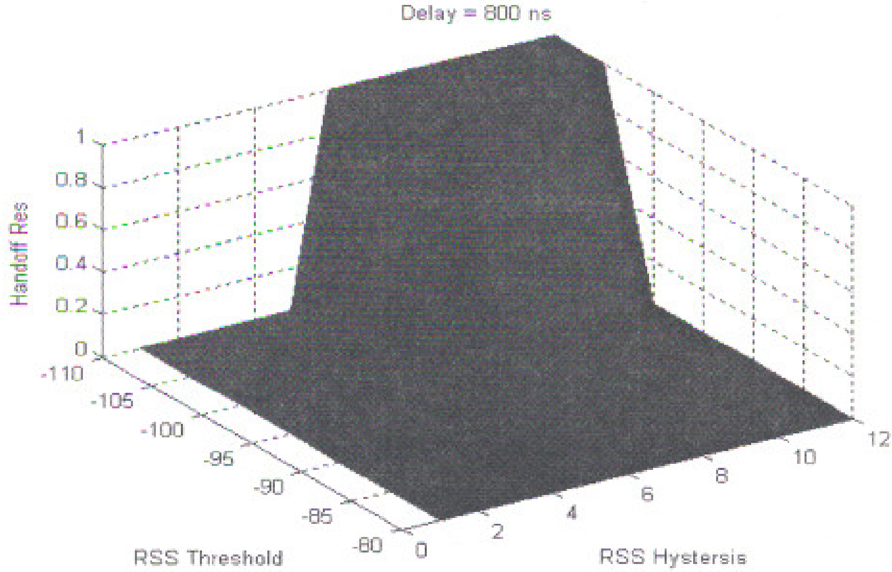


Figure 5.17 Surface plot of network parameters for 800 ns Delay

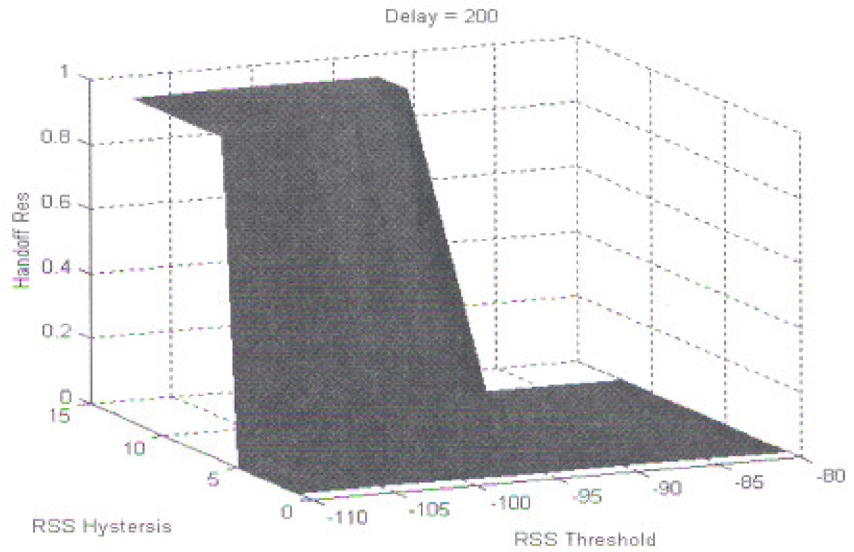


Figure 5.19 Surface plot of network parameters for 200 ns Delay

Synthesis Report

Design Statistics
IOs : 25

Macro Statistics :
Adders/Subtractors : 120
30-bit adder : 6
31-bit adder : 8
32-bit adder : 21
33-bit adder : 5
34-bit adder : 4
35-bit adder : 4
4-bit adder : 18
5-bit adder : 18
6-bit adder : 18
7-bit adder : 18
Multipliers : 46
13x18-bit multiplier : 4
21x13-bit multiplier : 24
21x8-bit multiplier : 18
Comparators : 563
31-bit comparator greater : 1
31-bit comparator less : 1
32-bit comparator greater : 270
32-bit comparator less : 11
32-bit comparator lessequal : 280

Device utilization summary:

Selected Device : v1000bg560-6

Number of Slices:	12001	out of	12288	97%
Number of 4 input LUTs:	21719	out of	24576	88%
Number of bonded IOBs:	25	out of	408	6%

Floorplanner View

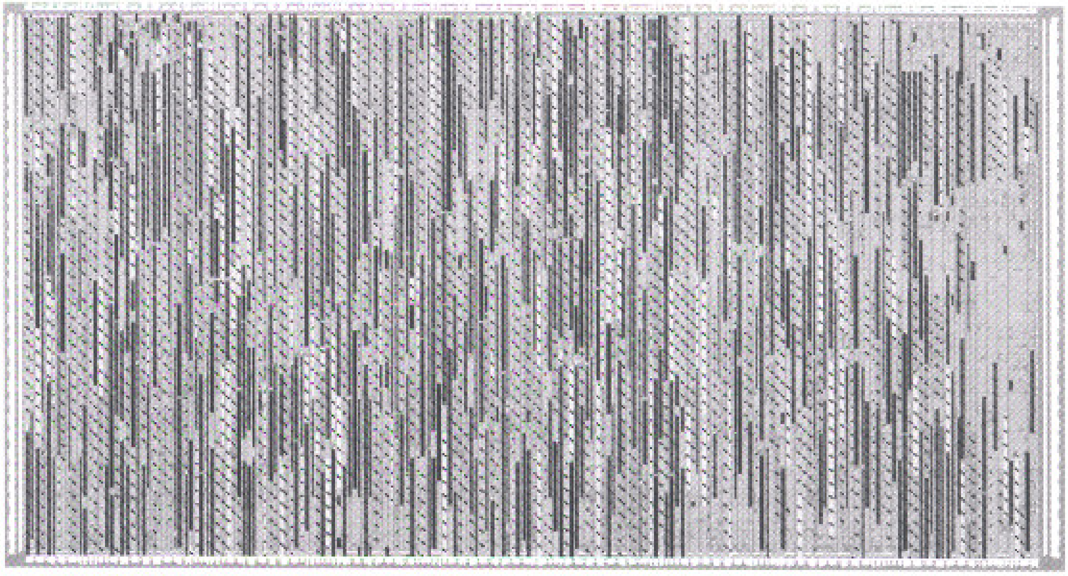


Figure 5.20 Floorplanner View

The figure 5.20 shows that the design fully utilizes the device and the figure 5.21 shows the detailed view of the floorplanner.

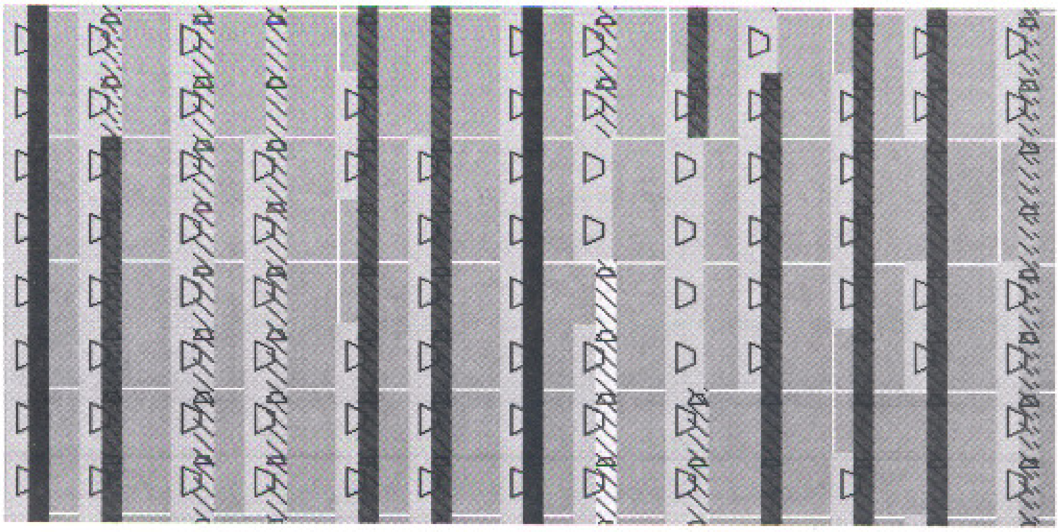


Figure 5.21 Detailed Floorplanner View

The designed chip can be viewed in figure 5.22 with the basic inputs and output.

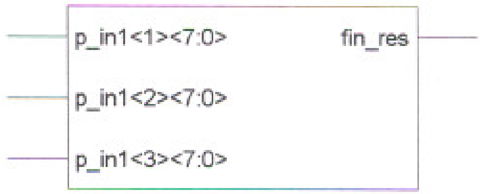


Figure 5.22 Main View of chip

The main chip consists of the following major blocks indicating the original structure of neural network shown in figure 5.23-5.25.

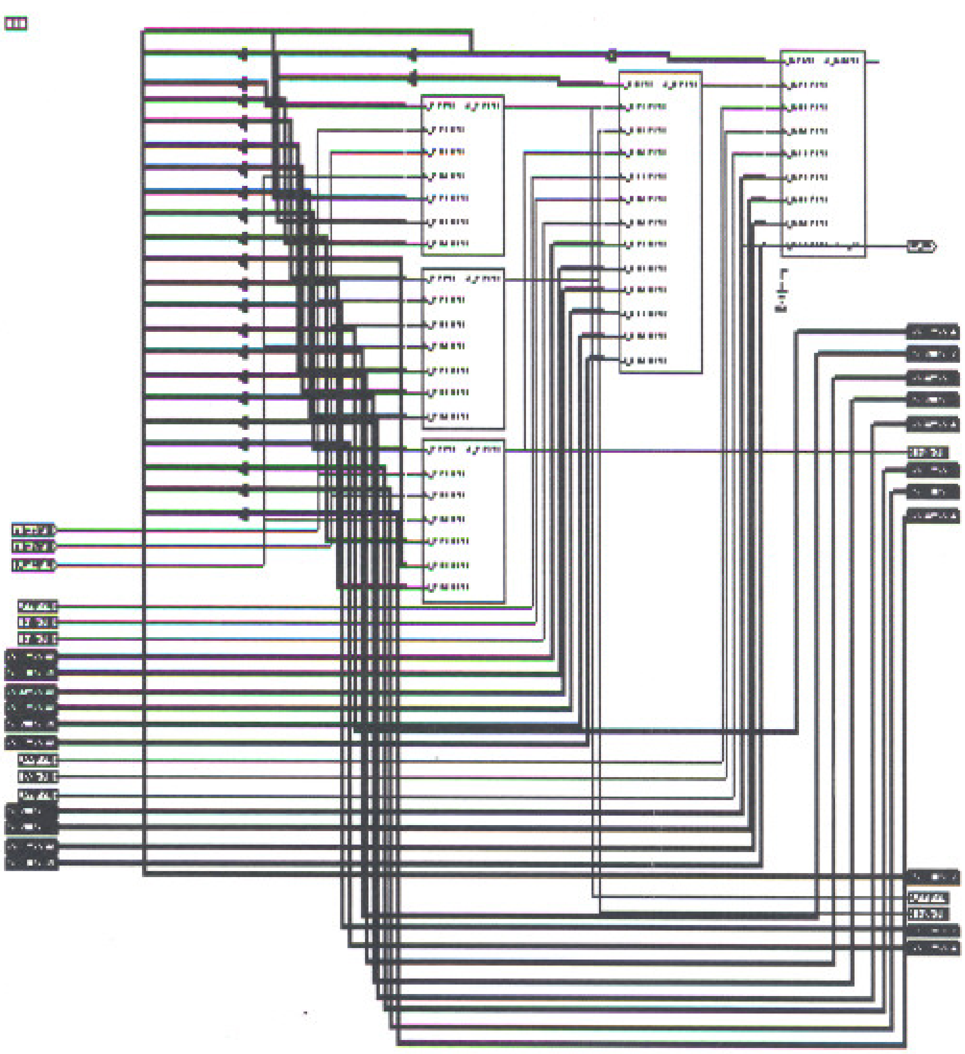


Figure 5.23 Detail_1 View of chip

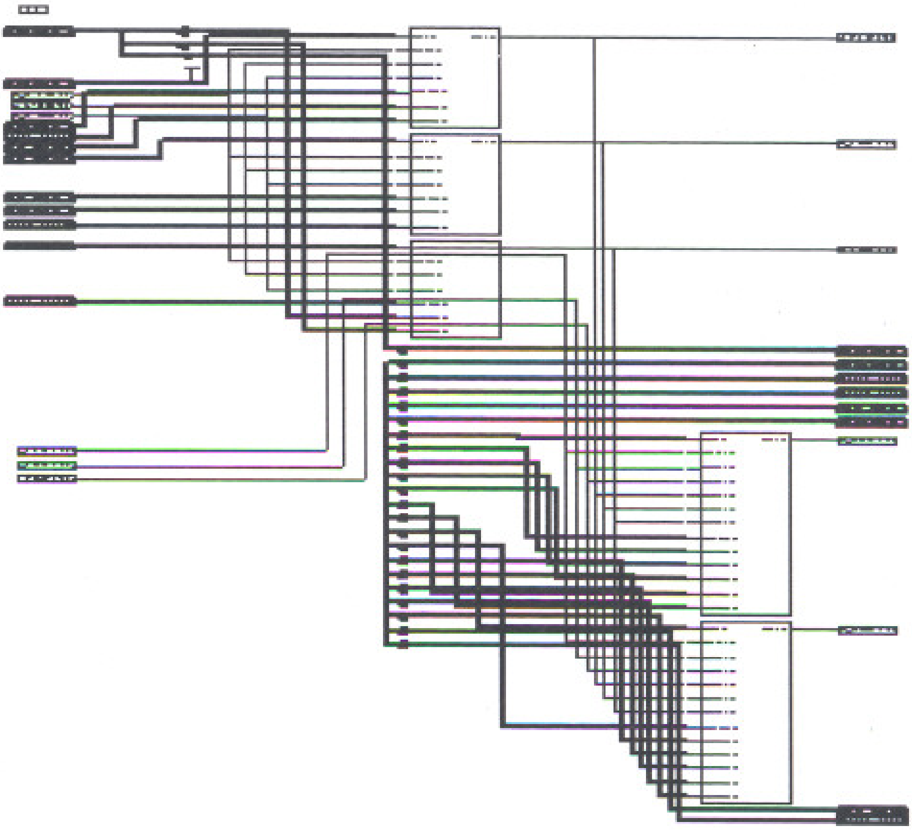


Figure 5.24 Detail_2 View of chip

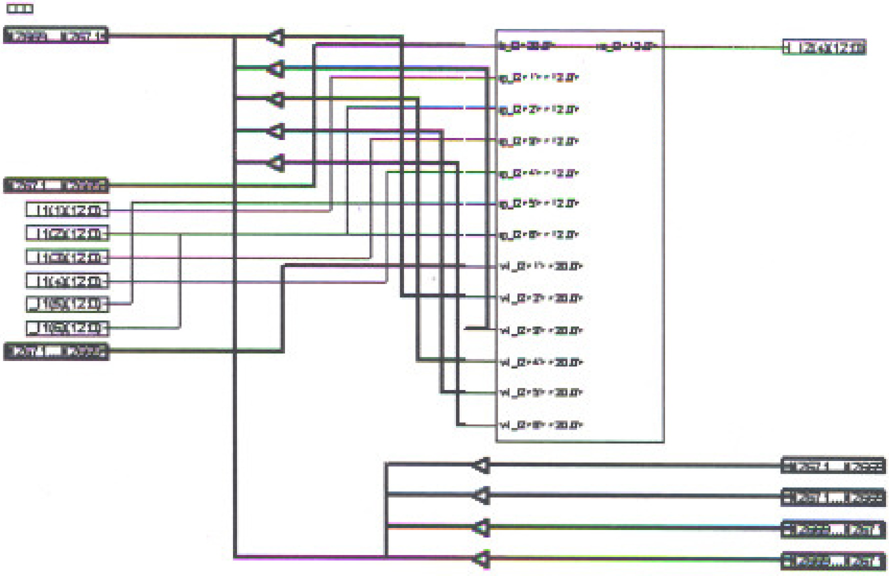
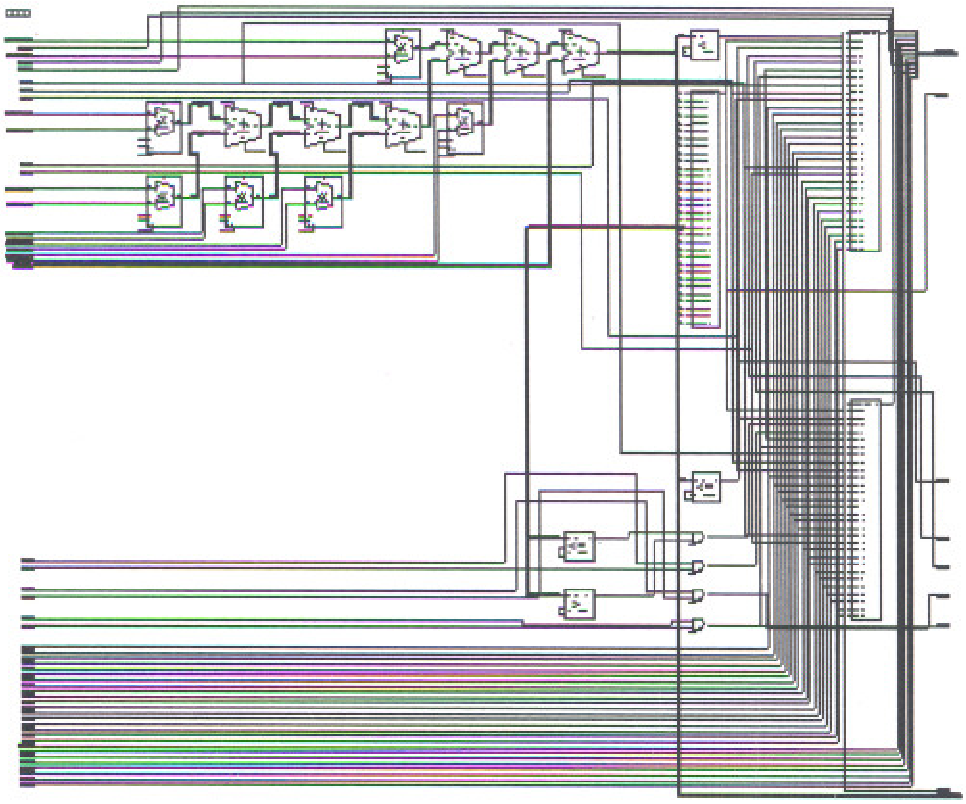


Figure 5.25 Detail_3 View of chip

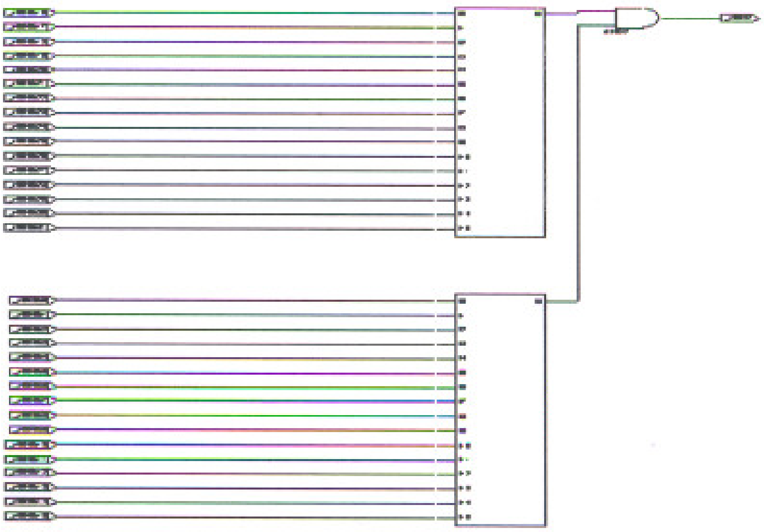
Detail Circuit for a Neuron of Logsigmoid Type

The neuron has basically four main circuits, which are further composed of a number of circuits, till the component level of the neuron is achieved.

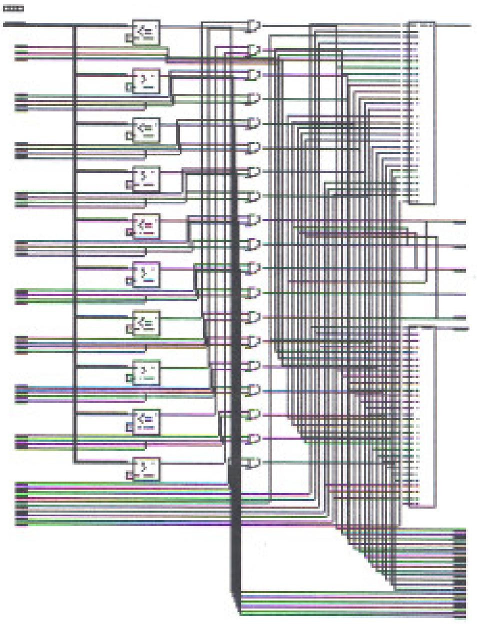
1.



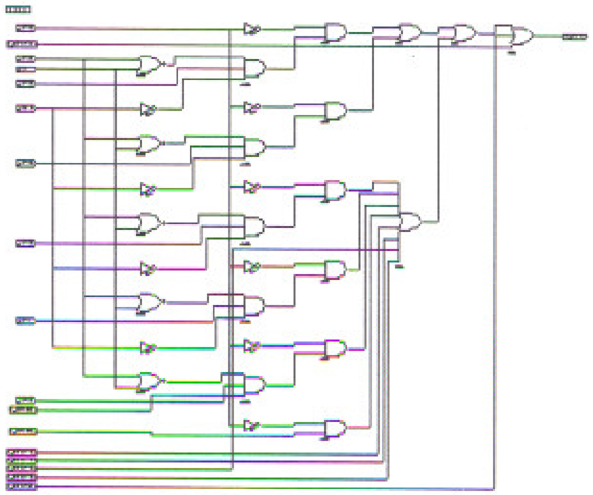
1_a.



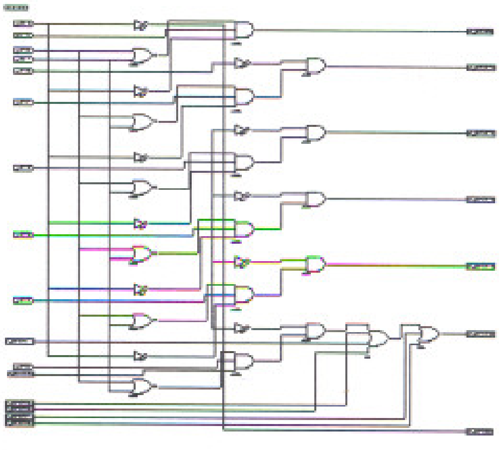
2.



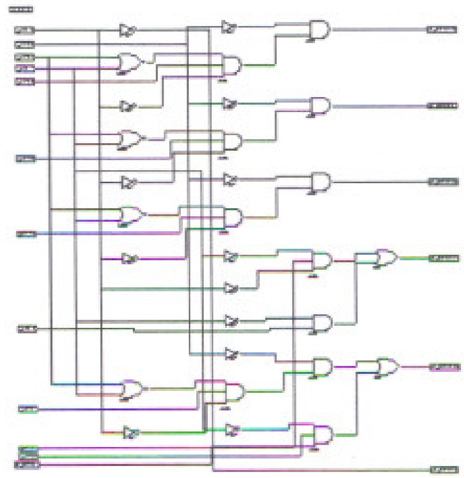
2.a



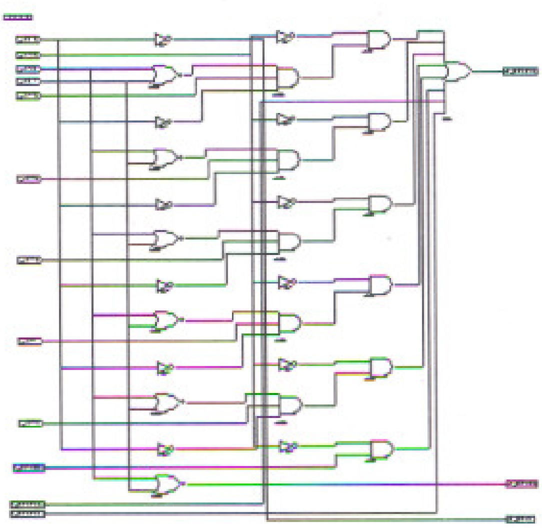
2.b



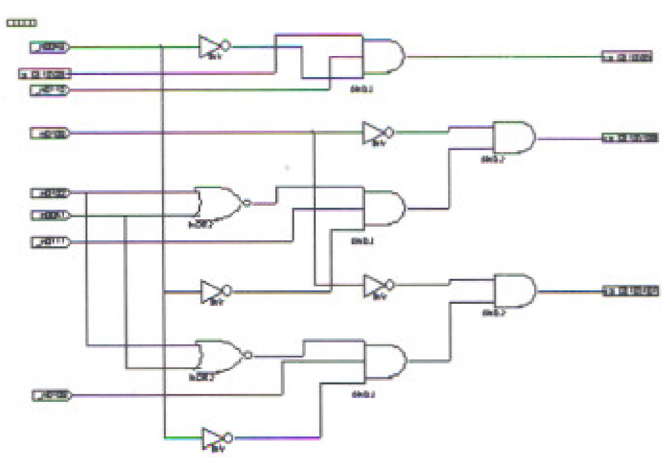
2.c



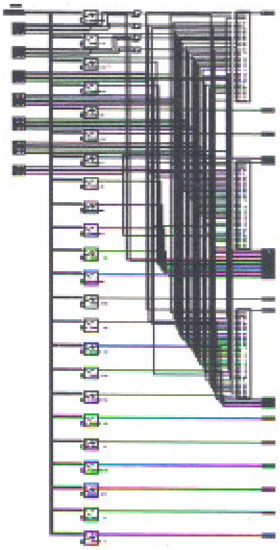
2.d



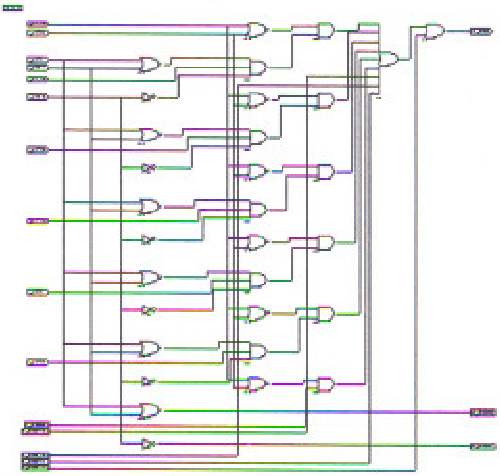
2.e



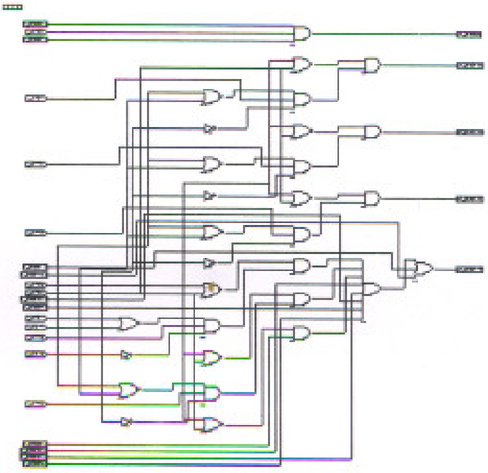
3.



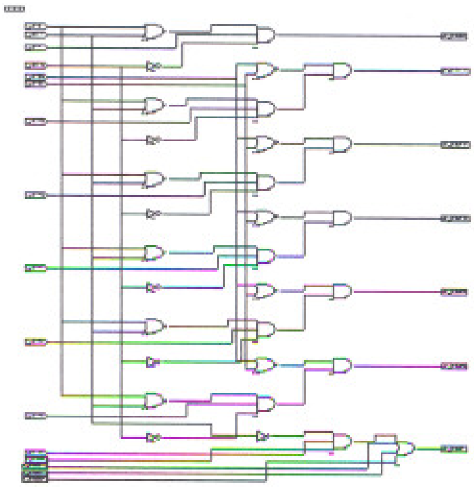
3.1.a



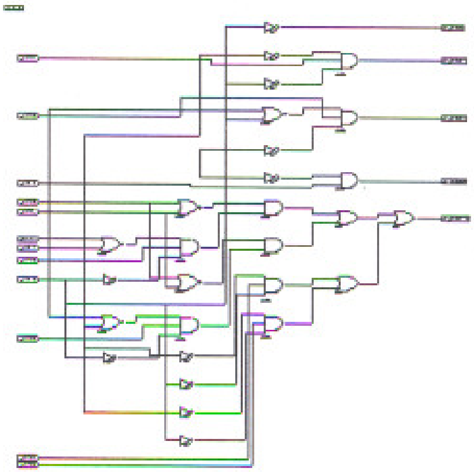
3.1.b



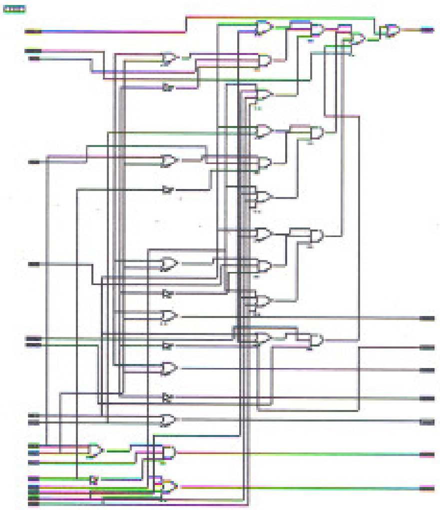
3.1.c



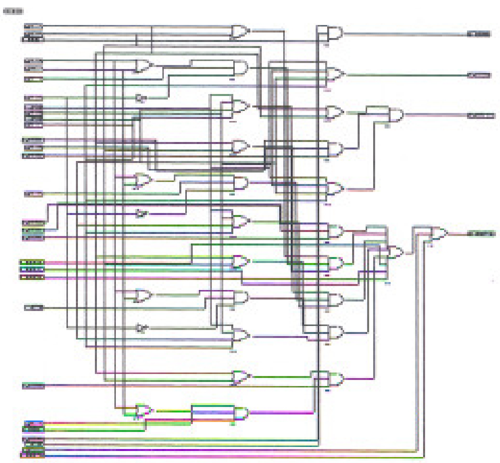
3.1.d



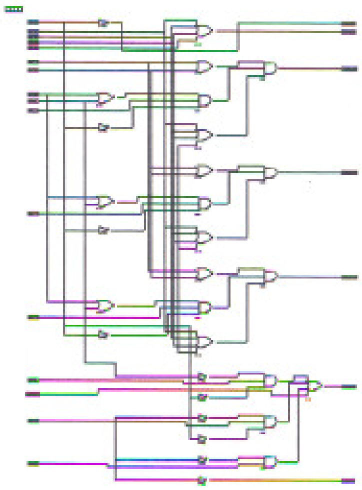
3.2.a



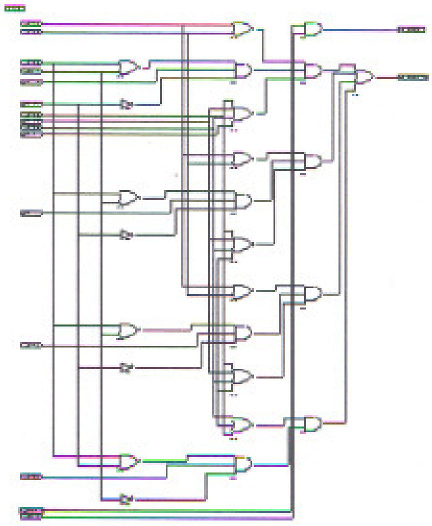
3.2.b



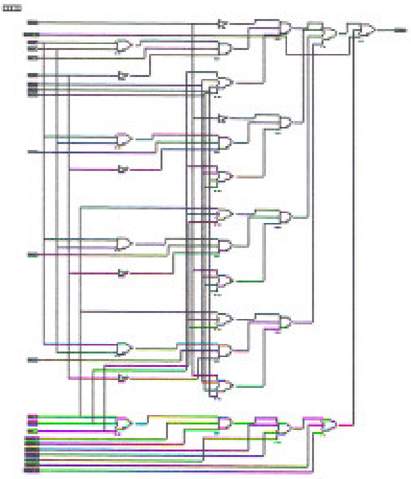
3.2.c



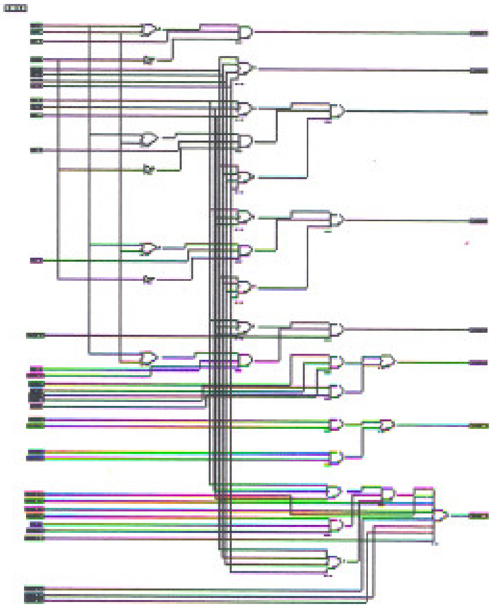
3.2.d



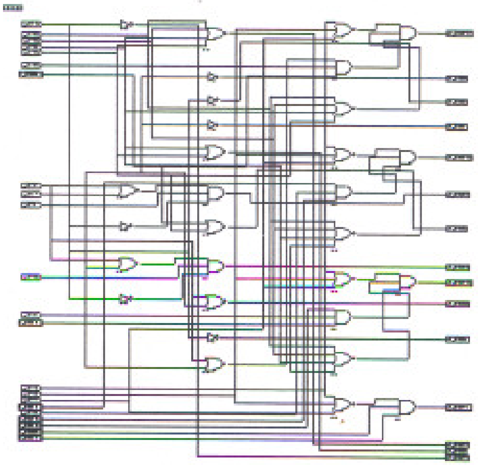
3.3.a



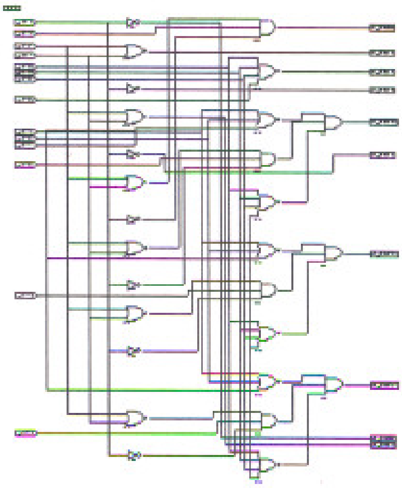
3.3.b



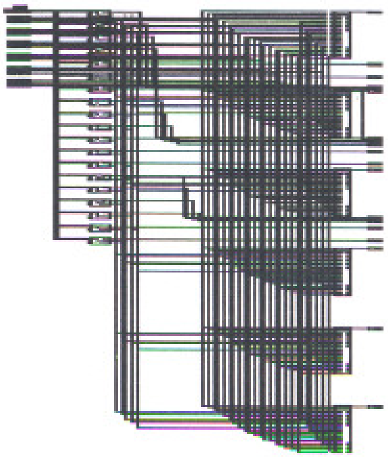
3.3.c



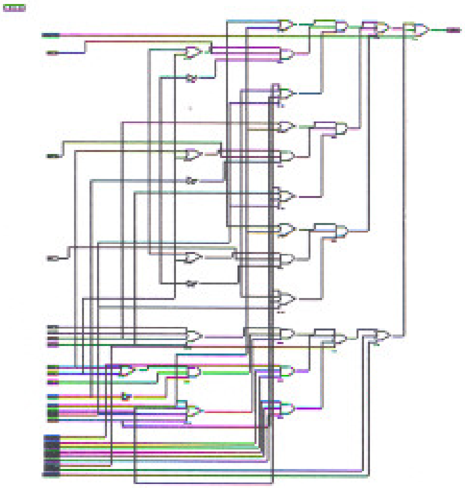
3.3.d



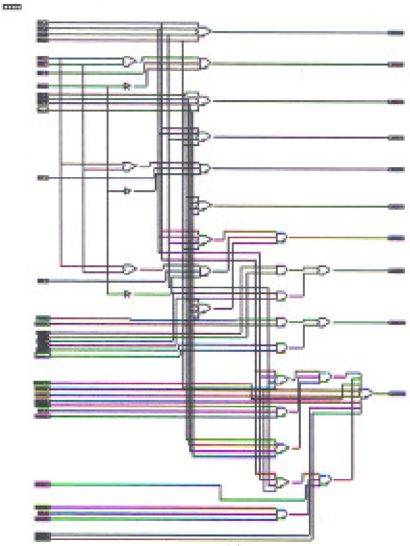
4.



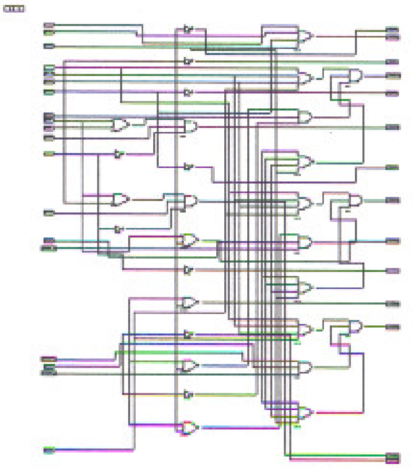
4.1.a



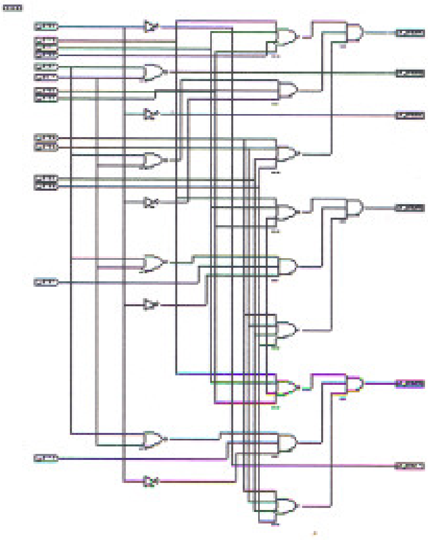
4.1.b



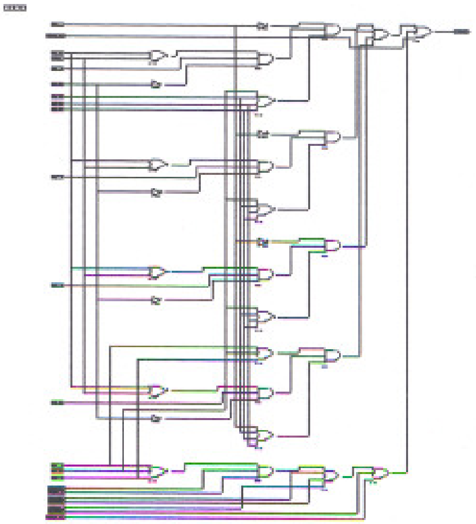
4.1.c



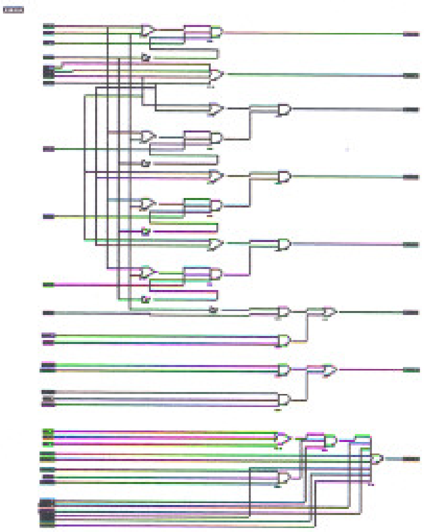
4.1.d



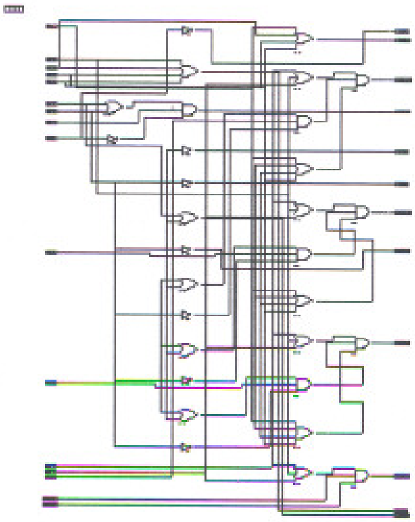
4.2.a



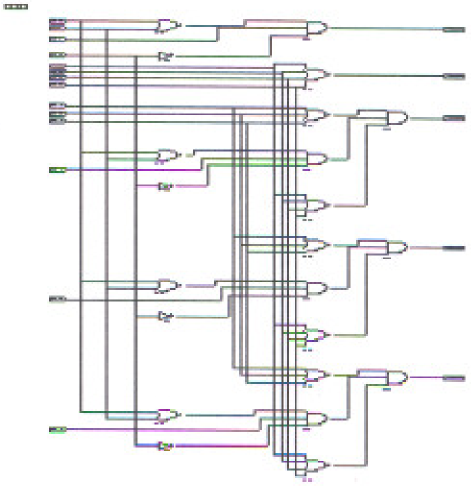
4.2.b



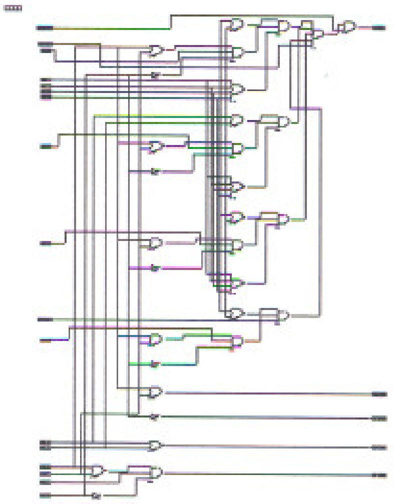
4.2.c



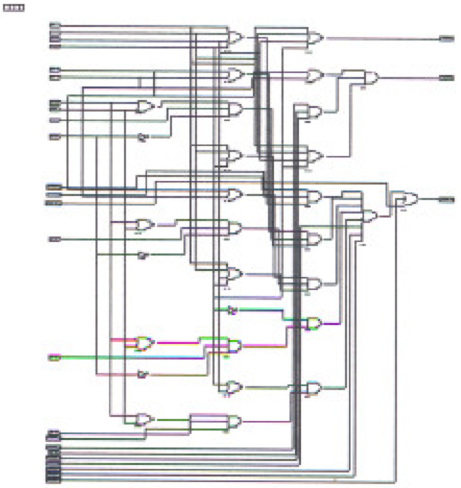
4.2.d



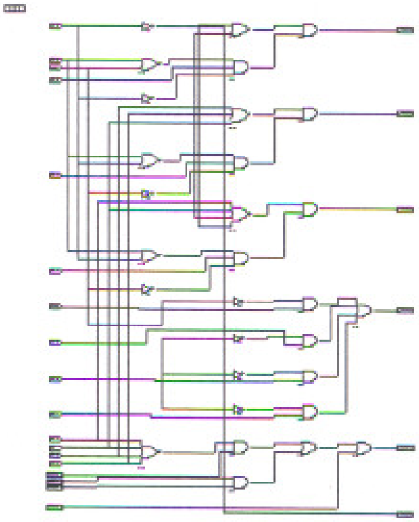
4.3.a



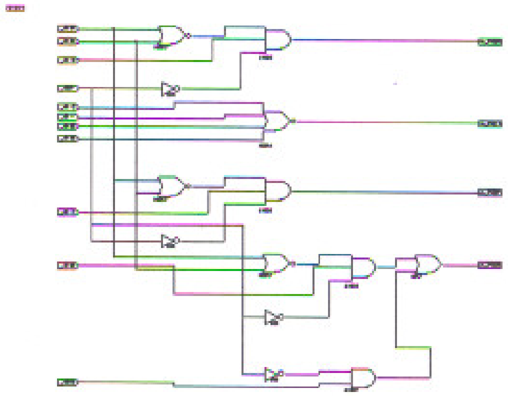
4.3.b



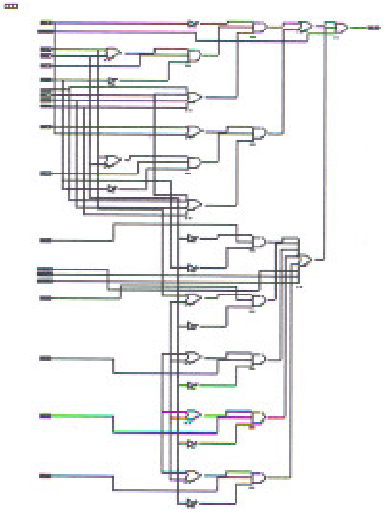
4.3.c



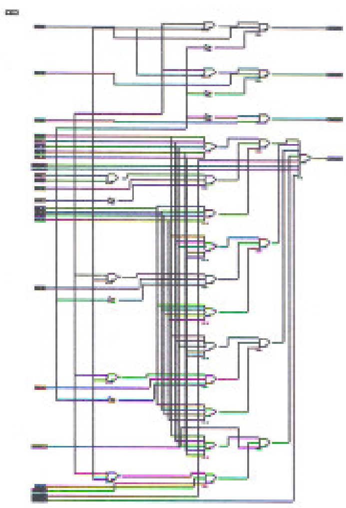
4.3.d



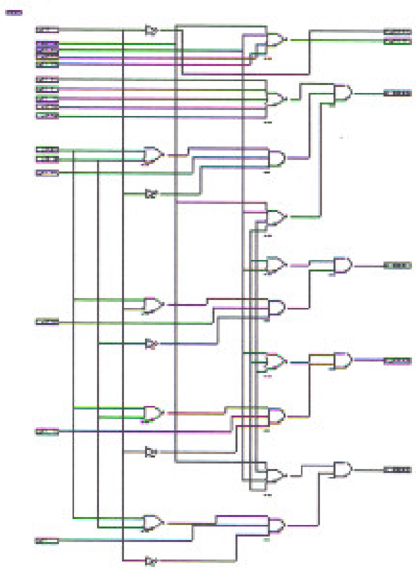
4.4.a



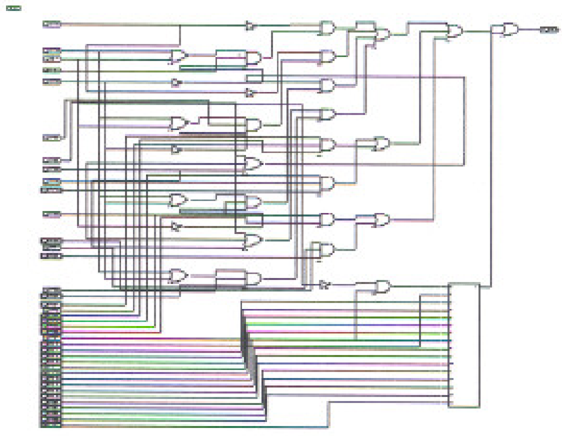
4.4.b



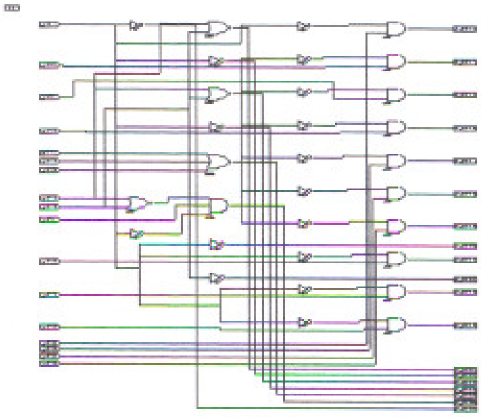
4.4.c



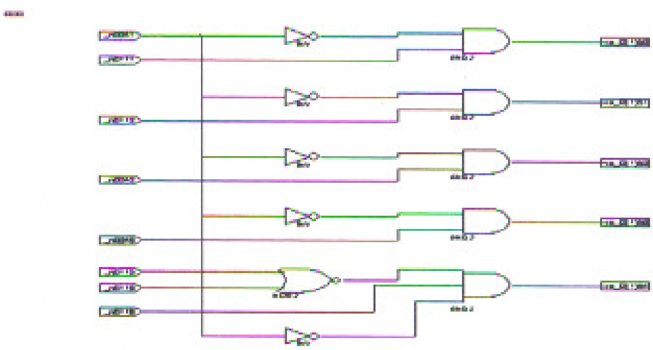
4.5.a



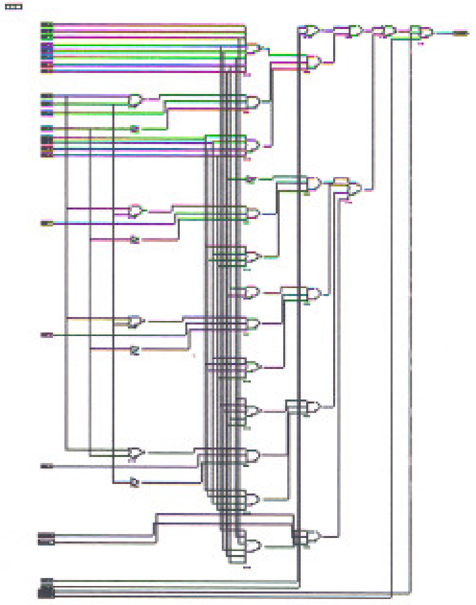
4.5.b



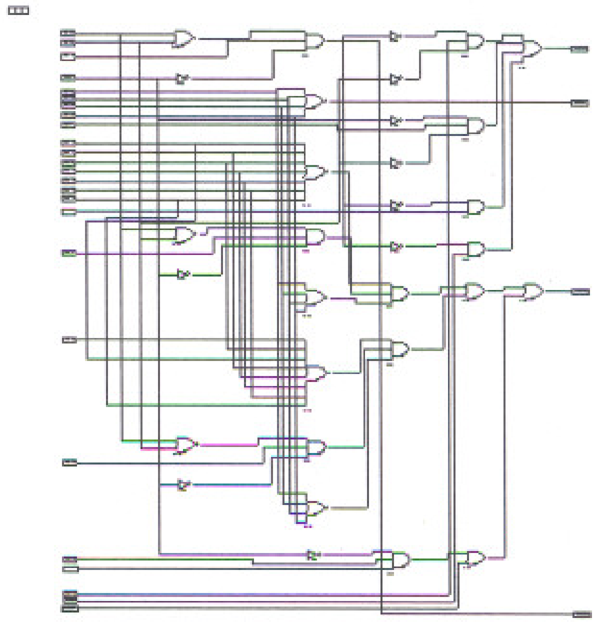
4.5.c



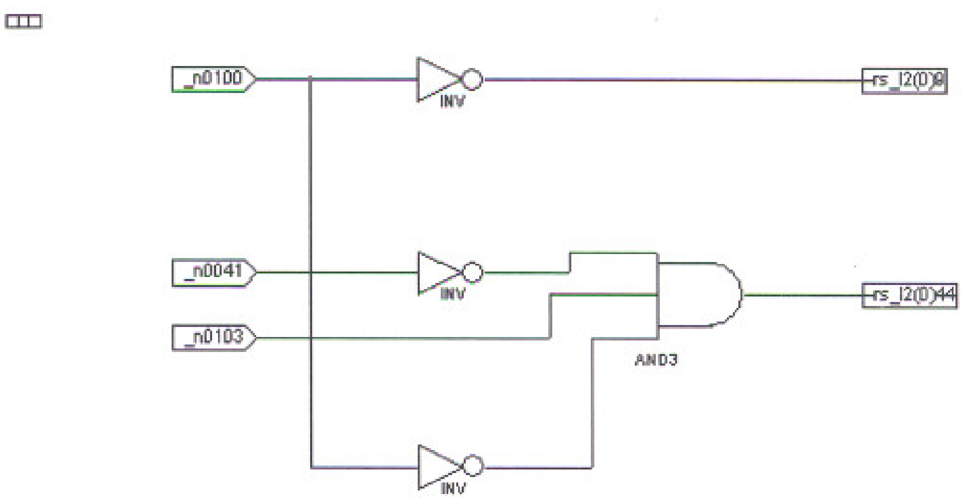
4.6.a



4.6.b



4.6.c



Chapter 6

Implementation of Handoff in CDMA

In this chapter, the basic concept of Code Division Multiple Access (CDMA) network is discussed briefly. The advantages of the CDMA over the TDMA/FDMA are also given. Further, the Soft Handoff (SHO) concept with the conceptual design in fuzzy logic and neural networks is discussed. The hardware implementation of the design is also given.

6.1 Technological Review

CDMA is a digital wireless air interface and networking standard based on the principle of spread-spectrum techniques, which allow multiple users to access the system simultaneously on the same carrier frequency. The intent of CDMA technology is to provide increased bandwidth in a limited frequency system, but has also other advantages including extended range and more secure communications. In a CDMA system, a narrow-band message signal is multiplied by a spreading signal, which is a pseudo-noise code sequence that has a rate much greater than the data rate of the message. CDMA uses these code sequences as a means of distinguishing between individual conversations. All users in the CDMA system use the same carrier frequency and may transmit simultaneously [1,2].

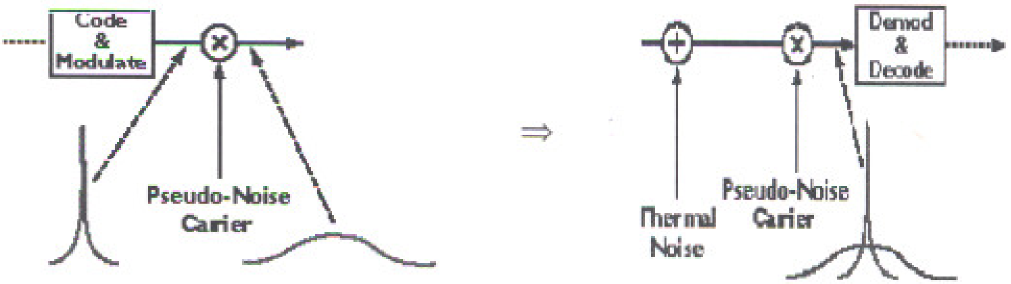
The use of CDMA for civilian mobile radio applications is novel. It was proposed theoretically in the late 1940's, but the practical application in the civilian marketplace did not take place until 40 years later. Commercial applications became possible because of two evolutionary developments. One was the availability of very low cost, high-density digital integrated circuits, which reduce the size, weight, and cost of the subscriber stations to an acceptably low level. The other was the realization that optimal multiple access communication requires that all user stations regulate their transmitter powers to the lowest that will achieve adequate signal quality.

CDMA changes the nature of the subscriber station from a predominately analog device to a predominately digital device. Old-fashioned radio receivers separate stations or channels by filtering in the frequency domain. CDMA receivers do not eliminate analog processing entirely, but they separate communication channels by means of a pseudo-random modulation that is applied and removed in the digital domain, not on the basis of frequency. Multiple users occupy the same frequency band. This universal frequency reuse is not fortuitous. On the contrary, it is crucial to the very high spectral efficiency that is the hallmark of CDMA.

Spread Spectrum

A transmission technique in which a pseudo-noise code, independent of the information data, is employed as a modulation waveform to spread the signal energy over bandwidth much greater than the signal information bandwidth is known as Spread Spectrum. At the receiver the signal is despread using synchronized replica of the pseudo-noise code.

Basic principle : Direct Sequence Spread Spectrum (DSSS)



spreading **despreading**

Figure 6.1 Spreading and despreading

Direct sequence is, in essence, multiplication of a more conventional communication waveform by a **pseudonoise (PN) ±1** binary sequence in the transmitter.

In this, the spreading of the signal takes place prior to any modulation, entirely in the binary domain, and the transmitted signals are carefully band limited. A second multiplication by a replica of the same ±1 sequence in the receiver recovers the original signal.

The noise and interference, being uncorrelated with the PN sequence, become noise-like and increase in bandwidth when they reach the detector. Narrowband filtering that rejects most of the interference power can enhance the signal-to-noise ratio. Thus the SNR of the system is enhanced by the so-called processing gain W/R, where W is the spread bandwidth and R is the data rate.

CDMA Receivers

A CDMA receiver separates the signals by means of a correlator that uses the particular binary sequence to despread the signal and collect the energy of the desired signal. Other users' signals, whose spreading codes do not match this sequence, are not despread in bandwidth and, as a result, contribute only to the noise. These signals represent a self-interference generated by the system [1]. The output of the correlator is sent to a narrow-bandwidth filter. The filter allows all of the desired signal's energy to pass through, but reduces the interfering signal's energy by the ratio of the bandwidth before the correlator to the bandwidth after the correlator. This reduction greatly improves the signal-to-interference ratio of the desired signal. This ratio is also known as the processing gain. The signal-to-noise ratio is determined by the ratio of the desired signal power to the sum of all of the other signal powers. It is enhanced by the processing gain or the ratio of spread bandwidth to baseband data rate.

CDMA Channel Assignments

A CDMA digital cellular waveform design uses a pseudorandom noise (PN) sequence to spread the spectrum. The sample rate of the spreading sequence (called the chip rate) is chosen so that the bandwidth of the filtered signal is several times the bandwidth of the original signal [2].

A typical system might use multiple PN sequences. In addition, it might use repeated spreading codes of known lengths to ensure orthogonality between signals intended for different users. The channel assignment is essentially determined by the set of codes that are used for that particular link. Thus, the signal transmitted at any time in a logical channel is determined by:

- The frequency of operation for the base station
- The current symbol
- The specific orthogonal spreading code assigned for the logical channel
- The PN spreading code

CDMA Signal Processing

In the demodulation of CDMA signals, the different paths may be independently received, which greatly reduces the severity of the multipath fading. However, multipath fading is not completely eliminated because occasionally there may be multiple paths that cannot be independently processed by the demodulator [24].

Different users in CDMA employ signals that have very small cross-correlation. Thus, correlators can extract individual signals from a mixture of signals even though they are transmitted simultaneously in the same frequency band. CDMA systems employ wideband signals with good cross-correlation properties, which means the output of a filter matched to one user's signal is small when it receives a different user's signal as input.

In direct-sequence spread-spectrum systems, a high-rate antipodal pseudorandom spreading sequence modulates the transmitted signal so that the bandwidth of the resulting signal is roughly equal to the rate of the spreading sequence. The cross-correlation of the signals is then largely determined by the cross-correlation properties of the spreading signals. Although CDMA signals overlap in both time and frequency domains, they can be separated, based on their spreading waveforms.

Spreading rates can be chosen to exceed the coherence bandwidth so that the channel becomes frequency selective. For instance, different spectral components are affected unequally by the channel, and only parts of the signal are affected by fades. Expressing the same observation in time domain terms, multipath components are resolvable at a resolution equal to the chip period and can be combined coherently, for example, by means of a rake receiver. Coherent combination of multipath components requires an estimate of the channel impulse response. Such an estimate can be calculated from a training sequence or by means of a pilot signal.

Fundamentals of the IS-95A CDMA System

This section describes the important features of the IS-95A CDMA system specifications. The CDMA Reference Blockset is designed to develop models to simulate different components of the IS-95A CDMA system.

Channel Schematics

The figure 6.2 illustrates an IS-95A forward channel. The transmitter section includes channel coding, modulation and spreading, and filtering. The receiver section includes filtering, despreading and demodulation, and channel decoding [24].

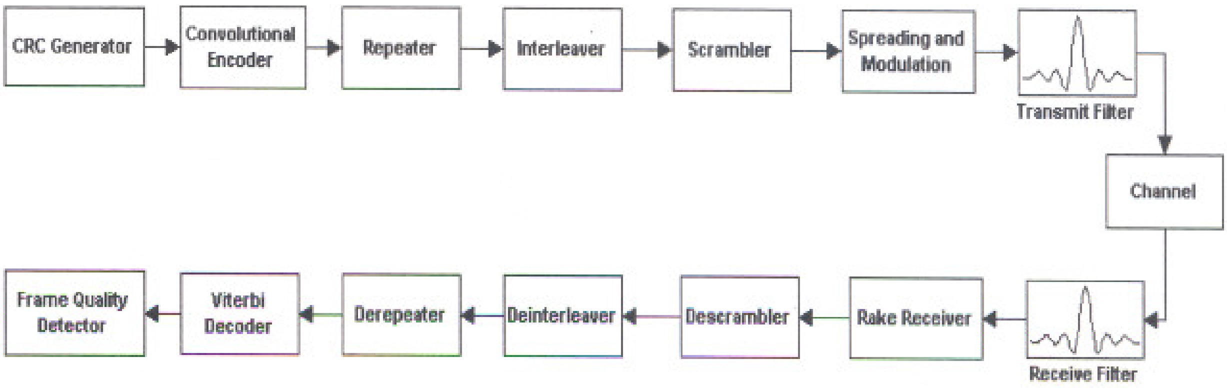


Figure 6.2. IS-95A Forward Channel Diagram

The Figure 6.3 illustrates an IS-95A reverse channel. It includes many of the same operations that are in the forward channel, but the functionalities of the blocks correspond to the reverse channel specifications.

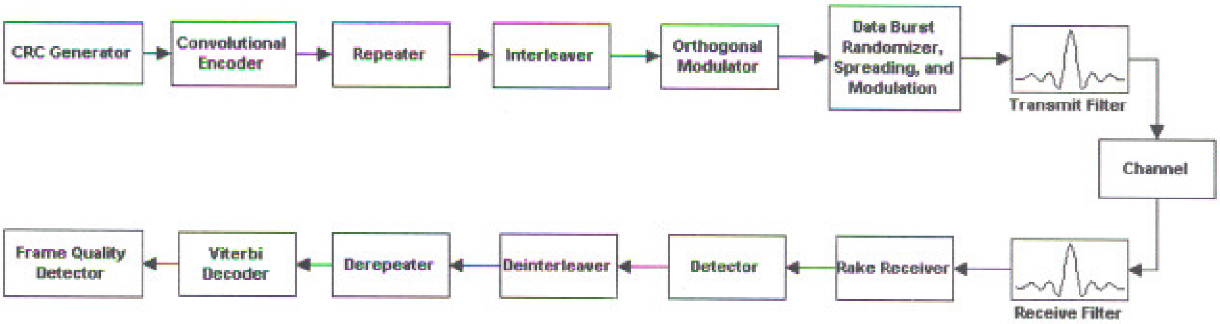


Figure 6.3 .IS-95A Reverse Channel Diagram

The details of the Forward and Reverse channel diagrams are however, not included here. Only the basic concept is discussed. Further, we discuss the advantages of CDMA over TDMA or FDMA in the next section and the concept of soft handoff in CDMA systems is discussed thereafter. Much emphasis is laid on the handoff process and the algorithm used is also briefly discussed.

6.2 Advantages of CDMA over TDMA/FDMA

CDMA technology enhances wireless communications and provides a number of benefits, which taken together represent an advance over FDMA wireless technology. The primary advantage of both CDMA and TDMA technology over FDMA technology is capacity [1,2,24].

Increased capacity: The major advantage of CDMA is increased capacity through more efficient use of the spectrum. Greater capacity enables your CDMA wireless network to handle higher call density at a lower cost. The number of users that can use the same CDMA carrier and still have acceptable performance is determined by the total interference power that all of the users generate in the receiver. There is no hard limit on the number of system users in a CDMA system. Spread spectrum takes advantage of the fact that at any given time, there will be enough open holes in the spectrum for enough information to get through. The odds of a conflict depend only on the likelihood of two or more users landing on the same frequency at the same time. The more users, the more collisions. Signal quality is measured as a bit error rate (BER).

Improved voice quality and system performance: CDMA technology increases the reliability and quality of service by reducing static and improving voice clarity. The soft handoff feature ensures that calls are connected before handoff is completed, minimizing dropped calls and speech disruption. Multipath signals are combined for increased signal integrity.

Increased mobile unit battery life: CDMA requires less RF power than the other access technologies. It allows the mobile to operate at a power level many times lower than analog and TDMA. This conserves the battery power of the mobile unit, increasing talk time, and as a result, revenues.

Multipathing: Whereas FDMA and TDMA suffer losses and interference due to naturally occurring multipath RF signals, CDMA signal quality actually improves under such conditions. This characteristic greatly improves in-building RF penetration, for example. CDMA receivers (called rake receivers) use three or four parallel correlators to receive and track separately the strongest of signals in multiple paths. The receiver then combines the signals constructively (in-phase) and uses the result to demodulate the signal. While there is fading on each arrival, the fades are usually independent of one another. A loss in performance occurs only when all correlators experience fades at the same time.

Soft handoff: A soft handoff permits a call to be carried by two to six cells or sectors at the same time while the mobile station is traveling through a handoff zone. The difference in arrival time of signals from the cells or sectors is treated just as multipathing—the mobile receiver combines the signals constructively.

A handoff in FDMA or TDMA, referred to as a hard handoff, is performed on a "break before make" basis: the old link is dropped before the new link is established. In contrast, a soft handoff in CDMA is performed on a "make before break" basis: the mobile chooses the best quality signal from multiple links (up to six).

A soft handoff virtually eliminates the interference, clipping, and clicks commonly associated with a hard handoff. Since every cell uses the same CDMA carrier, the only difference in transmission is the binary codes (sequences). Under normal circumstances there is no handoff from one frequency to another frequency. The probability that a CDMA call will be dropped because a handoff command has been received in error is greatly reduced.

Security: CDMA uses spread-spectrum technology in which the information content is spread over a wider bandwidth than the frequency content of the original information. Each subscriber is assigned a unique binary code that distinguishes that user from all other users simultaneously transmitting over the same frequency band. Would-be eavesdroppers hear only unintelligible noise and can only decode the signal with the appropriate equipment and binary codes.

Eliminates need to engineer reuse pattern: CDMA's N=1 frequency reuse pattern provides superior Radio Frequency (RF) coverage for a given transmit power level and simplifies frequency engineering when configuring the network.

In FDMA and TDMA, frequency management is both a critical and difficult task to carry out. Since the frequency reuse factor is 1 for CDMA, no frequency management is needed for CDMA.

Unlike current FDMA and TDMA access technologies, which require frequency engineering to avoid co-channel (same channel) interference in nearby cells, the same block of CDMA spectrum may be reused in every cell or sector. CDMA, by its very design, can decode the proper signal in the presence of high interference.

Enables seamless integration of data applications: Variable rate signal coding permits higher-rate voice coding and bandwidth-on-demand for data transmissions. CDMA's packetized communications structure is well suited to data transmissions and services. Packetized data transmission is also available to TDMA cells.

Suitability for microcell and in-building systems: CDMA is a natural waveform suitable for microcell and in-building wireless systems because of its tolerance to noise and interference.

6.3 Soft Handoff in CDMA

Introduction

In soft handoff, the link to the prior base station is maintained as the user is transferred to the new cell's base station. That is to say that the mobile is linked to more than one base station at a given time. Soft handoff is primarily used in CDMA [26]. In CDMA systems, all mobile stations transmit signals using the same frequency band. This makes it possible to use soft handoff. During the soft handoff, each mobile station can simultaneously connect to multiple base stations [1,2].

A 2-way, 3-way or 6 way soft handoff is one in which a mobile unit is communicating with two, three or 6 cells at the same time [11]. In Lucent Technologies' CDMA system implementations, each mobile station can connect up to 6 base stations. This is known as 6-way soft handoff [8]

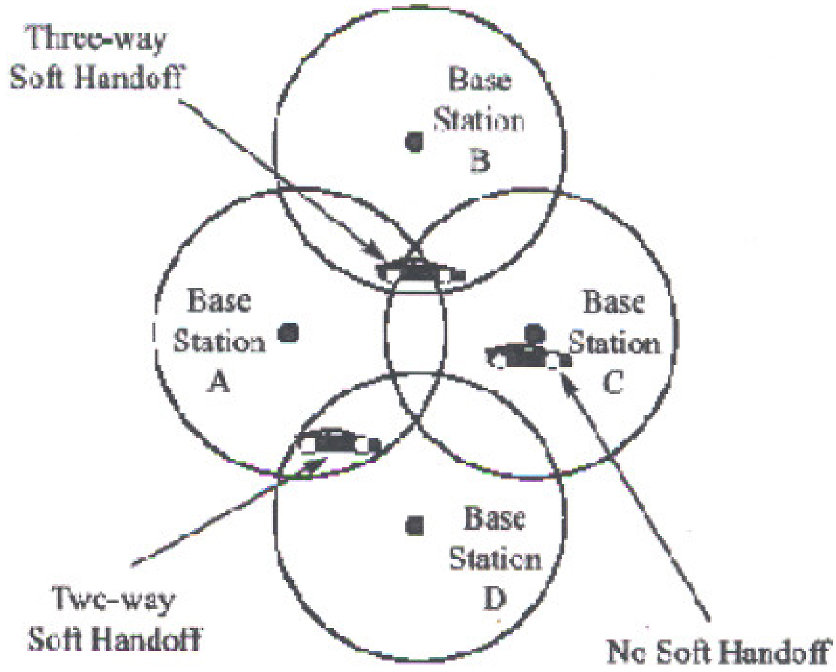


Figure 6.4: Soft Handoff Scenarios in a Cellular System

Figure 6.4 shows generic soft handoff scenarios in a cellular system. When a MS is close to a BS (e.g., near BS C), it communicates with only that BS (BS C).However, near cell borders, it is relatively far from all adjacent BSs, and the RSS from a single BS may not be sufficient to provide a good quality communication link. In such a case, the MS combines RSSs from different BSs to obtain a good quality signal using one of the diversity combining techniques such as equal gain combining, selection combining, and maximal ratio combining [1]. For example, in the overlap region between Cell A and Cell D, the MS is connected to both BS A and BS D, leading to a two-way soft handoff scenario. There may also be a three-way soft handoff in which the MS communicates with three BSs (e.g., BSs A, B, and C in Figure).

The simultaneous connections provide a diversity gain that improves link quality in fringe areas. The application of power control from neighbor base stations also ensures that a progressively distant mobile station will not unduly boost it's transmit strength and become a primary source of interference to a nearby base station.

A softer handoff is an intracellular handoff, occurring between sectors of a two sector or three sector "sub-cell." This type of handoff occurs only at the cell site and is independent of the MSC. When a call is in softer handoff, the second connection to the requested call circuit is made before the original connection is broken. The mobile will communicate with only one CDMA Channel Element (CE), which handles and combines the signals from both sectors. Softer handoffs will improve trunking efficiency and provide improved switch performance [8].

Figure shows Soft handoff between the mobile unit and base station.

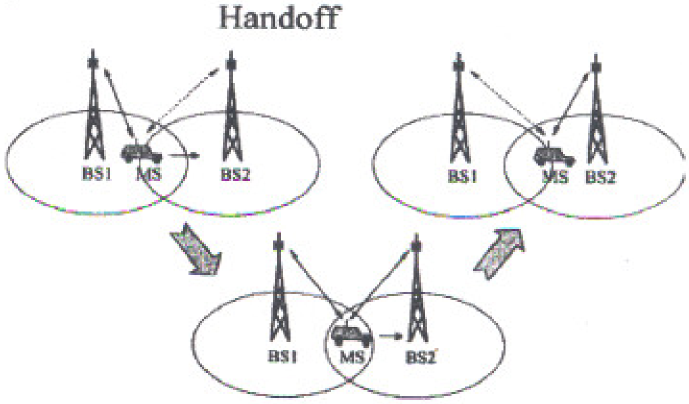


Fig. 6.5 Soft Handoff between Mobile unit and Base station

Soft handoff is beneficial because it reduces interference into other cells and improves performance by using macro diversity [2]. Also, it has been shown that soft handoff can extend coverage area by as much as a factor of 2.5. In hard handoff, a handoff is performed when the signal strength of an adjacent cell exceeds that of the current cell by some threshold. In CDMA, the adjacent cell frequencies are just the same as those of the current cell. Therefore, using this hard handoff technique would cause severe interference into neighbor cells and thus degrade capacity. In a CDMA system with soft handoff, each mobile user is connected to two or more base stations at a time. The base station with the highest relative strength seen from the mobile is given the control of the mobile user's call. Also, because a user in soft handoff is connected to several adjacent base stations, probability of a lost call is reduced. Soft handoff fits nicely into the structure of CDMA [22]. As was just mentioned, in the uplink several or more base stations may receive the user signal. This is because of CDMA's reuse factor of one. In the downlink, the signals from the base stations can be coherently combined as they are seen as multipath components.

CDMA Terminology

I. **Active set.** It is the set of base stations that are involved with the mobile station during the soft handoff.

II. **Active set update.** This is when a change in the active set occurs. An update occurs when a candidate base station exceeds the add threshold, when an old base station has been below the drop threshold for too long, or the active list becomes too large.

III. **Discard set.** These are the base stations that are currently members of the active set but will be dropped because they are no longer qualified as such.

IV. **Candidate set.** CS contains the channels which are almost as good as those in AS and any one of them can be chosen for SHO as a new member of AS.

V. **Neighbor set.** NS contains the set of channels, which are not included in the AS and CS, but are reasonably strong.

VI. **Remaining set.** RS contains other left over channels that are not the members of all other sets.

Procedure

Suppose that the mobile station is linked and communicating with base station 1. Every base station is sending a pilot signal, which among other things gives a measure of the signal strength to mobile users. When the signal strength of base station 2 exceeds the add threshold, base station 1 is notified to place base station 2 onto the candidate list. Further, when the signal strength of base station 2 becomes greater than that of base station 1 by some specified level, Base station 2 is placed on the active list and it also is allowed control of the call. Here, diversity combining is implemented. Now upon the signal level of base station 1 going below the drop threshold, the drop timer is activated. If it happens now that the signal level of base station 1 goes back above the drop level, the drop timer will be reset. However, if the signal strength level goes below the drop threshold and the drop timer expires, base station1 is dropped from activity with the call [8].

6.4 The Algorithm

Soft Handoff is a diversity handoff scheme that user attempts to have simultaneous traffic communication channels with more than one base station. Ref [18] proposed the technique based on fuzzy logic to increase the value of T_DROP in order to release the traffic at high traffic load for increasing the carried traffic.

Whenever MS feels the need for soft handoff, it searches for the other usable cells to which it could handoff the control when needed. If the NS pilot strengths become above T_ADD, the MS includes the pilot into the CS and removes it from the NS. The BS controller instructs the MS to add the new pilot and the MS will add the pilot to the AS, if it receives permission to add the pilot. If the current AS pilot strength decreases below T_DROP for T_TDROP seconds, the MS moves the pilot from the AS into the NS if it receives permission to drop the pilot. The AS is limited to have six pilots, while the CS is limited to five.

The inputs of the proposed SHO algorithm are the:

- Number of AS's pilots of any MS (no_{BS}) and
- Number of Remaining channels (CH_{rm}) of serving BS.

The outputs are the values of new T_DROP and T_ADD. However, the value of T_ADD for each MS is assigned as T_DROP + 2 db for constant SHO window size.

6.5 Conceptual Design

As discussed in the chapter 5, the system performance is checked using MATLAB. Using the FIS editor, the two inputs and single output FIS is framed.

6.5.1 Fuzzy Logic Approach

The two inputs considered are:

- Number of AS's pilots of any MS (no $_{BS}$) and
- Number of Remaining channels (CH $_{rm}$) of serving BS.

The outputs are the values of new T_DROP and T_ADD. However, the value of T_ADD for each MS is assigned as T_DROP + 2 db for constant SHO window size.

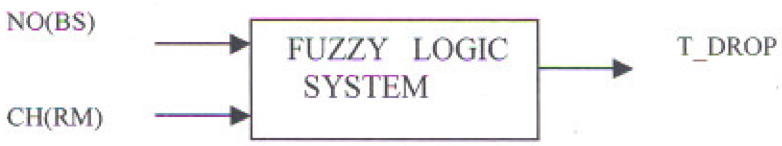


Figure 6.6 Fuzzy Logic Based Algorithm

The membership functions for the different inputs of the FLS are made followed by the fuzzy rule base, which gives the desired response Figure 6,7,6.8,6.9,6.10. The rule and surface views so obtained can be seen in the Figures 6.11, 6.12.

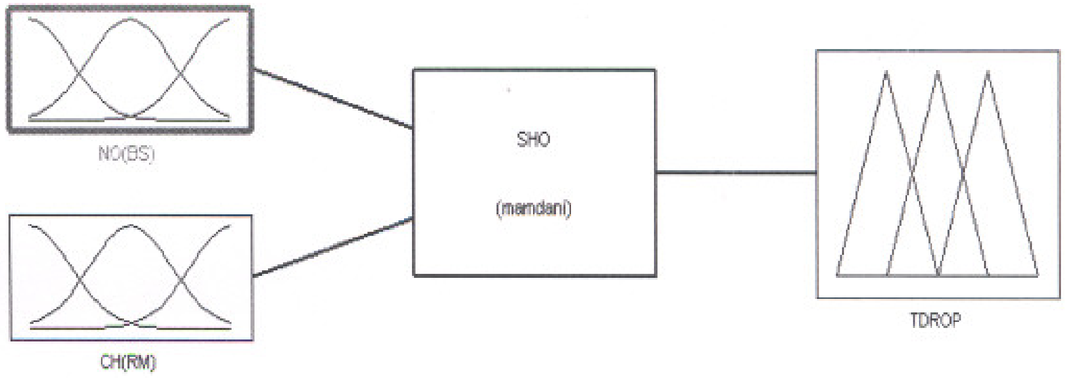


Figure 6.7. Basic FLS for SHO

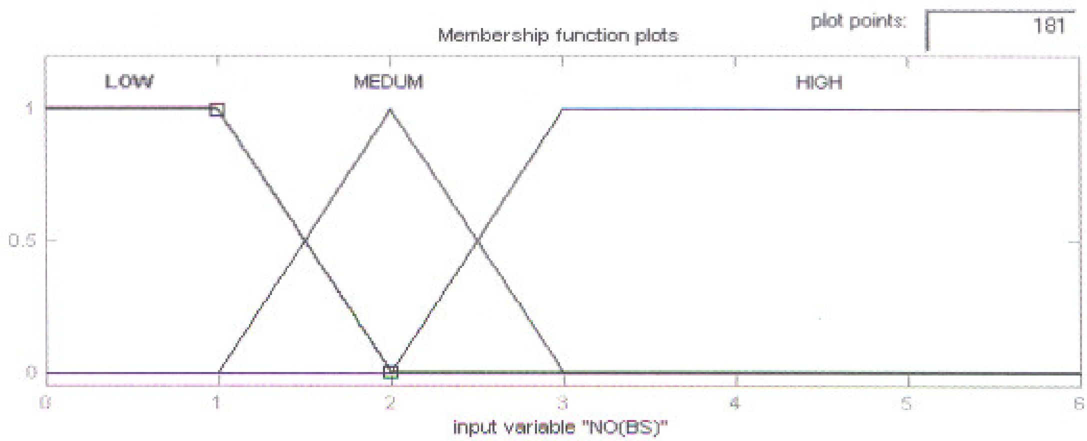


Figure 6.8 Membership function for NO(BS)

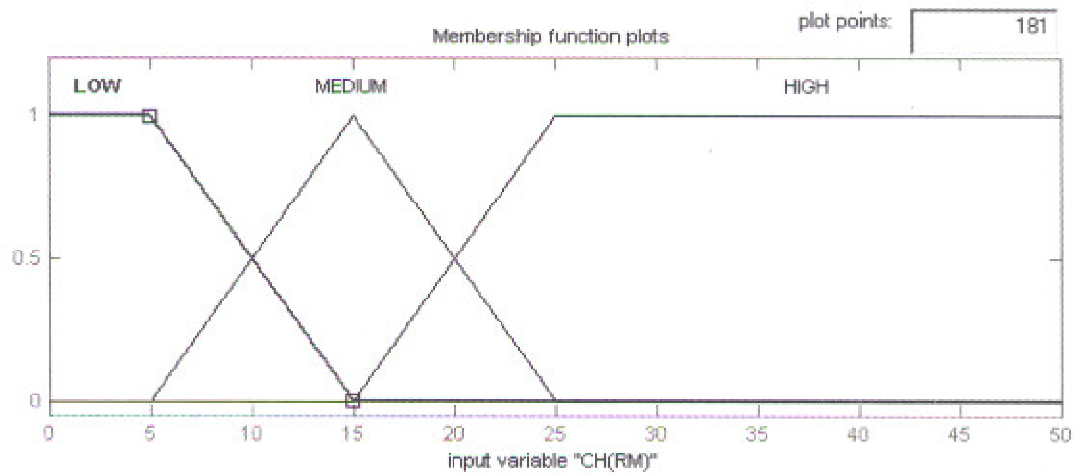


Figure 6.9 Membership function for CH(RM)

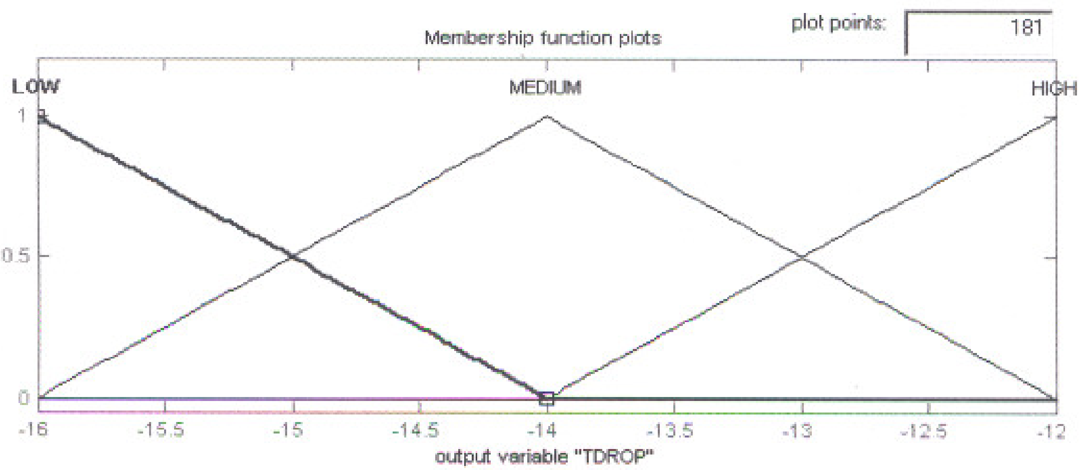


Figure 6.10 Membership function for T_DROP

he rule base for the designed system consists of 25 rules. Table 6.1 shows the system rule base:

Table 6.1 System Rule Base

CH(RM)	NO(BS)	T_DROP
Low	Low	High
Low	Medium	High
Low	High	High
Medium	Low	Medium
Medium	Medium	Medium
Medium	High	High
High	Low	Low
High	Medium	Low
High	High	Medium

After the system structure and Rule base is specified, we simulate the system response. The system performance can be obtained as a single crisp value for a given set of crisp inputs as specified in figure 6.11(rule view). Here the two inputs are {3,25} and the corresponding output is –14.

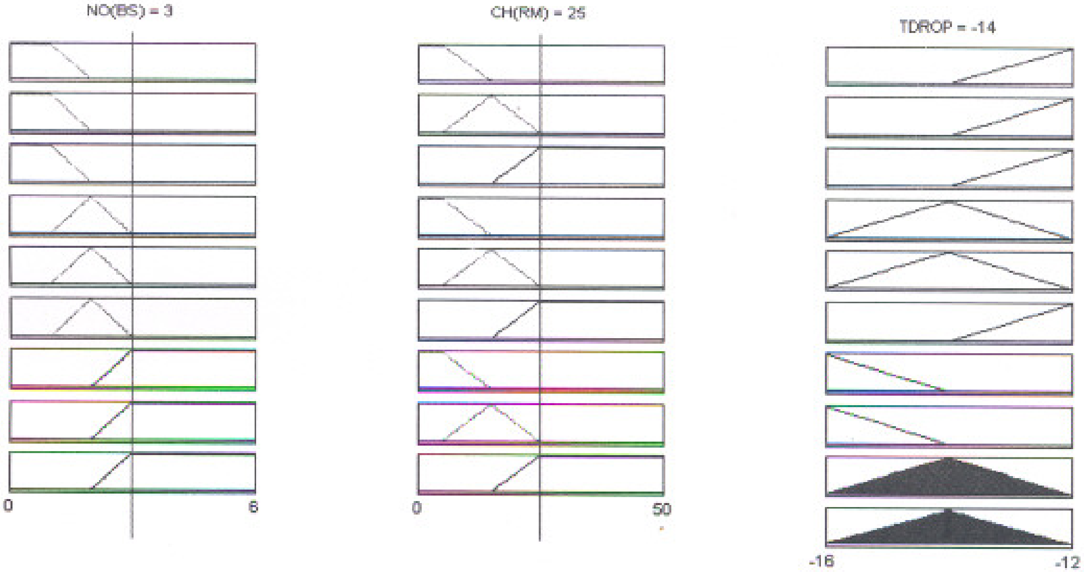


Figure 6.11 Rules for the Soft Handoff Threshold

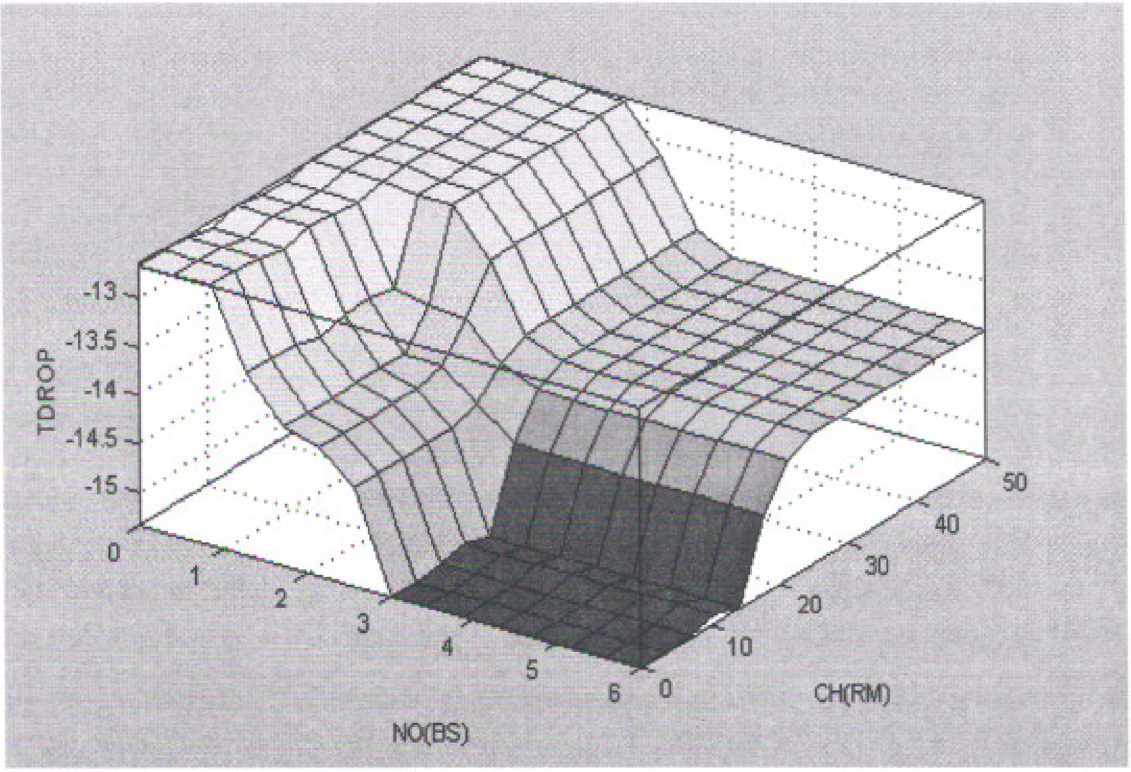


Figure 6.12. Surface view for the FLS

Observations

As discussed in article 5.2.1 all the rules shown in the view rule can be verified with the surface view of the system.

For the low values of CH(RM) and high values of NO(BS), the value of the T_DROP is minimum. Further, for the higher values of CH(RM) and low for NO(BS), the value of T_DROP is found to be the maximum. The rules defined are used to monitor the output in the figure 6.12.

6.5.2 ANN Approach

The above discussed system is further implemented in the neural networks using the NNTOOL of MATLAB. The procedure for this is same as discussed in the section 5.2.2 Thus, we get the following structure of 6-2-1 for this system.

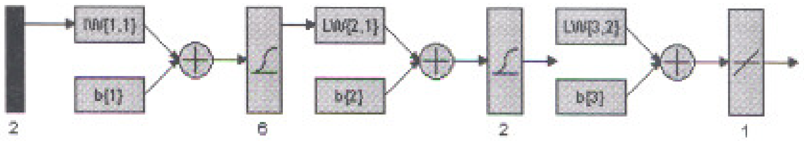


Figure 6.13 Structure of desired network

6.5.3 Hardware Implementation

Finally, the structure discussed above is implemented in VHDL as in section 5.2.3 and the simulation results and the synthesis reports are discussed in the section below.

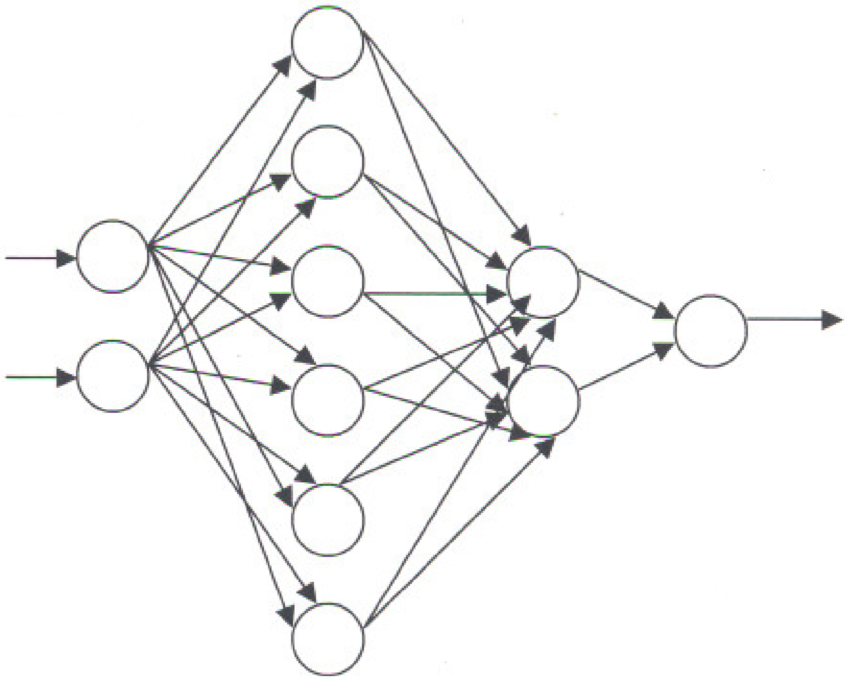


Figure 6.14 Detail structure of design

Simulation Results

The simulation results of the system so developed are shown in figure 6.15. The signals The signals 'P_in1<1>'and 'P_in1<2>' are taken as the inputs for the network and the signal 'fin_res' is the final result, giving the value of the desired threshold level i.e. the value of T_DROP.

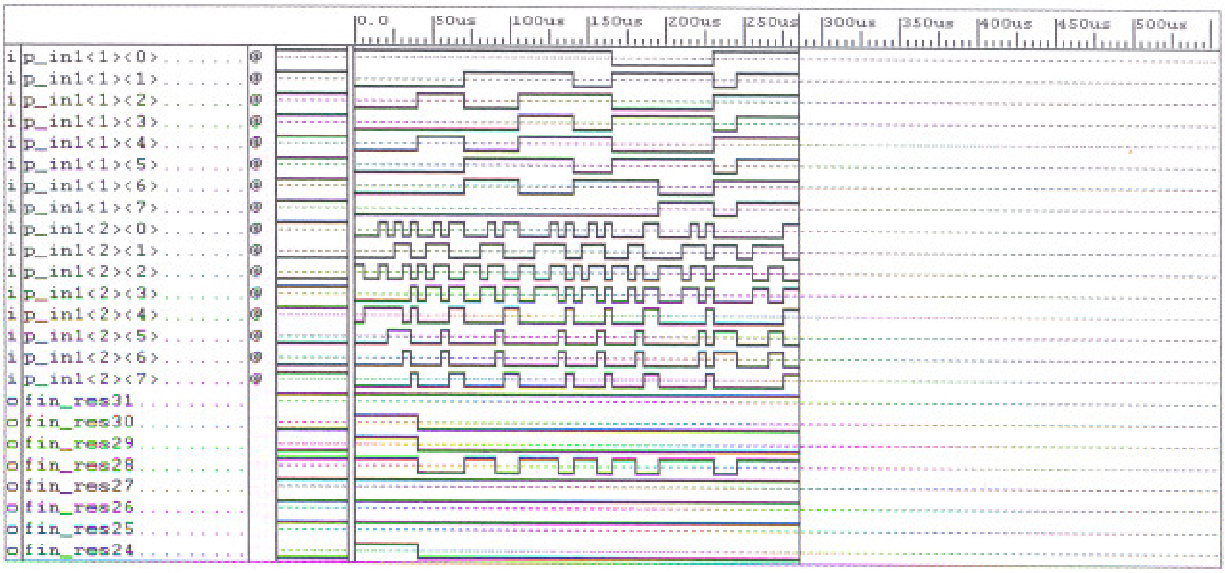


Figure 6.15 Simulation result_1 for network in XILINX

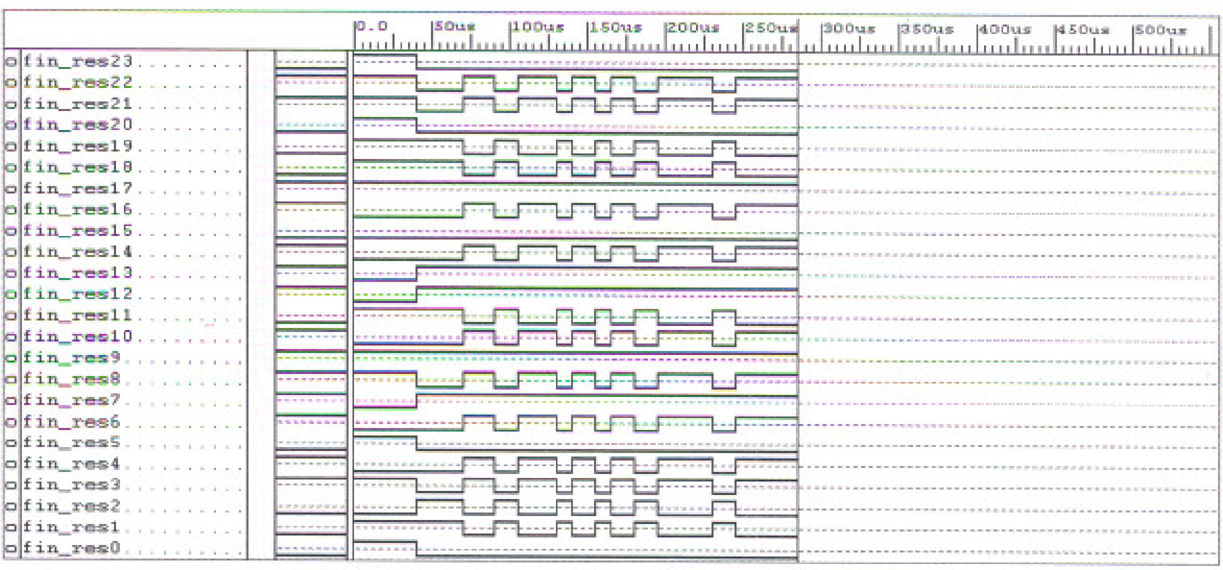


Figure 6.16 Simulation result_2 for network in XILINX

Synthesis Report

Design Statistics
IOs : 48

Macro Statistics :
Adders/Subtractors : 74
11-bit adder : 2
32-bit adder : 18
33-bit adder : 2
34-bit adder : 2

```
#     35-bit adder              : 2
#     4-bit adder               : 12
#     5-bit adder               : 12
#     6-bit adder               : 12
#     7-bit adder               : 12
# Multipliers                   : 26
#     13x21-bit multiplier      : 2
#     21x13-bit multiplier      : 12
#     24x8-bit multiplier       : 12
# Comparators                   : 448
#     32-bit comparator greater   : 216
#     32-bit comparator less      : 8
#     32-bit comparator lessequal : 224
```

Device utilization summary:

Selected Device : v1000bg560-6

```
Number of Slices:           8097  out of  12288   65%
Number of 4 input LUTs:    14737  out of  24576   59%
Number of bonded IOBs:        27  out of    408    6%
```

Floorplanner View

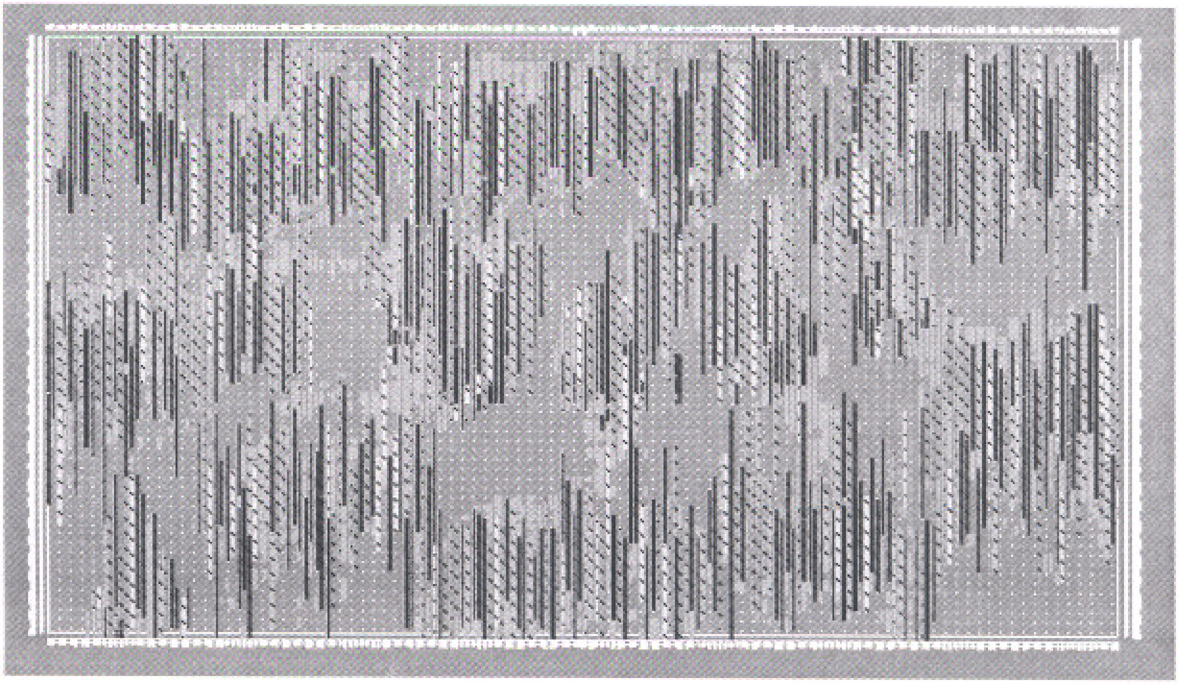


Figure 6.17 Floorplanner View

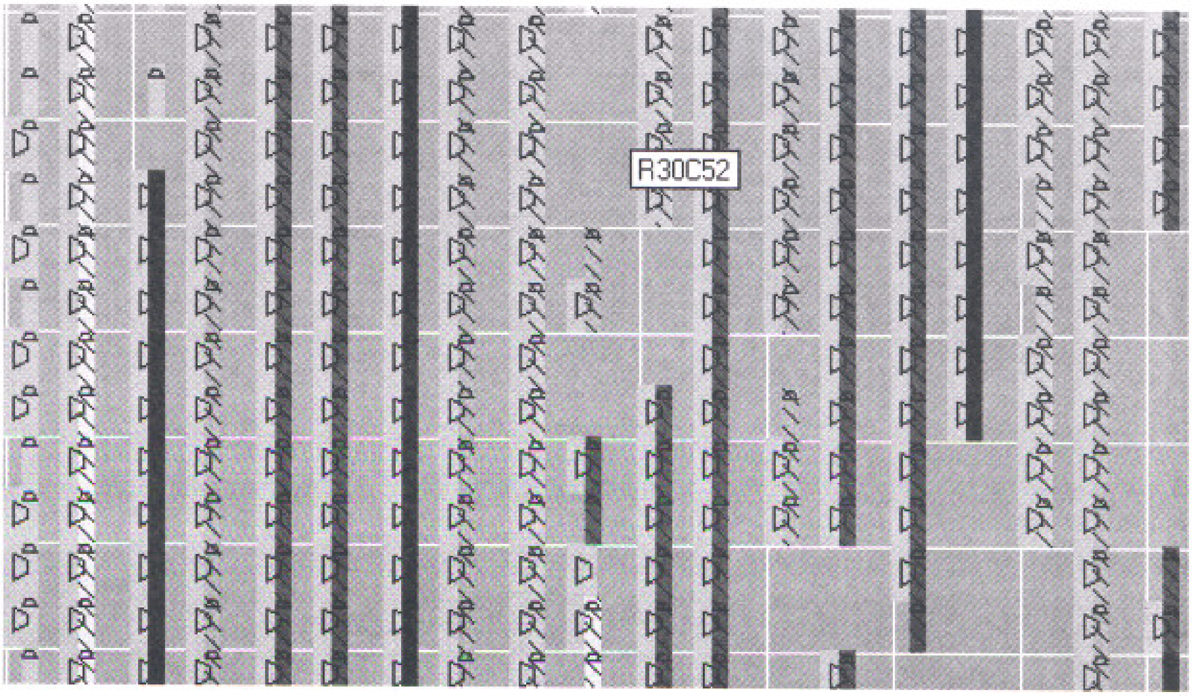


Figure 6.18 Detailed Floorplanner View

The chip view with the major blocks can be seen in the figure below:

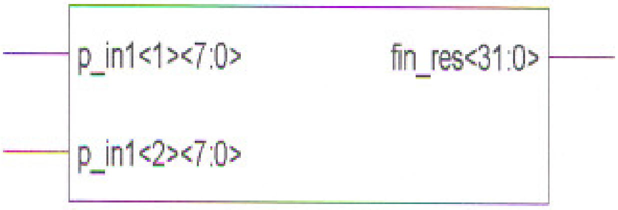


Figure 6.19 Main View of chip

The main chip consists of the following major blocks indicating the original structure of neural network shown in figures 6.19, 6.20, 6.21.

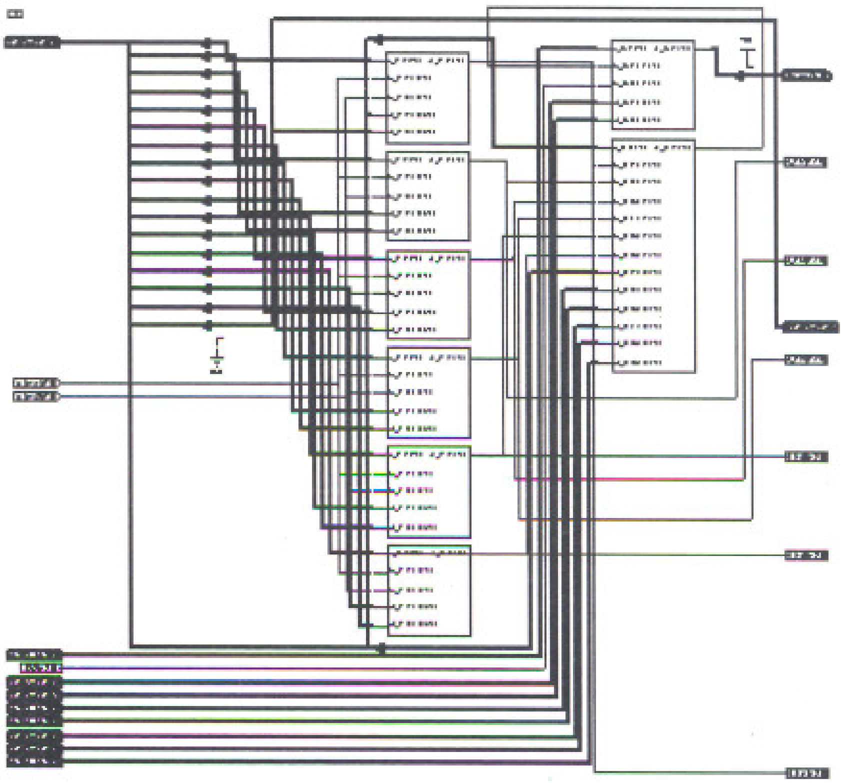


Figure 6.20 Detail_1 View of chip

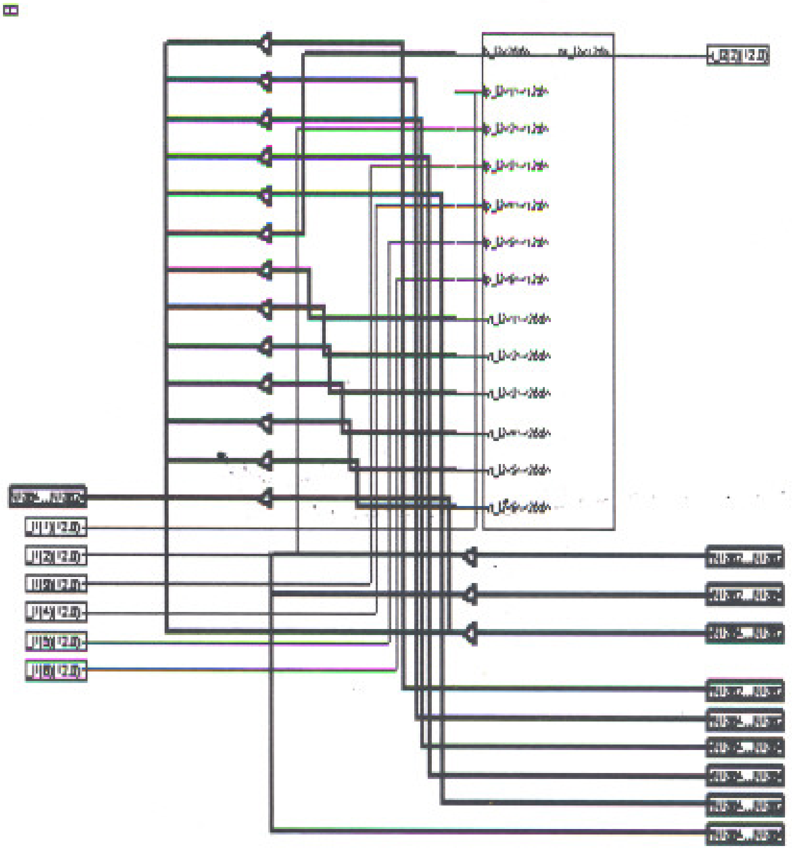


Figure 6.21 Detail_2 View of chip

Conclusions and Future Scope

In order to implement the handover circuit on an FPGA, we first created a fuzzy model for the handoff subsystem. Three inputs were selected as the inputs to model; these inputs are Received signal strength with threshold, Received signal strength with hystersis margin and time delay. These three inputs were observed to adequate to initiate a handoff decision. The rule base was generated with the help of experts from the field. The fuzzy model was simulated using Fuzzy Logic toolbox of MATLAB. The model performed adequately and was used to generate training data for our ANN based design. Section 5.2.1 discussed the design of fuzzy model.

In order to obtain the structure for ANN that could implement handoff initiation decision, we used Neural Network toolbox of MATLAB. We did numerous hit and trial observations and finally found that the performance of a 3 layered feedforward back propagation ANN with 2 hidden layers was adequate.

Once the computer simulation confirmed the performance, we implemented the ANN structure using VHDL. We designed the logsigmoidal neurons with activation function as $a = \frac{1}{1+e^{-n}}$. These VHDL entities were used as components to build structural system for ANN. The structural model was synthesized, simulated and finally the timing analysis was carried out. The performance of the ANN was found to be satisfactory. The design was implemented in Virtex II chip V1000bg560-6.

Further, another fuzzy inference system for the Soft handoff in CDMA is also designed so as to update the active set of mobile stations by evaluating the value of TDROP. The training data set required for the training of the ANNs is obtained from the FIS and the structure of the network is thus obtained. The design procedure for this application is similar to the one discussed above for the HO in GSM.

Once the structure is obtained, it is implemented in VHDL and we finally get the design on the FPGA.

However, a generalized feedforward backpropagation algorithm with 2 hidden layers is designed in VHDL, which can be use for implementing any desired application on an FPGA. The basic parameters required for this general network are the number of inputs to the system, the number and types of neurons in both the hidden layers and the number of outputs. This enables the user to get the design on a chip.

Future scope of the work

After we successfully designed, simulated and implemented the design we found that though this code can be adopted for any ANN with less than 12 neurons yet the designer has to design the structure by himself and make minor modifications in the VHDL code to use this code for his application. This becomes difficult for novices and for those who are not much familiar with both ANNs and VHDL.

To solve their problem the design process of ANNs need to be automated. A software can be developed that can automatically generate the structure of an ANN for a given problem (for a given set of input and output data). The VHDL code that can target the design on an FPGA can then be automatically generated. Automating this design process will make design of ANNs on an FPGA a fun. It will slash the design time drastically thus allowing a product to be launched in the market at its earliest. We have started the work in this direction but lot more need to be done!

References

[1] T.S. Rappaport, "Principles of Mobile Communications", Pearson Education Asia, 2nd ed., 2002.

[2] W.C. Y. Lee, "Mobile Communications Design Fundamentals" John Wiley & Sons Inc., 2nd ed., 1993.

[3] V.K.Garg, J.E.Wilkes, "Principles & Applications of GSM", Pearson Education Inc., 1st ed., 2002.

[4] B.J.Singh, S. Kumar, K.K.Aggarwal, "Handover Initiation Control Techniques in Mobile Cellular Systems" IETE Technical Review, Vol 20, pp 13-21, Jan-Feb 2003.

[5] N.D. Tripathi, N.Jeffrey, H.Reed, and H.F.Vanlandingham, "Handoff in Cellular Systems" IEEE Personal Communication, December 1998.

[6] G. Liodakis and P. Stavroulakis, "A novel approach in handover initiation for microcellular systems," Proc. 44th IEEE VTC, pp. 1820-1823, 1994.

[7] N. D. Tripathi, "Generic Adaptive Handoff Algorithms using Fuzzy Logic and Neural Networks" , Ph.D. Thesis, Blackburgs, Virginia, 1997.

[8] S. Abhyankar, "Handoff in Mobile Wireless Network" survey write-up, 2000.

[9] G.P.Pollini, "Trends in handover design", IEEE Communications Magazine, pp. 82-90, March 1996.

[10] G. Edwards and R. Sankar, "Microcellular handoff using neuro-fuzzy Techniques ", Wireless Networks 4 ,401–409, 1998.

[11] B.J. Singh, S. Kumar and K.K.Aggarwal, "A Fuzzy Based Multicriteria Handover Algorithm for Cellular Systems", CODEC-2001, REC, Jalandhar, pp 8-15, February 21-23 2001.

[12] W.Mende, "Evaluation of a proposed handover algorithm for the GSM cellular system", Proc. 40th IEEE VTC, pp. 264-269, 1990.

[13] B.J. Singh, S. kumar and K.K.Aggarwal, "Hysteresis plus Timer based Handover Initiation Algorithm for Microcellular Systems", IETE Technical Review, Vol. 21, No. 3, pp 175-182, May-June 2003.

[14] Jacek M. Zurada, "Introduction to Artificial Neural System", Jaico Publishing House, 1997.

[15] M. T. Hagan, H.B.Demuth, M.Beale, "Neural Network Design", Vikas Publishing house, 2002.

[16] S.V.Kartalopoulos, "Understanding Neural Networks and Fuzzy Logic", Prentice Hall of India, 1st ed., 2000.

[17] S.Haykin, "Neural Networks", Pearson Education, Asia, 2nd ed., 2002.

[18] B. Homnan, V. Kunsriruksakul and Benjapolakul, "Fuzzy Inference System based Adaptation of CDMA Soft Handoff Thresholds with Different Defuzzification Schemes", proc Singapore ICC, 1994.

[19] Douglas Perry, "VHDL", Tata McGraw Hill, 3rd ed, 2001.

[20] S. Sjoholm and L. Lindh, "VHDL for Designers", Prentice Hall, Europe, 1st ed., 1997.

[21] J.Bhasker, "VHDL Primer", Pearson Education, Asia, 2rd ed., 2001.

[22] D.Wong and T. Joon Lim, "Soft Handoffs in CDMA Mobile Systems", IEEE Trans. on Personal Communications, vol. 4, no. 6, pp. 6 – 17, Dec 1997.

[23] Web Pages
http://www.stanford.edu/group/wcs
www.global-source-mkt.com/cellphonefacts.html
www.cellphones.about.com
http://www.cdg.org

[24] S.Brown and J.Rose, "Architecture of FPGA and CPLDs : A Tutorial", Dept. of Electrical and Computer Engineering, University of Toronto.

List of Publications

[1] C.Dhawan, R.Khanna," Neural Networks Based Adaptive Handoff Algorithm", National Conference-TICE , feb. 2004, T.I.E.T., Patiala.

[2] C.Dhawan, R.Khanna ,S.Kumar, "ANN Based Mobile Controlled Handover System", National Conference on Electronic Circuits & Communication Systems, sept. 2004, T.I.E.T., Patiala.

[3] C.Dhawan, R.Khanna ,S.Kumar, "Design of ANN based Handover chip using VHDL", National Conference onTrends of Computational Techniques in Engineering, oct. 2004,S.L.I.E.T., Longowal.