Use of beamforming in cross-layer design for wireless communication systems

by

Deepali Arora

B.E, Karnataka University, 1996
M.A.Sc, University of Victoria, 2001

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

in the Department of Electrical and Computer Engineering

Use of beamforming in cross-layer design for wireless communication systems

by

Deepali Arora

B.E, Karnataka University, 1996
M.A.Sc, University of Victoria, 2001

Supervisory Committee

Dr. P. Agathoklis, Supervisor (Department of Electrical and Computer Engineering)

Dr. A. Antoniou, Department Member (Department of Electrical and Computer Engineering)

Dr. H-C. Yang, Department Member (Department of Electrical and Computer Engineering)

Dr. F. Diacu, Outside Member (Department of Mathematics)

Supervisory Committee

Dr. P. Agathoklis, Supervisor (Department of Electrical and Computer Engineering)

Dr. A. Antoniou, Department Member (Department of Electrical and Computer Engineering)

Dr. H-C. Yang, Department Member (Department of Electrical and Computer Engineering)

Dr. F. Diacu, Outside Member (Department of Mathematics)

Abstract

Wireless systems that have traditionally been designed using a layered approach have seen a recent paradigm shift to a cross-layered approach where the interactions between two or more layers are considered explicitly in an integrated framework. This dissertation presents new methodologies that aim to improve the performance of wireless systems through consideration of cross-layer based design.

The physical (PHY) and the medium access control (MAC) layers are the primary layers responsible for data transmission and user selection/control, respectively, in wireless systems. This dissertation begins with an analysis illustrating the use of multiple antennas and antenna arrays at the PHY layer. A framework combining space-time block coding and beamforming for uplink in a wireless systems is considered for studying the trade-offs between antennas and antenna arrays at the receiver. Results indicate that in high noise environments the diversity achieved by using a large number of antennas combats bit error rate (BER) more efficiently than

beamforming. On the other hand, in low noise environments beamforming plays an important role in reducing BER by minimizing the effect of interference from other co-channel users.

Two approaches of cross-layer design that are currently available are the bottom-up and top-down approaches. The bottom-up approach uses the PHY layer information at the MAC and higher layers to make decisions that affect the system performance. Following a bottom-up approach, a new scheduling algorithm is designed that uses the channel state information and direction of arrival information of mobile users to efficiently schedule users for service. Both semi-analytical (based on the probability density and cumulative distribution functions) and numerical frameworks are used to compare the performance of this algorithm with the traditional round-robin and greedy scheduling algorithms. Both the numerical and semi-analytical frameworks which are shown to be consistent with each other yield improved system capacity for the proposed algorithm compared to the traditional algorithms. This is the result of explicitly considering the angular location of mobile users around the base station that results into the reduced interference between simultaneously served users on one other. The effect of channel availability on the scheduling algorithms is also investigated within a queuing framework and the results indicated that the system performance is also dependent on channel availability and traffic conditions.

A top-down approach is based on modifying the PHY layer to support the requirements or protocols used at the MAC layer to improve system throughput. Following a top-down approach, a new methodology is presented that reduces null depths of a given beam to address the hidden beam problem in IEEE 802.11 systems. The hidden beam problem occurs in carrier sensing multiple access (CSMA) systems when mobile users lying in deep nulls are unable to sense an ongoing downlink transmission and start transmitting data in an uplink. The modified beam with reduced null depths is compared with the original beam in terms of reducing the hidden beam problem when

used in non-persistent CSMA systems. The modified beam is shown to improve the throughput of a slotted non-persistent CSMA system significantly when compared to original beam with relatively small changes to directivity and half-power beamwidth.

The bottom-up and top-down approaches used in this dissertation illustrate that by jointly addressing the PHY and MAC layer issues in an integrated cross-layer framework the performance of wireless systems can be significantly improved.

Table of Contents

List of Tables

List of Figures

List of Symbols

2G	Second-generation
3G	Third-generation
ACK	Acknowledged
ADSL	Asymmetric digital subscriber line
AMC	Adaptive modulation and coding
AMPS	Advanced mobile phone service
APTS	Acceptable power to send
ARQ	Automatic repeat request
AWGN	Additive white Gaussian noise
BER	Bit-error rate
BF	Beamformer
BPSK	Binary phase shift keying
BS	Base station
CDMA	Code division multiple access
CLD	Cross layer design
CSI	Cannel state information
CSMA	Carrier sense medium access
CSMA/CD	CSMA with collision detection
CTS	Clear to send
DAB	Digital audio broadcasting
D-Amps	Digital-Advanced Mobile Phone Service

dB	Decibel
DoA	Direcion of arrival
EDCA	Enhanced distributed channel access
EGPRS	Enhanced general packet radio service
ESPRIT	Estimation of signal parameters via rotational invariance technique
EV-DO	Evolution-Data Optimized
FDMA	Frequency division multiple access
GSM	Global System for Mobile communications
HCCA	HCF controlled channel access
HCF	Hybrid coordination function
HDR	High data rate
HOL	Head of line
HPBW	Half power beam width
HSDPA	High-Speed Downlink Packet Access
IS-95	Interim Standard 95
ISI	Inter symbol interference
ISO	International Organization for Standardization
LAN	Local area network
LoS	Line of sight
MAC	Medium access control
MACAW	Multiple access with collision avoidance for wireless
MAN	Metropolitan area network
MIMO	Multiple input multiple output
M-LWDF	Modified largest weighted delay first
MUSIC	Multiple signal classification
MVDR	Minimum variance distortionless response
NDMA	Network-assisted diversity multiple access

NLoS	Non line of sight
OFDM	Orthogonal frequency division multiplexing
OSI	Open system interconnection
PCM	Power control MAC
PCMA	Power controlled multiple access
PDF	Probability density function
PHY	Physical
QoS	Quality of service
RF	Radio frequency
RPTS	Request power to send
RR	Round-robin
RTS	Request to send
Rx	Receiver
SDMA	Space division multiple access
SINR	Signal to interference and noise ratio
SNR	Signal to noise ratio
STBC	Space-time block coding
TC	Traffic classes
TDMA	Time division multiple access
TX	Transmitter
UMTS	Universal Mobile Telecommunications System
USDC	U.S. digital cellular
UWB	Ultra wideband
VoIP	Voice over internet protocol
VoWLAN	Voice over wireless LAN
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access

WLAN	Wireless local area network

Acknowledgements

I owe my gratitude to several people who have made this dissertation possible and because of whom my graduate experience will be a memorable one.

My deepest gratitude is to my advisor, Dr. Panajotis Agathoklis who taught me how to question thoughts and express ideas. I am deeply grateful to him for the long discussions that helped me sort out several technical details of my work and for holding me to a high research standard. His patience and support in completing this dissertation is greatly appreciated.

I am grateful to Dr. W. S. Lu and Dr. K. F. Li for their continuous encouragement and guidance. I am also thankful to Dr. Andreas Antoniou for encouraging the use of correct grammar and for carefully reading and commenting on my reports and this manuscript.

I am grateful to the members of my supervisory committee for reading this manuscript and helping me to improve it. I am also indebted to the members of the digital signal processing research group with whom I have interacted during the course of my graduate studies. I also appreciate the efforts of the many graduate students for their contributions to the various domains.

I am grateful to Mary-Anne for providing me teaching assistant opportunities, Vicky for always helping me and answering my questions and emails so efficiently, Lynne Barrett for always encouraging me and listening to me when I needed someone to talk to and Moneca for helping me with administrative stuff. I am also thankful to the system staff Eric Laxdal and Steve Campbell.

Most importantly, none of this would have been possible without the love and patience of my family.

Chapter 1

Introduction

1.1 Introduction

Wireless communications is one of the most exciting areas in the field of communications today. The past decade has seen a surge of research activities due to increasing demand to support not only voice but also data and multimedia services over wireless systems. Some of the recent technologies designed or proposed in the wireless area include the development of third generation (3G) systems (CDMA 2000, UMTS), 3.5G systems(HSDPA, EV-DO), wireless LAN (IEEE 802.11 a/b/g), ultra-wideband (UWB) systems and Wi-MAX (IEEE 802.16) systems. These technologies have spurred a lot of research in the signal processing and cross-layer design for wireless communication areas. Providing reliable communications over a wireless medium is a challenging task. Some of these challenges include the rapid channel fluctuations due to fading, interference from other users, limited power, inability to work efficiently in systems following layered approaches and network security. Several new techniques (orthogonal frequency division multiplexing, adaptive modulation and coding, adaptive beamforming, multiple antenna systems and multiuser diversity within cross-layer system design) are available for improving wireless system performance. Discussion of all these techniques in detail is beyond the scope of this dissertation and the focus is on the techniques summarized below. The limita-

tions of these techniques are also briefly mentioned and they form the basis of this dissertation.

1.2 Motivation

1.2.1 Multiple antennas/antenna array systems

To improve the performance of wireless systems, two promising technologies, multiple antenna systems and beamforming have emerged that attempt to reduce the effect of fading and interference respectively. Both these techniques improve the system capacity by exploiting spatial diversity. In the case of multiple antenna systems, multiple versions of the same signal are transmitted and/or received at the receiver based on the fact that individual channels between the transmitter and receiver experience different fading environments and this increases the reliability of the transmitted signal. The beamforming technique aims at minimizing the effect of interference and saves power by directing the signal in the desired direction and forming nulls in other directions. Both these techniques have been studied extensively over the last few years and different techniques have been developed for their design and implementation [1], [2], [3]. The performance of the system where these two techniques are used together to reduce bit error rate for a given environment (combining both fading and interference from other users) remains an active research topic [4].

1.2.2 Cross-layer design

In early communication models such as the open system interconnection (OSI) model, the network functionality was divided into several sub-modules, also referred to as layers, where each layer was assigned a pre-defined task with no cross-layer interactions. Techniques such as multiple antenna systems and beamforming at the physical layer (PHY) and various scheduling algorithms at the medium access control (MAC) layer were designed by considering other layers as black boxes. The layered approach worked efficiently for wired networks that did not suffer from channel variations.

However, when used in wireless systems the layered approach proved to be inefficient. For example, the use of space-time block coding schemes at the PHY layer when scheduling delay sensitive traffic over the MAC layer is a perfect example illustrating the failure of layered approach in wireless systems. This is because using space-time block coding, although a diversity gain can be achieved but the data is transmitted in two time steps [5] thus leading to increased delay. It has slowly become evident that to improve the reliability of wireless systems it is necessary to have some sort of interactions between different layers of the network architecture and this led to the design of cross-layer systems. The scheduler [6] in this sense becomes one of the most important components to achieve improved cross-layer performance as it acts as a focal point of interaction between PHY and higher layers.

Recent cross-layer based scheduling algorithms exploit channel state information (CSI) available at the PHY layer to make scheduling decisions at the MAC layer [7]. These algorithms attempt to improve the system capacity by serving the user with the best channel conditions and thus exploits multiuser diversity (the greedy approach). Such algorithms work efficiently when a single user is served as in time division multiple access (TDMA) systems. However, when used in systems that serve more then one user, the greedy approach suffers from performance degradation due to the interference created by simultaneously served users on one another. Consideration of CSI alone at the MAC layer does not provide an optimal solution for improving system capacity in cross-layered systems and new algorithms are needed that can reduce the interference that simultaneously served users cause on one another.

1.2.3 Effect of channel availability

Another important issue to be considered while assessing the performance of scheduling algorithms is the impact of channel availability. Most scheduling algorithms are based on the assumption that once users are selected based on certain scheduling criteria they are served immediately regardless of channel availability. A channel is

said to be unavailable if its instantaneous signal to noise ratio (SNR) is below some minimum specified threshold. Serving users when the channel is in poor state leads to performance degradation. Tsibonis et al. [8], for example, designed a scheduling policy for a wireless system with time-varying connectivity, for general arrival rates. They found that scheduling policies proposed for maximizing throughput under various assumptions on the arrival rates may fail when channel availability is taken into account. To assess the behaviour of scheduling algorithms in a more comprehensive manner it becomes necessary to take channel availability into account.

1.2.4 Effect of beamforming in carrier sense medium access (CSMA) based systems

The increased use of wireless systems to support multimedia services requires careful consideration of the advancements at the PHY layer in designing higher layer protocols within a cross-layered framework. The implementation of advanced technologies at the PHY layer is not straightforward and presents its own share of problems. One such example is the implementation of beamforming at the PHY layer when using medium access control protocols such as the CSMA. Users in CSMA systems rely on sensing the medium before accessing it and when the base station uses beamforming to serve a desired user in a downlink mode the deep nulls in the direction of undesired users can lead to hidden beam, hidden terminal and deafness problems. These problems arise because of the inability of users to sense an ongoing transmission which leads to collisions and packet loss.

1.3 Scope and contribution of the dissertation

The main objective of this dissertation is to investigate the effect of techniques that improve the performance of wireless networks at the PHY layer and MAC layer within a cross-layer framework to support mutiple applications using either fixed (TDMA) or random (CSMA) channel assignment techniques. The dissertation consists of seven

chapters which are organized as follows.

In Chapter 2, a brief introduction to cross-layer systems is presented. Some of the current cross-layer research areas are investigated, possible future research problems are presented and the limitations of using cross-layered systems are also summarized.

In Chapter 3, the effect of using multiple antennas and beamforming in a combined framework for systems affected by both fading and interference is presented. The limitations of multiple antennas and beamforming for environments characterized by different SNR and signal to interference ratio (SIR) are also highlighted. Chapter 3 provides a detailed understanding of the effect of using multiple antennas and beamforming systems at the PHY layer. The simulation model presented in Chapter 3 confirmed that the use of multiple antennas in high noise and fading environments and the use of beamforming in multiuser environment are beneficial in reducing BER.

In Chapter 4, following a bottom-up approach, a new scheduling algorithm is designed and presented, that attempts to improve system capacity by considering both the channel state information and angular location of mobile users around the base station while scheduling users such that the simultaneously served users causes less interference on one another. Both numerical and semi-analytical (based on probability density and cumulative distribution function) approaches are used in Chapter 4 and the performance of the proposed scheduling algorithm is compared with that of the traditional greedy and round-robin algorithms in terms of system capacity. The results from both the numerical and semi-analytical model indicated that the system performance can be improved by considering not only the CSI alone but also the direction of arrival (DoA) information when scheduling users.

In Chapter 5, the effect of channel availability on the performance of cross-layered systems presented in Chapter 4 is further investigated within a queuing framework. An M/M/1 queuing system with exponentially distributed inter-arrival and inter-departure times is used to model the arrival and departure of the packets. The

performance of three scheduling algorithms presented in Chapter 4 is further assessed both numerically and semi-analytically (based on probability density and cumulative distribution function) in terms of signal to interference plus noise ratio in a model that considers channel availability. The effect of traffic conditions on system performance is assessed and finally the buffer size of the queue is fixed to some finite value to compare the performance of three scheduling algorithms in terms of packet loss and delay. The results obtained in Chapter 5 show that the improvement in system capacity obtained by considering DoA information along with CSI while scheduling users in Chapter 4, is still maintained when the channel availability is taken into account.

In Chapter 6, a top-down approach is used to modify the original beam for addressing the hidden beam problem in CSMA systems. The modified beam is compared with the original beam in terms of beam parameters such as directivity and half-power beam width as well as the effect of implementation of the modified beam in non-persistent CSMA systems is investigated in terms of system throughput. It was found that the modified beam when implemented in non-persistent CSMA systems exhibits the same throughput as that obtained using omni-directional antennas while still maintaining the benefits of directional antennas in terms of power savings when compared to omni-directional antennas.

Finally in Chapter 7, the work presented in Chapters 2 to 6 is summarized. Chapter 7 also presents some of the possible future research directions.

Chapter 2

Cross-layer design for wireless systems

2.1 Introduction

Current demand for wireless technology in several applications varying from Bluetooth enabled mobile devices, cameras and printers to wireless local area network (WLAN) enabled computers and handheld devices places a stringent demand on the quality of service (QoS). In order to support both real time traffic such as voice, multimedia conferences and games, and data traffic such as web browsing, messaging and file transfer with sufficient QoS, the traditional layered protocol architecture has to be re-designed to allow interaction between the layers, thus leading to cross-layer design (CLD). In this chapter, a brief overview of CLD including its emergence and advantages and disadvantages are presented. An overview of current research in the area of CLD is presented and some open research problems that formed the basis of this dissertation are highlighted.

2.2 Historical background

The International Organization for Standardization (ISO) began in the early 1980s to work on an open set of protocols that would enable multi-vendor computers to interact and communicate with one another. This work eventually led to the design of the open system interconnection (OSI) network stack which was sought to become

the building block of all network based communication worldwide [9]. The seven layered architecture of an OSI model divided the overall networking task into layers and defined a hierarchy of services to be provided by the individual layers. At different layers the services were provided using specific protocols. This architecture forbids direct communication between non-adjacent layers; communication between adjacent layers is limited to procedure calls and responses [10]. The OSI model was eventually replaced by a simplified five layered model as shown in Figure 2.1 but the task of the layers in the five layered model was still pre-defined like the OSI model. These layered architectures were easy to design and upgrade since the designers had to design protocols such that a higher-layer protocol only makes use of the services at the lower layers and is not concerned about the details of how that service is being provided. It has been realized [10] that the traditional layered approach is inefficient in meeting QoS to support current traffic (that is a mix of both real and non-real time applications) in wireless systems thus leading to the development of the cross-layer design.

2.3 Cross-layer design for wireless communications

Cross-layer design is based on the interaction between different layers to improve the overall system throughput. The layered architecture used in wireline systems has proven to be inefficient in wireless systems because they have limited channels and frequencies available for transmission and also suffer from severe variations in channel due to multipath fading or interference from users operating at the same or adjacent frequencies referred to as co-channel and adjacent channel interferences respectively. In order to mitigate the effect of fading and interference, multiple antennas [11], [12] ,[13] and antenna array systems [14], [15], [16], [17] were developed at the physical (PHY) layer and they offer promising results in terms of minimizing bit error rate (BER) but they have their own limitations. One such example, as mentioned earlier,

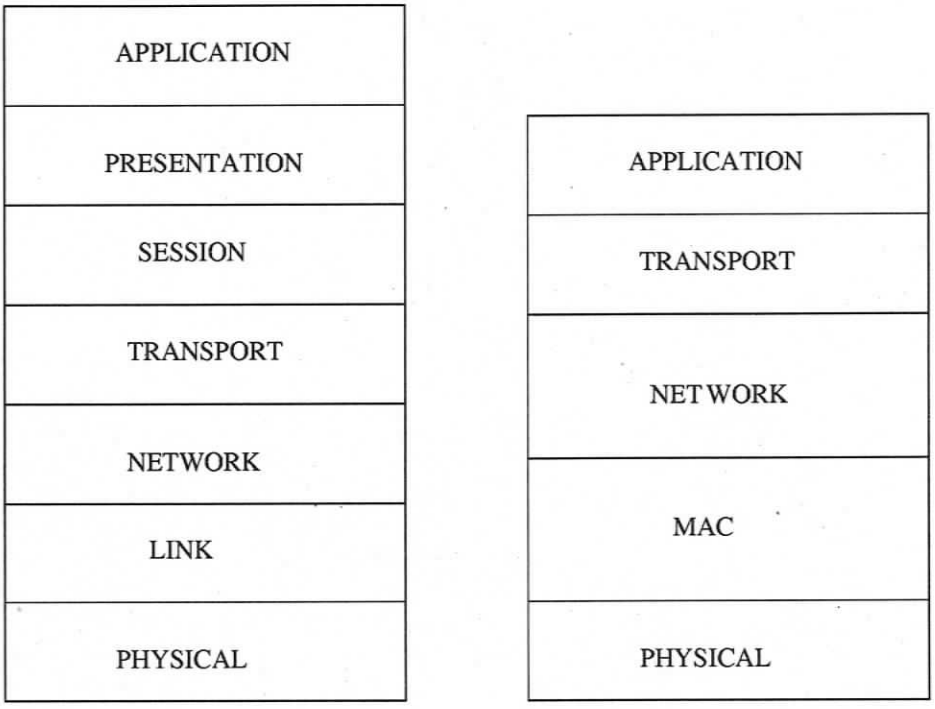


Figure 2.1: Layered architectures of OSI model and simplified wireless network model.

is scheduling delay sensitive traffic over PHY layer that uses multiple antennas for space-time block coding.

In order to minimize BER while providing sufficient QoS to both real and non-real time traffic the wireless system design needs to consider cross-layer design. CLD exploits the interaction between the different layers that are traditionally not possible in wired networks and promotes adaptability at all layers based on information exchange between them. The exchange of information between layers and their consideration in management of wireless systems leads to overall improvement of QoS.

2.3.1 Role of different layers of simplified wireless network layered model

The role of different layers of a five layered reference model consisting of application, transport, network, link (comprising of data link control and medium access control (MAC) sub-layers), and physical layers [18] are briefly discussed in this section.

1. The PHY layer: The PHY layer is special compared to other layers of the model because it is the only one where data is physically moved across the network interface. All of the other layers perform useful functions to create messages to be sent but they must all be transmitted down the protocol stack to the physical layer where they are actually sent out over the network. Some of the important functions of the PHY layer include defining hardware specifications, coding and modulation functions that transform the data from bits that reside within a computer or other device into signals that can be sent over the network, and transmitting and receiving the data.

2. The MAC layer: The MAC data communication sub-layer acts as an interface between the network's physical layer and higher layers. It provides addressing and channel access control mechanisms that are required for several terminals or network nodes to communicate within a multipoint network, typically a local

area network (LAN) or metropolitan area network (MAN).

3. The network layer: The network layer defines how interconnected networks function. Some of the other key functions of the network layer are logical addressing, routing, datagram encapsulation, and fragmentation and reassembly.

4. The transport layer: The transport layer provides the necessary functions to enable communication between software application processes on different computers. The transport layer is also responsible for defining the means by which potentially large amounts of application data are divided into smaller blocks for transmission. Another key function of the transport layer is to provide connection services for the protocols and applications that run at the levels above it.

5. The application layer: The application layer is used by network applications. Application protocols defined at this layer implement specific user applications and other high-level functions. Since they are at the top of the stack, application protocols are the only ones that do not provide services to a higher layer and they make use of services provided by the layers below.

2.4 Different methodologies to design cross-layer based systems

The introduction of the cross-layered approach in wireless systems has led to avenues for exploration of opportunistic communication that could not be exploited sufficiently in a strictly layered design. For instance, the time-varying link quality allows the opportunistic usage of the channel, whereby the transmission parameters can be dynamically adjusted according to the variations in the channel quality. Additionally, the wireless medium offers some new modalities of communication the layered architectures do not accommodate. For instance, the physical layer can be made

capable of receiving multiple packets at the same time. The nodes can also make use of the broadcast nature of the channel and cooperate with one another in involved ways. To explore the advantages offered by the cross-layered approach, it becomes essential to understand some of the design methodologies. A number of cross-layer design proposals have been recently proposed in the literature [19]. In this section a brief overview of some of the possibilities of coupling different layers in a cross-layer design are presented. These include creation of new interfaces, merging of adjacent layers, design coupling without new interfaces and vertical calibration across layers. As mentioned in [19], most of the current cross-layer design proposals in the literature can easily fit into one of these basic categories.

2.4.1 Creation of new interfaces

Several cross-layer designs require creation of new interfaces between the layers. These interfaces are used for information sharing between the layers in real time. The architecture violation here is obviously the creation of a new interface not available in the layered architecture. Information flow between the layers can be from higher to lower layers or vice-versa, or bi-directional and can be categorized into three main categories.

Upward information flow

In this case the higher layer protocol requires some information from the lower layers resulting into the creation of a new interface between the lower and the higher layer. One prominent example of upward information flow is multiuser diversity where the PHY layer information is passed to the MAC layer to make scheduling decisions.

Downward information flow

Some cross-layer design proposals rely on setting parameters on the lower layer of the stack using a direct interface from higher layers. For example, the application layer can inform the MAC layer about their delay requirements and the MAC layer

can then treat packets from delay-sensitive applications with priority [20].

Back and forth information flow

Two layers, performing different tasks, can collaborate with each other. Often, this manifests in an iterative loop between the two layers, with information flowing back and forth between them. For example, in the network-assisted diversity multiple access (NDMA) proposal [21] the PHY and MAC layers collaborate in collision resolution in an uplink wireless LAN system. Another example of back and forth information flow between layers is seen in algorithms performing joint scheduling and power control in wireless adhoc networks [22].

2.4.2 Combining adjacent layers

Cross-layer systems can be designed by combining two or more adjacent layers together such that the new superlayer provides all services of the constituent layers. This does not require any new interfaces to be created in the stack. Architecturally speaking, the superlayer can be interfaced with the rest of the stack using the interfaces that already exist in the original architecture. For example, the collaborative design between the PHY and MAC layers tends to blur the boundary between these two adjacent layers.

2.4.3 Vertical calibration across layers

As the name suggests, this approach adjusts parameters that span across layers. Basically, the performance seen at the application layer is a function of the parameters at all the layers below it. An example of vertical calibration is presented in [23] where the delay requirements dictate the persistence of link-layer automatic repeat request (ARQ) which in turn becomes an input for deciding the rate selection through a channel-adaptive modulation scheme.

2.5 Role of the PHY and MAC layers in CLD

While each layer of the protocol stack plays a specific role, the PHY and MAC layers are considered the most important in improving system reliability and throughput [21]. Higher layers interact with the PHY layer to tune their parameters such as modulation and coding scheme [24], [25], transmission power [26] or data rate. In this section some of the recent advancements made at both the PHY and MAC level to improve system performance are discussed briefly.

2.5.1 Advancements at the PHY layer

The role of the PHY layer is reliable transmission/reception of data over the wireless medium. Data is transmitted from source to destination through a channel which can be accessed using either fixed or random channel assignment techniques. In fixed channel assignment techniques users share a given channel resource e.g., a frequency band, a time-slot, or both. These methods make relatively efficient use of radio resources. Some of the earlier fixed channel assignment techniques developed were time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA) as shown in Figure 2.2 . In TDMA time is divided into intervals of regular length and each interval is subdivided into slots. Each user is assigned a slot number and can transmit over the entire bandwidth during its slot within each interval. U.S. digital cellular (USDC) (also called IS-54/IS-136) and global system for mobile (GSM) are examples of TDMA. In FDMA, the available radio spectrum is divided into channels of fixed bandwidth, which are then assigned to different users e.g., advanced mobile phone service (AMPS). In CDMA, all users share the available bandwidth all the time. The manner in which available bandwidth is shared is based on a code or pattern that is unique to each user. The receiver knows the pattern of time/frequency use of the users and can distinguish them accordingly. US CDMA cellular (IS-95), wideband CDMA (W-CDMA) and wireless LANs (IEEE

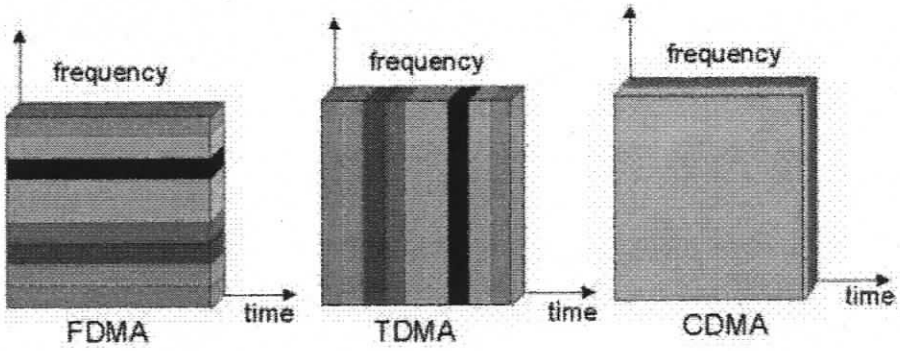


Figure 2.2: Fixed channel assignment technique (FDMA, TDMA, CDMA).

802.11b, 802.11g) are standards based on the CDMA technology.

In the case of packet switching, the channel assignment at the PHY layer is not fixed but instead it is random in which a data sequence from a digital source is broken down into smaller pieces which are organized into data packets. Packets are transmitted to a destination through a shared radio network without explicit channel assignment. Since the channel is shared, protocols must be observed to assure the fair and orderly transfer of data. One of the earliest multiple access protocol developed was the ALOHA protocol developed at the University of Hawaii for bursty low-data-rate transmission over satellite systems. According to ALOHA protocol a user transmits as soon as a packet is ready to go. If two or more users starts transmitting simultaneously then collision occurs. This protocol is simple, but it exhibits low throughput due to high collisions. Other forms of ALOHA are slotted ALOHA in which transmission can occur only at the beginning of specific time slots (doubles throughput) and reservation ALOHA in which a transmitter with a long file can reserve slots. Other types of multiple access protocols are based on sensing the medium. For example, in the carrier sense medium access (CSMA) protocol the transmitter "listens" to the medium to see if the channel is idle (i.e., no

carrier is detected). If the channel is idle, the user transmits. Collision occurs if two users start transmitting simultaneously. CSMA is used in wireless LANs. Variants of CSMA include, 1-persistent CSMA in which a packet is transmitted as soon as the channel is idle, non-persistent CSMA in which negative acknowledged packets are retransmitted only after a random amount of time and CSMA with collision detection (CSMA/CD) in which the transmitter listens while transmitting to see if anyone else is also transmitting. If so, transmission is aborted immediately. The CSMA/CD protocol is used in Ethernet.

The medium over which data are transmitted is not error free and the data received are not the same as those transmitted due to corruption from fading and/or interference. Various new technologies have been introduced at the PHY layer both from the signal processing and information theory point of view [5]. Some of these technologies are discussed below.

Adaptive modulation and coding (AMC)

In cellular communication systems, the quality of a signal received by a mobile depends on a number of factors such as the distance between the desired and interfering base stations, path loss exponent, shadowing, fading and noise. In order to improve system capacity, peak data rate and coverage reliability the signal transmitted to and by a particular user is modified to account for the signal quality variation through a process commonly referred to as link adaptation. Traditionally, CDMA systems have used fast power control as the preferred method for link adaptation. Recently, AMC have offered an alternative link adaptation method that promises to raise the overall system capacity [27]. AMC provides the flexibility to match the modulation-coding scheme to the average channel conditions for each user. With AMC, the power of the transmitted signal is held constant over a frame interval, and the modulation and coding format are changed to match the current received signal quality or channel conditions. The main advantages of AMC are that it increases the overall through-

put and reduces the interference. AMC is now being used both in GSM and CDMA based systems as well as wireless LANs.

Orthogonal frequency division multiplexing (OFDM)

Orthogonal frequency-division multiplexing is a frequency-division multiplexing (FDM) scheme utilized as a digital multi-carrier modulation method [28]. A large number of closely-spaced orthogonal sub-carriers are used to carry data. The data are divided into several parallel data streams or channels, one for each sub-carrier. Each sub-carrier is modulated with a conventional modulation scheme at a low symbol rate, maintaining total data rates similar to conventional single-carrier modulation schemes in the same bandwidth. OFDM has developed into a popular scheme for wideband digital communication. The primary advantage of OFDM over single-carrier schemes is its ability to cope with severe channel conditions, narrowband interference and frequency-selective fading due to multipath, without complex equalization filters. Channel equalization is simplified because OFDM may be viewed as using many slowly-modulated narrowband signals rather than one rapidly-modulated wideband signal. OFDM is capable of handling time-spreading and thus eliminates intersymbol interference (ISI). OFDM is also used in 802.11a WLAN, 802.16 and WiMAX technologies.

Antenna diversity

To mitigate the effect of fading, a technique has been developed in which the information carrying signal is transmitted along different propagation paths. This can be achieved by using multiple receiver antennas (diversity reception), by using multiple transmitting antennas (transmit diversity) or both (multiple input multiple output (MIMO), Figure 2.3). The signals can be sent using different antennas either in the same or different time slots. Signals undergoing both spatial and temporal diversity are mainly designed following space time block or Trellis coding [11]-[13]. Some of

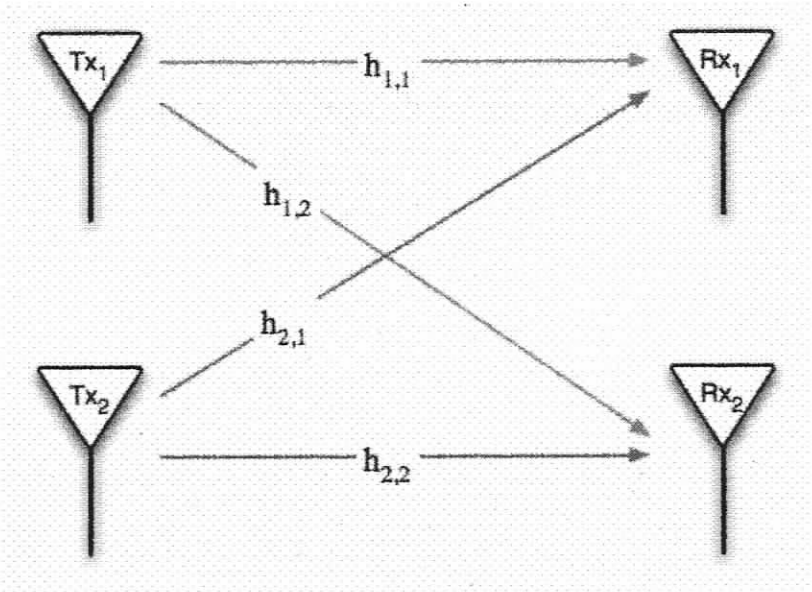


Figure 2.3: Multiple input multiple output (MIMO) antenna systems.

the applications where antenna diversity is used include Wi-Fi, GSM and CDMA based mobile phones and the IEEE 802.11n standard.

Use of adaptive antenna arrays or smart antennas for beamforming

Smart antennas technology offers a significantly improved solution to reduce interference levels and improve system capacity. With this technology, each user's signal is multiplied by complex weights that adjust the magnitude and phase of the signal to and from each antenna. This causes the output from the array of antennas to form a transmit/receive beam in the desired direction and minimizes the output in other directions. This drastically reduces the overall interference in the system. This method of transmission and reception is called beamforming and is made possible through smart (advanced) signal processing at the base station. Beamforming can be either switched or adaptive. If the complex weights are selected from a library of weights that form beams in specific, predetermined directions, the process is called switched beamforming. Here, the base station essentially switches between the different beams based on the received signal strength measurements. On the other hand,

if the weights are computed adaptively in real time the process is called adaptive beamforming. Through adaptive beamforming, the base station can form narrower beams towards the desired user and nulls towards interferring users, considerably improving the signal-to-interference-plus-noise ratio. Beamforming is used in GSM and CDMA systems and in wireless LANs IEEE 802.11b/g standard.

2.5.2 Advancements at the MAC layer

While the PHY layer is mainly responsible for data transmission, it is the role of the MAC layer to decide which user will be served based on the traffic load. Analyzing the network traffic, designing scheduling algorithms and re-designing MAC layer protocols has been the main focus of the research community working on the MAC layer. In this section some of the recent developments at the MAC layer are discussed in brief.

Analyzing network traffic

Network traffic is mix of both real and non-real time services but modelling network traffic that resembles realistic behaviour is a challenging task. The simplest traffic model used earlier in wired systems and now in wireless systems yield Poisson distributed traffic. However, it has been realized that the Poisson model does not characterizes the network traffic realistically which exhibits self-similar behavior. This led to the development of self similar traffic models which exhibit similar behaviour at all time scales. Although self similar models were developed earlier for wired networks they are now also being used for wireless networks [29].

Power control MAC (PCM)

Power savings is one of the important requirements of current wireless systems and has been associated with the PHY layer in the past. Recently, it has been realized that an efficient MAC layer protocol design can lead to significant power savings. Some of the techniques developed at the MAC layer are briefly discussed in this

section.

- PCM: A PCM protocol is proposed in [30] to allow per-packet selection of transmit power. In PCM, request to send/clear to send (RTS/CTS) packets are transmitted with maximum power level, P_{max} but the data packets are transmitted with a lower power level. In order to avoid a potential collision caused by the reduced carrier sensing zone, during the data packet transmission PCM periodically increases the transmission power to P_{max}. Acknowledged (ACK) packets are transmitted with the minimum required power to reach the source node. By periodically increasing the power level for data transmission, PCM effectively reduces the number of possible collisions. This way, retransmission is avoided as much as possible, and correspondingly, the goal of energy savings is achieved. Results show that PCM can achieve a throughput comparable to that of the IEEE 802.11 but with less energy consumption. However, PCM requires a frequent increase and decrease in transmission power levels.

- Power Controlled Multiple Access (PCMA): The PCMA Protocol [31] proposes a flexible "variable bounded power" collision suppression model and allows variable transmit power levels on a per-packet basis. Similar to IEEE 802.11, PCMA uses request-power-to-send/acceptable-power-to-send (RPTS/APTS) handshake to determine the minimal transmission power required for successful packet reception. The difference lies in that PCMA introduces a second channel, the busy tone channel, to implement the noise tolerance advertisement. During data transmission periods, each active receiver will periodically send a busy tone to advertise the maximum additional noise power it can tolerate. Any potential transmitter must first sense the channel for busy tones to determine the upper bound of its transmit power for a minimum time period (determined by the frequency with which the busy tones are transmitted). PCMA thus uses

the signal strength of a received busy tone message to bound the transmission power of neighboring nodes. In this manner, power control mechanism is realized and spatial reuse is achieved.

Supporting advanced services over MAC layer

Voice over internet protocol (VoIP) is becoming increasingly popular. However, it is a delay sensitive service and the traditional MAC layer cannot support this functionality efficiently. New standards are continuously designed that provide QoS enhancements for wireless LAN applications through modifications to the MAC layer. IEEE 802.11e used in Wi-Fi multimedia is one such example. The 802.11e defines a new coordination function known as the hybrid coordination function (HCF) [28]. Within the HCF, there are two methods of channel access, HCF controlled channel access (HCCA) and enhanced distributed channel access (EDCA). Both EDCA and HCCA define traffic classes (TC). For example, emails could be assigned to a low priority class, and voice over wireless LAN (VoWLAN) could be assigned to a high priority class.

Modifying the random access control techniques

New protocols such as multiple access with collision avoidance for wireless (MACAW) were developed to support services over wireless medium. MACAW [32] is based on the previously proposed MACA (multiple access with collision avoidance) scheme [33], which advocated that each transmission be preceded by a handshake between the sender and receiver to reserve the channel for a period of time. Before sending data a node sends a request to send (RTS) message as long as it has not heard any clear to send (CTS) message in the recent past. If the intended receiver hears this message and has not heard any other RTS then it responds with a CTS. The original RTS message includes the amount of time that the sender wishes to reserve the channel, as does the CTS (subtracting out the RTS time). A node that does not

get a response to its RTS sets an exponential backoff timer and retries the RTS after a random period of time chosen from the backoff interval.

This RTS-CTS-data approach reduces the number of collisions because a sender can send only if it has not heard any recent CTS (so it's own transmission cannot interfere with another node's reception), and a sender can send only if the receiver has not heard any RTS (so there is no other sender in the receiver's vicinity). Of course, in practice, radio frequency (RF) data reception may not be smooth and other factors can corrupt packet delivery, but the approach does have its merit. Moreover, it does not require any carrier sensing support.

2.6 CLD examples of PHY and MAC layer integration

As mentioned earlier, the two layers that are considered to play a vital role in cross-layer design are the PHY and the MAC layers. These two layers have been merged in an integrated framework using several design methodologies. While some of these design methodologies are based on using PHY layer information at the MAC layer, others design approaches exploited the MAC layer information at the PHY layer to improve system performance. In this section some of the examples based on these design methodologies are presented.

2.6.1 Multiuser diversity

In order to improve the system throughput, the users take advantage of the channel state information available at the PHY layer to make scheduling decisions at the MAC layers by prioritizing the user with best channel conditions, leading to multiuser diversity gain. Algorithm that gives higher priority to users with good channel conditions over users with bad channel conditions is also sometimes referred to as the opportunistic or the greedy scheduling [34]. Examples of systems using greedy algorithms include high data rate (HDR) versions of CDMA 2000 [35] and enhanced general packet radio service (EGPRS) in the EDGE extension of GSM [36]. In [37], Tse et

al. proposed a scheduling rule called the proportionally fair rule, which explicitly makes use of the channel state information (CSI) and also provides fair allocation of bandwidth across users. Since the development of greedy and proportional fair schedulers, a number of cross-layer designs have been proposed [38], [39], [40], [41], [42] that use the CSI in scheduling users following either the greedy or proportional fair approach at the MAC layer and use single or multiple antenna systems at the PHY layer.

2.6.2 Multiuser diversity with QoS guarantees

One important issue that must be taken into consideration while designing cross-layered systems is the QoS requirements of different users. QoS is assessed in terms of SNR/SINR for single/multiple users and the delay suffered by the packets. While multiuser diversity gain based algorithms offer promising results in terms of throughput they do not address delay requirements for delay sensitive traffic such as voice or video. For example, scheduling algorithms presented in [38]-[43] consider only the PHY layer QoS issues. These algorithms consider the CSI available at the PHY layer to schedule users at the MAC layer with the aim of maximizing SNR/SINR. The performance of algorithms designed specifically for single user case [42]-[43], however, degrades significantly in the presence of multiple users due to interference. The algorithms that do focus on maximizing SINR [38]-[40] in the presence of multiple users suffer from increased computational complexity associated with finding the best combination of users that may be served simultaneously. None of these algorithms consider the effect of delay at the MAC layer. Scheduling algorithms that jointly consider the channel state information from the PHY layer and delay information from the MAC layer [44], [45], [46] are also available. The authors of [44] proposed modified largest weighted delay first (M-LWDF) algorithm that uses the product of CSI and delay associated with the packets waiting at the head of line of queues into consideration while prioritizing the users for service. This metric ensures that users

with either higher SNR or experiencing large delay or both are preferentially served.

2.6.3 Multiple access protocols based on multiuser diversity

The traditional ALOHA [47] and CSMA [48] protocols have been re-designed to take advantage of the channel state information available at the PHY layer by allowing the user with the best CSI to access the medium. Results indicate that by using CSI available at the PHY layer, both these protocols exhibit higher throughput compared to the scenarios when CSI is not used.

2.6.4 Combining beamforming at the PHY layer with multiple access protocols at the MAC layer

Traditional MAC protocols (e.g., CSMA) are designed based on the assumption that the underlying PHY layer uses the omni-directional antennas. However, with the advancements in antenna technology the omni-directional antennas are being replaced by directional antennas which when used in systems relying on carrier sensing can no doubt result in huge power savings and better directivity but can also cause severe performance degradation due to the hidden node problem, the hidden beam problem and deafness [49]. Deafness occurs when the transmitter fails to communicate to its intended receiver, because the receiver's antenna is oriented in a different direction. The directional hidden terminal problem occurs when one mobile terminal fails to hear a prior RTS and CTS exchange between another mobile terminal and base station and starts transmitting thereby causing collision. The hidden beam problem occurs due to simultaneous uplink/downlink transmission where a mobile user fails to see an ongoing downlink transmission and starts transmitting in uplink thus causing collision. To address the problems caused by directional antennas in systems based on carrier sensing, it becomes necessary to consider changes either at the PHY or MAC layer within a cross-layered system. A MAC protocol in which request to send and clear to send signals are sent omni-directionally while data and acknowledgment

are sent directionally have been proposed [50] which reduces the hidden node and deafness problems. Recently, the hidden beam problem also has been addressed [51], [52] using a beamforming methodology that uses directional antenna to send signals to desired users and omni-directional antenna to send the broadcast signal.

2.7 Cross-layer design skepticism

While CLD in wireless adhoc networks and cellular networks has received a lot of attention in the last few year, there are critics who argue to exercise caution against use of CLD approaches [53] for the following reasons:

1. A modular architectural design has proven its reliability over time. Modularity provides the very essence of abstraction necessary for researchers and engineers to fully understand any system. The layered OSI reference model is used as an example over which the Internet and its tremendous success is based on. A modular design can also accelerate development since designers can focus their efforts on different subsystems with the assurance that the entire system will interoperate once it is brought together. This is not the case for CLD.

2. CLD by nature creates interactions between processes and layers. Some interactions are intentional, others might be unintentional. CLD can thus create loops, and it is well known from control theory that stability becomes an issue under such conditions.

3. Standardization allows subsystems to be used across many applications, thus resulting in lower development cost which in turn increases usage. In contrast, a CLD based system will need to be adapted for every application and this will likely increase the cost.

4. Once the layering is broken, the possibility to review and redesign parts of the system is lost since everything is interconnected one way or the other. Protocols

have to be re-implemented in a cross-layered fashion, taking into account several layers as opposed to earlier where a protocol could be developed in isolation.

5. A system-wide CLD can lead to a "spaghetti" implementation, which in turn will hamper further innovation and be difficult to maintain. Future design improvements may become impossible because it will be difficult to foresee exactly how a new modification will affect the overall system operation.

The main point raised by [53] is that while CLD gives a short term performance gain, good architectures are usually based on longer term consideration. Clearly the CLD concept is relatively new and it is prudent to use CLD with caution. While the authors of [53] raise valid concerns about CLD, there is no denying that the CLD can provide significant gains in improving the performance of the wireless systems and its use is increasing significantly over time.

2.8 Open challenges in cross-layer design

Cross-layer design is a relatively new area of research and offers several research possibilities some of which are listed below.

2.8.1 Important cross-layer couplings

One of the most important questions while designing CLD systems is the choice of the layers that should be merged to yield the highest throughput with minimum power consumption. A thorough cost-benefit analysis of different cross-layer design proposals in terms of implementation complexity versus performance improvement yields insight into this. Generally speaking, the following inferences can be drawn from the current literature. Cross-layer design is needed between the network and MAC layers for adhoc networks since the functionalities of the two layers interact [54]; explicit notifications by new interfaces to the transport layer improve end-to-end performance [19]; making use of channel knowledge at the MAC layer allows opportunistic usage

of the channel and improves performance [54]; and energy, delay and security related issues need to be handled across the layers in a holistic manner. To move ahead from these general insights to specific solutions requires comparative quantitative study of the different cross-layer design proposals. For example, opportunistic scheduling leads to higher signal to noise ratio (SNR) in a single user system but when implemented in a multiuser system can lead to severe performance degradation. Design of scheduling algorithms that can serve more than one user while considering the impact of interference they create on one another is an important research topic.

2.8.2 Maintaing QoS in cross-layer designs

The channel conditions in a wireless network are time-varying. One of the stated motivations behind cross-layer is to make the system responsive to variations in the underlying channel conditions so that minimum QoS is always maintained. Designing a system model that considers channel fading before carrying out data transmission in order to maintain sufficient QoS is a challenging task.

2.8.3 Role of the PHY layer

In wired networks the role of the physical layer has been rather small; sending and receiving packets when required to do so from the higher layers. As mentioned earlier, the advances in signal processing allow the PHY layer to play a bigger role in wireless networks. For example, designing algorithms and protocols at the MAC layer by taking PHY layer parameters into consideration or re-designing the PHY layer based on MAC layer information to increase system throughput. Cross-layer designs relying on advanced signal processing at the PHY layer provide a number of interesting opportunities.

2.8.4 The right communication model

Wired networks are essentially a collection of well defined point-to-point communication links. The same cannot be said about wireless networks because the wireless

channels suffer from both fading and interference. This gives rise to a fundamental question as to which technologies offer more benefits in yielding high system performance when implemented in wireless networks. Design of the right communication model while taking into account the type of application, the channel access scheme, the power saving requirements, and channel variations is an important research issue.

2.9 Conclusions

This chapter focussed on some of the shortcomings of the layered architecture when used in wireless systems that led to the development of cross-layer systems. Some of the current cross-layer proposals are discussed and some cautions required in designing cross-layer systems are also investigated. Finally, some of the open problems in the area of cross-layer systems that formed the basis of research in this thesis are also highlighted.

Chapter 3

Role of multiple receive antennas and antenna arrays at the physical layer

3.1 Introduction

The physical (PHY) layer plays an important role since it is responsible for the actual transfer of data. However, the data transferred through the PHY layer are not error free because of the signal attenuation between the source and destination path (path loss), signal absorption in local structures such as the mountains, buildings and trees (shadowing), signal variation due to the time varying nature of the channel (fading) and signal interference due to the angular distribution of subscribers (co-channel and adjacent channel interference) [4], [55]. To improve the signal quality over the PHY layer, there has been increased use of multiple antennas [11]-[13] and antenna arrays [14]-[17]. Multiple antennas can be used to exploit both spatial and temporal diversity based on space-time coding (STC) techniques [11], [13]. Antenna systems are usually spaced 3λ to 30λ (where λ is the wavelength) apart from one another and are used to mitigate the effect of fading by transmitting the signal through different propagation paths. Antenna array systems are mainly used to minimize the effect of interference and the inter-array spacing is usually around half the wavelength i.e., $\lambda/2$. Studies have demonstrated that the use of beamforming alone at the base station does not

yield the best performance [56] and it is practical to use both antennas and antenna arrays in an integrated framework to combat noise, fading and interference.

This chapter deals with the behaviour of space-time block coding (STBC) and beamforming, the two primary technologies to combat noise, fading and interference in a combined framework. The behaviour of antennas (that are used to combat noise and fading) and antenna arrays (that are used to combat interference) is investigated in a framework of varying noise, fading (characterized by line of sight (LoS) and non-LoS (NLoS) component) and interference environments (characterized by the number of co-channel users interfering with the desired signal). The system considered is a combination of Alamouti's STBC algorithm (with two transmit and variable receive antennas) and a minimum variance distortionless response (MVDR) based beamforming algorithm (with a variable number of antenna arrays) to study relative merits of antennas and antenna arrays in the presence of channel noise and co-channel interference in an uplink environment. The objective is to understand how a combination of antennas and antenna arrays can be used efficiently given computational and other resources.

A number of studies have exploited various aspects of combining STC and beamforming based techniques [57], [58], [59], [60], [61], [62], [63] for both downlink [57]-[61] and uplink systems [62]-[63]. Zaragoza and Ghavami [57] adopted a flexible approach, where they used either Alamouti's [11] or Tarokh's [13] STC schemes, depending on the number of beams calculated using beamforming for downlink transmission. Katz et al. [58] used space-time coded beams for downlink transmission of the signals in the directions determined using beamforming in outdoor environments. Lei et al. [59] used a beamforming technique at the base station for space-time coded downlink transmission to maximize the signal to noise ratio (SNR) at the mobile. Their results indicated that the combined beamforming and space-time coded model outperforms pure diversity or pure beamforming based models in terms of bit error rate (BER)

especially in multiuser environments. Tarokh et al. [60] proposed a combined space-time coding and array processing approach in which the antennas at the transmitter are partitioned into small groups using individual space-time codes. At the receiver the antenna arrays are used to accept a single group as the desired group in a given time slot while other groups are treated as interference sources whose effect is reduced using antenna arrays. Wang et al. [61] combined beamforming with STB encoder at the transmitter to steer the beam in the direction of the desired mobiles and used antenna arrays at the receiver to accept the desired signal while minimizing interference from other users for space-time decoding. These and other studies such as [64], [65], [66], [4] have focussed on analyzing relative merits/demerits of using antennas and antenna arrays but considered only either antenna or antenna array configuration at a time.

In this chapter a system model that uses various combinations of antennas and antenna arrays is used at the PHY layer and its performance is investigated under various conditions. Although an increased number of transmit (Tx) and receive (Rx) omni-directional antennas reduce the effect of fading and an increased number of antenna elements in directional antennas reduce interference, there is an upper limit beyond which adding more antennas/antenna arrays for a given environment leads to diminishing returns in terms of improvement in the system capacity or reducing BER as shown later in this chapter. The chapter is organized as follows. The proposed model with Alamouti's two transmitter (Tx) and N receiver (Rx) STBC scheme and the MVDR scheme is presented in section 3.2. Simulations results from different combinations of antennas and antenna arrays are presented and their performance in terms of BER in the presence of both noise and co-channel interference is discussed in section 3.3. The trade-offs between antennas and antenna arrays are discussed in section 3.4 and conclusions are presented in section 3.5.

Table 3.1: Combinations of number of receive antennas and antenna elements used.

Receive antennas N	Receive antenna elements L
1	8
2	4
4	2
8	1

3.2 System model

A multiuser uplink model with K co-channel users is considered. The signal from only one out of K users (the p^{th} user) is considered as the signal of interest and the other $K-1$ users are treated as interference sources. All users are assumed to be randomly distributed between -90^0 and $+90^0$ around the base station. It is assumed that each mobile has two transmit antennas and the base station has N receive antennas. The antennas are assumed to be separated by a distance of more than 10λ metres where λ is the wavelength. Both the number of antennas (N) and the number of antenna array elements in each receive antenna, L, at the base station, are varied such that their product remains constant equal to 8 ($L = 8/N$) as shown in Table 3.1. The objective of using larger L for smaller N is to assess the relative advantage of beamforming in mitigating the effect of interference in the presence of reduced receive diversity. Similarly, using larger N for smaller L helps in analyzing the relative advantage of receive diversity versus beamforming. The two receive antennas at the base station each consisting of four antenna elements is denoted by 2Rx/4L and a similar notation is used for all other combinations. The multiple antennas are used to perform space-time encoding using Alamouti's STBC scheme and the antenna elements are used for beamforming. The block diagram of the proposed model is shown in Fig. 3.1. Binary phase shift keying (BPSK) modulation is used to modulate the signals.

Let the signals transmitted by the mobile of interest at time t over its antennas 1 and 2 be s_0 and s_1, respectively, followed by their conjugates $-s_1^*$ and s_0^* at time

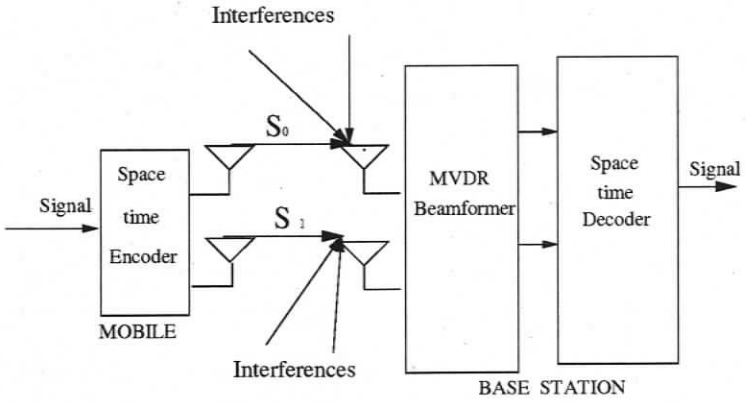


Figure 3.1: Combined beamforming and co-channel interference cancellation model for a space time coded system

$(t+T)$ where $(*)$ denotes the complex conjugate. These signals arrive at the base station via channel h, where $h = \alpha e^{j\theta}$, α is the channel amplitude response and θ is the phase shift. The amplitude response α of the channel is modelled using Rayleigh or Ricean distribution depending on the absence or presence of line of sight (LoS) component. The probability density function (pdf) of α is given by [67]

$$p(\alpha) = \frac{\alpha}{\sigma^2} \exp\left(\frac{-\alpha^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{A\alpha}{\sigma^2}\right) \tag{3.1}$$

where the parameter A denotes the peak amplitude of the dominant signal and $I_0(*)$ is the modified Bessel function of the first kind and zeroth order and σ^2 is the variance. Higher values of A imply stronger LoS component compared to other multipath components. As $A \rightarrow 0$, the dominant signal fades away and the Ricean distribution degenerates to the Rayleigh distribution. The phase shift θ is assumed to be uniformly distributed between 0 and 2π. The channels between each pair of transmit and receive antennas are assumed to be independent and identically distributed (i.i.d) due to enough separation between them. It is also assumed that the total average power at the transmitter is equally divided between the two transmitter antennas. In the presence of co-channel interference arriving at the base station from various

directions, the received signals can be obtained by multiplying both the signal of interest and the interfering signals with their respective steering vectors $\mathbf{v}$ and $\mathbf{u}$, given by [2]

$$\mathbf{v} = [1 \;\; e^{j\kappa d \sin(\eta_p)} \cdots \;\; e^{j\kappa(L-1)d\sin(\eta_p)}]^T \tag{3.2}$$

$$\mathbf{u_i} = [1 \;\; e^{j\kappa d \sin(\eta_i)} \cdots \;\; e^{j\kappa(L-1)d\sin(\eta_i)}]^T \quad \text{where } i = 1, 2, \ldots, K,\ i \neq p \tag{3.3}$$

where η_p represents the direction of arrival (DoA) of the primary signal, η_i, $i = 1, 2, \ldots, K$, $i \neq p$ represents the direction of arrival of $K-1$ interferring signals, d represents the inter-antenna element spacing usually taken as $\lambda/2$, L is the number of antenna elements and $\kappa = 2\pi/\lambda$.

The signals received at the N receive antennas are the combinations of the signal of interest (s_0, $-s_1^*$, s_1 and s_0^*), interferring signals from $K-1$ co-channel users and additive white Gaussian noise (AWGN) noise and can be expressed as

$$[\mathbf{Y}] = \sqrt{\frac{E_s}{2}}[\mathbf{H}][\mathbf{S}] + [\mathbf{U}] + [\mathbf{N}] \tag{3.4}$$

where $$[\mathbf{Y}] = \begin{bmatrix} \mathbf{y_0}(t) & \mathbf{y_1}(t+T) \\ \mathbf{y_2}(t) & \mathbf{y_3}(t+T) \\ \vdots & \vdots \\ \mathbf{y_{(2N-2)}}(t) & \mathbf{y_{(2N-1)}}(t+T) \end{bmatrix}$$

$$[\mathbf{H}] = \begin{bmatrix} \mathbf{v}h_0 & \mathbf{v}h_1 \\ \mathbf{v}h_2 & \mathbf{v}h_3 \\ \vdots & \vdots \\ \mathbf{v}h_{(2N-2)} & \mathbf{v}h_{(2N-1)} \end{bmatrix}, \quad [\mathbf{S}] = \begin{bmatrix} s_0(t) & -s_1^*(t+T) \\ s_1(t) & s_0^*(t+T) \end{bmatrix}$$

$$[\mathbf{U}] = \begin{bmatrix} \sum_{i=1,i\neq p}^{K} \zeta_{0_i}(t)\mathbf{u_i} & \sum_{i=1,i\neq p}^{K} \zeta_{1_i}(t+T)\mathbf{u_i} \\ \sum_{i=1,i\neq p}^{K} \zeta_{2_i}(t)\mathbf{u_i} & \sum_{i=1,i\neq p}^{K} \zeta_{3_i}(t+T)\mathbf{u_i} \\ \vdots & \vdots \\ \sum_{i=1,i\neq p}^{K} \zeta_{(2N-2)_i}(t)\mathbf{u_i} & \sum_{i=1,i\neq p}^{K} \zeta_{(2N-1)_i}(t+T)\mathbf{u_i} \end{bmatrix}$$

and $$[\mathbf{N}] = \begin{bmatrix} \mathbf{n_0}(t) & \mathbf{n_1}(t+T) \\ \mathbf{n_2}(t) & \mathbf{n_3}(t+T) \\ \vdots & \vdots \\ \mathbf{n_{(2N-2)}}(t) & \mathbf{n_{(2N-1)}}(t+T) \end{bmatrix}$$

where E_s is the average energy available at the transmitter over a symbol period. The interfering signals ζ_{0_i}, ζ_{1_i},...,ζ_{2N-1_i} $(i = 1, 2, \ldots, K,\ i \neq p)$ are generated on the basis of specified SIR and their DoAs are randomly generated. The effect of co-channel interference is minimized by steering the main beam towards the primary user by estimating an optimum weight vector ($\mathbf{w}$) such that the power contributed by interferring signals is minimized. For the model considered here, the MVDR beamforming algorithm is used to find this optimum weight vector. The signals received at the base station are multiplied with the optimum weight vectors and the resultant signals are written as

$$x_{2k}(t) = \mathbf{w}_{2k}^{H}(t)\mathbf{y}_{2k}(t) \tag{3.5}$$

$$x_{2k+1}(t+T) = \mathbf{w}_{2k+1}^{H}(t+T)\mathbf{y}_{2k+1}(t+T) \quad \text{where } k = 0, 1, \ldots, N-1 \tag{3.6}$$

where $(\cdot)^H$ represents the conjugate transpose. $\mathbf{w}$ for each receive antenna at each time step, following the MVDR beamforming algorithm [68], is obtained as

$$\mathbf{w_n}(t) = \frac{\mathbf{R_{U_n}}(t)^{-1}\mathbf{v}}{\mathbf{v^H}\mathbf{R_{U_n}}(t)^{-1}\mathbf{v}} \quad \text{where } n = 0, 1, \ldots, 2N-1 \tag{3.7}$$

where $\mathbf{R_{U_n}}(t)$ is the covariance matrix of the interferring signals given by

$$\mathbf{R_{U_n}}(t) = E[\mathbf{U_n}\mathbf{U_n}^H(t)] \tag{3.8}$$

and $\mathbf{U_n} = \sum_{i=1,i\neq p}^{K} \zeta_{n_i}(t)\mathbf{u_i}$.

In practice, the covariance matrix $\mathbf{R_{U_n}}(t)$ is rarely known and $\mathbf{w_n}$ is calculated using an estimate of the covariance matrix $\hat{\mathbf{R}}_{\mathbf{U_n}}(t)$

$$\hat{\mathbf{R}}_{\mathbf{U_n}}(t) = \frac{1}{K}\sum_{i=1,i\neq p}^{K} \mathbf{U_{n_i}}(t)\mathbf{U_{n_i}}^H(t) \tag{3.9}$$

The resultant signals obtained after interference cancellation (Eqns. (3.5)-(3.6)) at each receive antenna are combined according to the Alamouti's STBC scheme.

$$\begin{bmatrix} \hat{\mathbf{s}}_0 \\ \hat{\mathbf{s}}_1 \end{bmatrix} = \begin{bmatrix} h_0^*(t) & h_1(t+T) & \cdots & h_{2N-2}^*(t) & h_{2N-1}(t+T) \\ h_1^*(t) & -h_0(t+T) & \cdots & h_{2N-1}^*(t) & -h_{2N-2}(t+T) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1^* \\ x_2 \\ x_3^* \\ . \\ . \\ . \\ x_{2N-2} \\ x_{2N-1}^* \end{bmatrix} \tag{3.10}$$

Finally, the signals transmitted from the mobile using two transmit antennas are estimated at the receiver using Alamouti's space-time block decoder as $\hat{\mathbf{s}}_0$ and $\hat{\mathbf{s}}_1$, respectively.

3.3 Numerical analysis

The system model presented in the previous section was simulated and the results are used to illustrate the performance of the combined STBC and beamforming using different combinations of number of receive antennas and antenna arrays at the base station. Simulations were carried out for 10^5 symbols. The total number of users K is assumed equal to 10.

Space-time block encoding is applied at the mobile while beamforming and space-time decoding is carried out at the base station. It is assumed that the receiver has perfect knowledge of the channel state and the DoA for each user which can be obtained using pilot signals respectively [69], [70]. The performance is measured in terms of bit error rate (BER) for different signal to interference ratio (SIR) and SNR values.

3.3.1 Effect of antenna diversity on the system performance

Fig. 3.2 shows the simulation results for two transmitters and number of receive antennas equal to 1, 2, 4 and 8 for a Rayleigh faded environment. SNR is assumed to be equal to 20 dB and beamforming is not performed ($L = 1$). The effect of spatial diversity associated with increasing receive antennas is seen here. As expected, the BER decreases with an increased number of receive antennas. An increase in number of receive antennas between the mobile (transmitter) and the base station (receiver) reduces the probability of all branches experiencing simultaneous fading leading to an improvement in BER. BER is maximum for the pure transmit diversity case with two transmitter and one receive antenna.

3.3.2 Effect of antenna arrays

Fig. 3.3 shows the simulation results for two transmitters and one receiver case with variable number of antenna elements (L) for SNR equal to 20 dB, varying SIR and a Rayleigh faded environment. These simulations are performed to assess

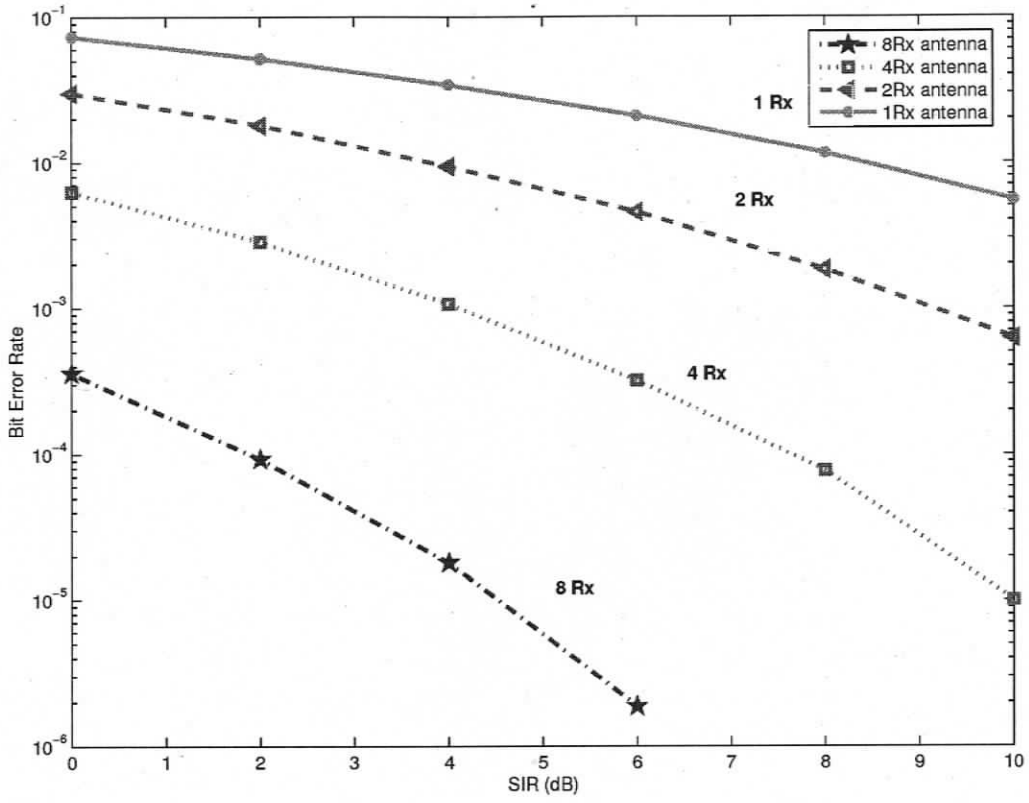


Figure 3.2: Effect of diversity using two transmitters and variable number of receive antennas (Tx=2, Rx=1,2,4,8) on BER for SNR=20 dB and varying SIR.

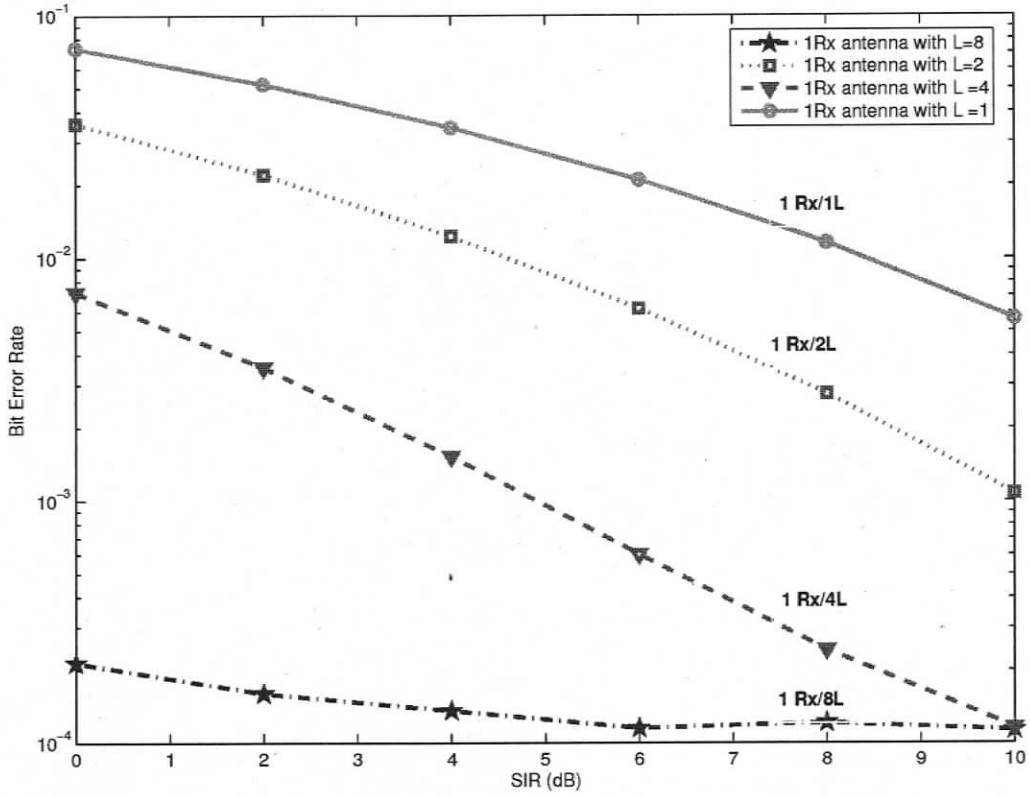


Figure 3.3: Effect of beamforming on BER for SNR=20 dB using two transmitters and one receive antenna with variable number of antenna array elements for varying SIR.

the beamforming-based gain. As expected, an increase in antenna array elements allows more interference cancellation leading to a reduction in BER for all SIRs. In the case of 8 antenna elements, beamforming using MVDR can effectively minimize interference completely from 7 users. The resulting BER is due to interference from the remaining users and/or fading and background noise.

3.3.3 Combined effect of antennas and antenna arrays

The simulation results discussed in sections 3.3.1 and 3.3.2 illustrate how the increasing number of antennas combat noise and how the increasing number of antenna array elements combats interference, respectively. In this section, simulation results are pre-

sented for combinations of varying number of antennas and antenna array elements as shown in Table 3.1 to illustrate the trade offs between beamforming and receive diversity. Fig. 3.4 shows the BER for these cases when SNR is equal to 0 dB. Fig. 3.5 shows the results for SNR equal to 20 dB. As expected, a reduction in BER is seen with increasing number of receive antennas. Beamforming offers some but relatively little improvement in reducing BER in the case of 0 dB SNR while the improvement using beamforming is much higher for the case of 20 dB SNR. This indicates that in an environment with high noise relative to co-channel interference most of the improvement comes from increase in receive diversity while in an environment with high co-channel interference relative to noise most of the improvement comes from beamforming.

The comparison of 2Rx/4L and 4Rx/2L cases is interesting to illustrate this. For SNR = 0 dB, i.e. in the presence of significant noise (Fig. 3.4), 4Rx/2L performs better than 2Rx/4L. For SNR = 20 dB (Fig. 3.5), on the other hand these two combinations yield similar BER for all SIRs. Note that 1Rx/8L in Fig. 3.5 yields similar BER for different SIR because large number of antenna elements are able to reduce the effect of interference significantly and thus the remaining BER is due to constant noise and fading present in the system. However, as the total number of users increases from 10 to 20, 8 antenna elements are unable to combat all the interference thus resulting into higher BER as shown in Fig. 3.6.

3.3.4 Effect of the fading environment

It is interesting to ask if the trade offs observed in the previous simulations depend on the fading environment. In Fig. 3.7 a comparison of the performances for Rayleigh ($A = 0$) and Ricean ($A = 2$) fading environments for the two cases, 4Rx/2L and 2Rx/4L, is shown. The SNR in these simulations is 10 dB. As expected, BER is lower for Ricean fading environment due to the presence of line of sight components that strengthen the signal. However, the relative advantage of the 4Rx/2L case

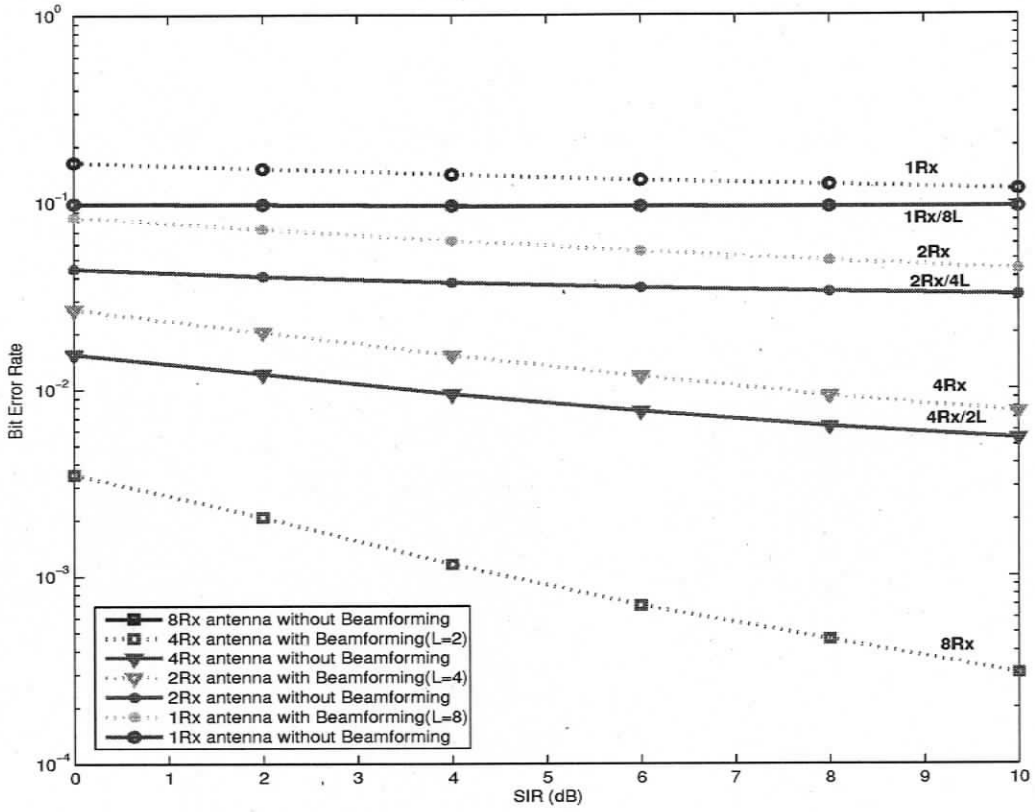


Figure 3.4: Effect of diversity and beamforming on BER for SNR=0 dB. Solid lines show simulation results with beamforming and dotted lines are for results without beamforming (L=1).

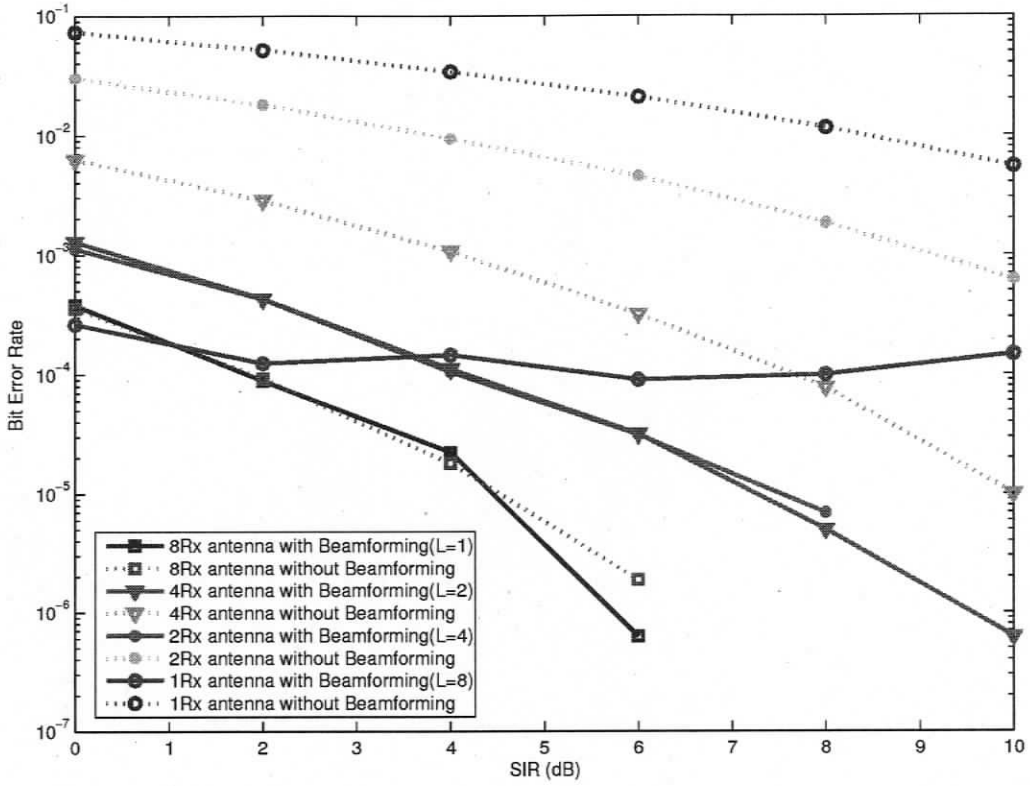


Figure 3.5: Effect of diversity and beamforming on BER for SNR=20 dB. Solid lines show simulation results with beamforming and dotted lines are for results without beamforming (i.e., L =1).

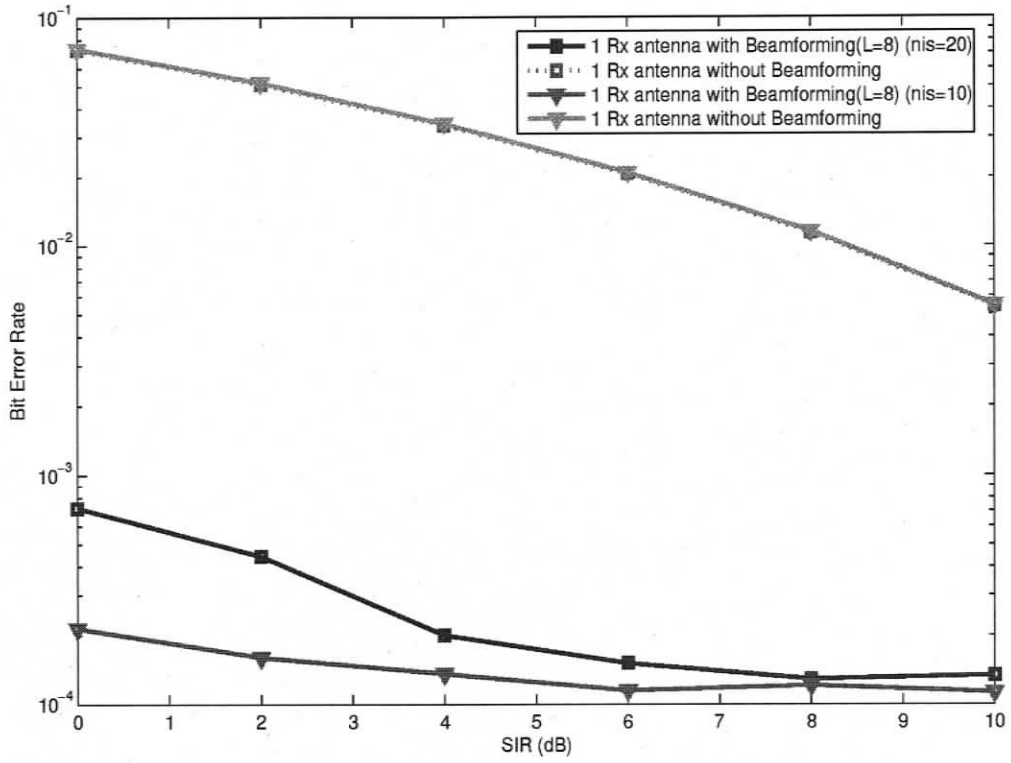


Figure 3.6: Effect of beamforming on BER for SNR=20 dB, for different numbers of total users. Solid lines show simulation results with beamforming and dotted lines are for results without beamforming (i.e., $L = 1$).

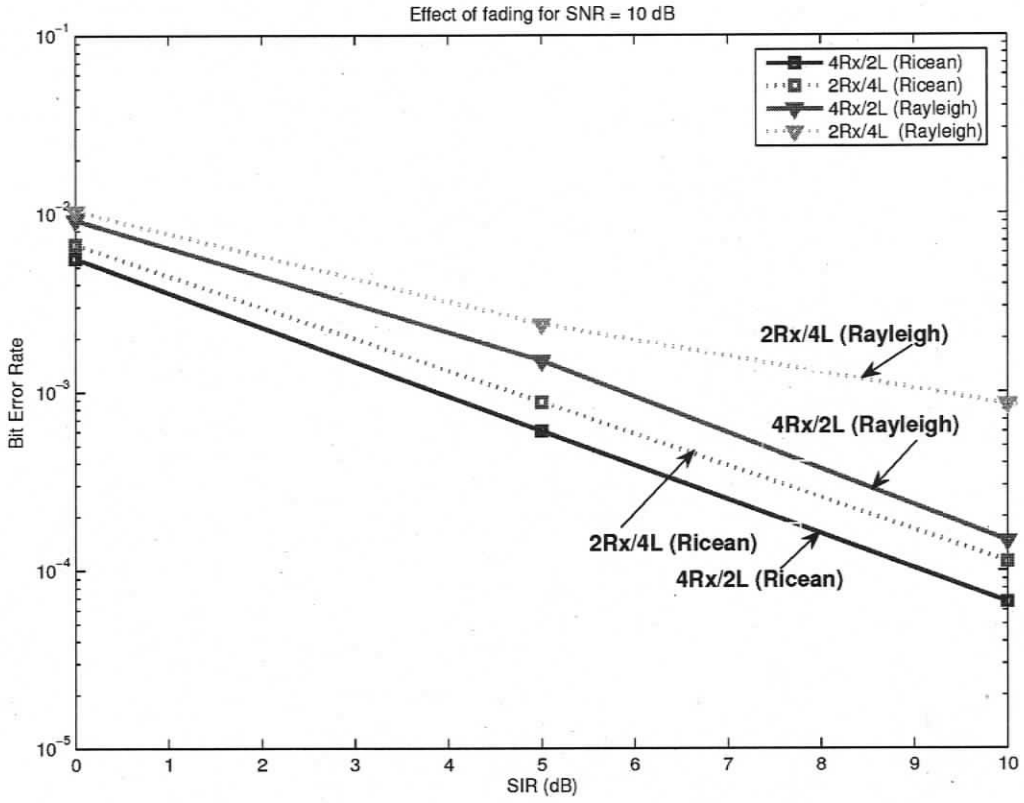


Figure 3.7: Effect of Rayleigh and Ricean fading environments on BER for SNR=10 dB and varying SIR.

over the 2Rx/4L is still maintained independently of the fading environment. The behaviour observed in the previous simulations, i.e that the diversity gains in high noise environments play a primary role in reducing BER and that beamforming is more effective when SNR increases, is not affected by the fading environment which only changes the absolute magnitude of BER.

3.3.5 Effect of the number of users

Further insight into the trade offs between various combinations of antenna and antenna arrays at the receiver is gained when simulations are performed for different number of users and when mobiles are located in narrow ($\pm 5°$) and wide ($\pm 90°$)

angular range around the base station. In Fig. 3.8 the BER for 2Rx/4L and 4Rx/2L cases is presented for Rayleigh faded environment, SNR = 20 dB and varying number of mobile users. Fig. 3.8 shows that 2Rx/4L combination performs better for small number of interferring users then the 4Rx/2L combination for both narrow and wide range of angular spread of mobile users around the base station because of higher interference cancellation capacity. For large number of interferring users, however, the 2Rx/4L combination performs better only for narrow angular spreads while for wider angular spreads the performance of both combinations is similar. This is expected because as the angular spread becomes larger, then the relative advantage of more antenna elements disappears. In addition, the effect of angular spread on BER is larger for 4Rx/2L combination and the 2Rx/4L combination is relatively insensitive to changes in angular spread. These results are consistent with earlier studies [4], [64]-[66] which have considered the trade offs between antennas and antenna arrays independently.

3.4 Discussion

The results presented in this chapter have yielded insight into the performance of integrated antenna and antenna array systems and how it depends on the fading environment, the presence or absence of interferring sources and the angular spread between the interferring users. In the case of high noise, high fading environments and absence of interferring sources, the diversity achieved by using large number of antennas plays a dominant role in minimizing the BER. In the case of multiple interferring users, the role of beamforming becomes important in minimizing the effect of interference from other co-channel users by directing the beam towards the desired user. Small angular spreads between the users lead to higher interference incurred by undesired users on the desired user and in such scenarios it becomes desirable to use beamforming. For large angular spread between the users, on the

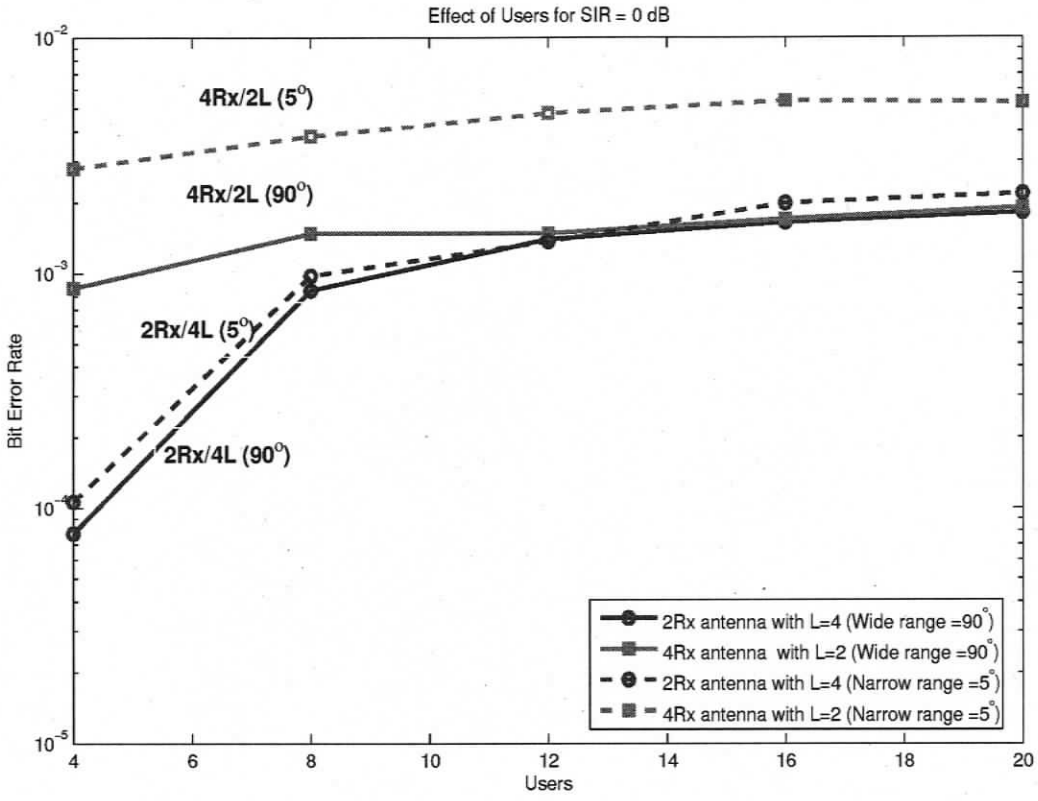


Figure 3.8: Effect of number of users on BER for SIR=0 dB, SNR = 20 dB

other hand, the role of diversity in minimizing the BER caused due to fading and noise becomes important.

The analysis presented in the previous section is based on the assumption of knowledge of the DoA and channel state information. DoA information is needed for beamforming while channel state information is required for receiver diversity. Any implementation of a system like the one presented in the previous section has to include techniques for obtaining such information. Blind DoA estimation techniques such as MUSIC or ESPRIT [71] are quite computationally demanding and also the number of antenna elements limits the number of users whose DoA can be estimated. Recently, a DoA estimation technique which uses the pilot signal of each user has been developed [70]. This technique is computationally efficient and the number of users whose DoA is estimated is not limited by the number of antenna array elements. Channel estimation at the receiver is an important part of using diversity since without channel knowledge it is difficult to estimate the desired signal. There are several techniques to obtain channel state information using pilot signal, or feedback from transmitter or using the reciprocity principle [1].

The consideration of a system consisting of antennas and antenna arrays also entails antenna spacing requirements. The inter-antenna element spacing is usually taken as $\lambda/2$, where λ is the wavelength. A value larger than $\lambda/2$ results in large lobes pointed towards other users which can degrade the performance of a system. On the other hand, a value less then $\lambda/2$ leads to reduction in the directivity of an array thereby making its performance similar to omni directional antennas with little or no interference cancellation capability (depending on λ) [4]. In the case of antenna systems the inter antenna spacing is usually very large compared to inter antenna element spacing; typically 3λ to 30λ. This is to ensure that the multiple copies of the signal with independent fading models arrive at the base station thereby increasing the probability of receiving at least one signal with minimum distortion due to the

presence of fading and scatterers. It is very unlikely that it will be possible to have more than two antennas on a mobile while it may be easier, depending on the location, to have several antennas at a base station. Finally, there is also a tradeoff between cost and computational complexity associated with using multiple antennas and antenna arrays at the base station. Implementation of large number of antennas at the base station requires more space and is thus expensive and using large number of antenna elements requires more computation but less space.

3.5 Conclusions

A combination of variable receive antennas in a STBC scheme and a minimum variance distortionless response beamforming algorithm with variable number of antenna array elements is used to investigate the relative merits and demerits of antennas and antenna array elements in an uplink environment. The objective of the work presented was to achieve an understanding of the role of a variable number of receive antennas and antenna arrays at the PHY layer in reducing the effect of fading, noise and interference. Simulation results showed that in a high noise and high fading environment, diversity achieved by using space-time block coding was more efficient in reducing BER compared to beamforming, while in the presence of high interference environment beamforming performed efficiently in terms of reducing BER. In addition, it was investigated how the role of different number of receive antennas and antenna arrays is affected in varying noise and fading environments and by the number of interferring users and their angular locations. For practical applications, the number of antennas/antenna array elements that can be deployed at the PHY layer is more likely to be determined by the balance between the improvement in capacity (or reduction in BER) and other implementation issues including cost, available space and computational complexity.

Chapter 4

A new cross-layer based multiuser scheduling algorithm

4.1 Introduction

The increased demand on wireless systems to support high data rates and multiple users has led to new research that focuses on the combined issues of the physical (PHY) and the medium access control (MAC) layers in a cross-layer design for wireless networks. Allocating resources at the PHY layer or designing algorithms and protocols at the higher layers by taking PHY layer parameters into consideration in order to satisfy the quality of service (QoS) requirements for both real and non-real time services are some of the challenges in designing cross-layer based systems. An important research question is what changes should be made at these two layers such that the performance objectives at both these layers are jointly optimized. Laroia et al. [6] suggest that one of the most important design components that can result into improved QoS for different kinds of traffic is the scheduler that acts as a focal point between PHY and higher layers and leads to cross-layer optimization. The main job of the scheduler is to prioritize or select users to be served in a way such that the system throughput is maximized.

A number of scheduling policies have been proposed that use the PHY layer in-

formation [38]-[43] in making scheduling decisions. Most of these algorithms schedule users based on their channel state information (CSI) using either the greedy scheduler [34] in which user(s) associated with the best channel conditions are scheduled in a given time slot thus exploiting multiuser diversity gain or the proportionally fair scheduler (PFS) [37] which makes use of the CSI but also provides fair allocation of bandwidth across users. Both these algorithms are currently used in time division multiple access (TDMA) and code division multiple access (CDMA) systems. When using the greedy algorithm, TDMA systems that serve only one user in a given time slot select the user associated with the best channel conditions which are measured in terms of instantaneous signal to noise ratio (SNR). To improve the capacity of TDMA systems by enabling them to serve multiple users in a given time slot, a combined TDMA/SDMA (spatial division multiple access) approach is followed in some studies [72]. These studies combine beamforming capability with TDMA systems to serve multiple users in a given time slot using the designated beams. One of the first standards that is based on combined SDMA using beamforming with TDMA (SDMA/TDMA) based technique is the wireless metropolitan area network IEEE 802.16 standard (also referred to as the WiMAX (Worldwide Interoperability for Microwave Access)) [73]. However, when multiple users are served following the greedy approach in combined TDMA/SDMA systems, the simultaneously served users cause significant interference on one another when their angular locations are close to one another. Examples of implementation of greedy algorithm include high data rate versions of CDMA 2000 and enhanced general packet radio service in the EDGE extension of GSM [19] based on the TDMA technology. When using the proportional fair approach, instead of using instantaneous SNR values, the TDMA or CDMA systems, select the users with the highest value of R/T (where R is the current data rate requested by a user and T is its average throughput over the past t time slots). The PFS is currently used in the emerging 3G data standards such as

CDMA2000 1xEV-DO (evolution data optimized) and high speed downlink packet access (HSDPA) [74].

Some of the existing system models also use multiple antenna systems at the PHY layer and simple scheduling algorithms such as the round-robin (RR) (which schedules users randomly), the greedy or the proportional fair at the higher layers to serve multiple users. Aktas et al. [39], for example, proposed one such scheduling algorithm for MIMO systems. In their downlink model the base station is equipped with n_T transmit antennas while users have n_R receive antennas. It is assumed that the users or the base station have perfect channel knowledge characterized by signal to interference and noise ratio (SINR) and this information is fed back to the base station. The base station then assigns the number of transmit antennas based on this information. Another algorithm supporting multiple users over multiple antennas is proposed by Shin et al. [75] in which the mobile user sends signal to interference plus noise ratio (SINR) information to the base station. The base station uses this information to group the users that are served using the round-robin (RR) scheduling algorithm. Kogiantis and Ozarow [76] also aimed at serving two users simultaneously by maximizing the sum capacity on the basis of their instantaneous channel gains. The effect of the interference created by simultaneously served users on one another, however, was not explicitly considered in these models. While the use of multiple antennas at the PHY layer can help in combating noise and fading, and the use of antenna arrays can reduce the effect of interference at the PHY layer, the selection of users plays a vital role in enhancing system performance beyond the use of antennas and antenna arrays. Careful selection of users based on their SNR combined with an attempt to reduce interference results in increased system throughput. Consideration of the CSI alone does not provide information about the interference that simultaneously served users may cause on each other. For example, if users selected to be served simultaneously are close to one another in an angular

sense then the antenna arrays used at the PHY layer cannot minimize the interference between simultaneously served users in an effective manner.

In this chapter, a new scheduling algorithm named as the combined SNR and angular separation based algorithm is presented that serves two users simultaneously and attempts to minimize the interference between them. The proposed scheduler takes the channel conditions into consideration while scheduling the users as well as their angular location around the base station based on their direction of arrival (DoA). This information is then used to schedule users at the MAC layer in an attempt to minimize the interference. Channel state [69] and DoA [70] information are readily obtained from pilot signals in an uplink environment. The pilot signal based DoA [70] estimation technique donot increase the computational load extensively in the proposed scheduling algorithm compared to other techniques such as MUSIC and ESPIRIT. Moreover the technique proposed in [70] is fast and efficient and the number of signals whose DoA can be estimated is not constrained by the number of antenna elements and thus the angular separation between the available users can be computed easily. The proposed algorithm can also be used within the framework of the TDMA and the CDMA systems.

This chapter is organized as follows. The system model used to study the performance of the proposed algorithm is presented in section 4.2. This system model is used to compare the performance of the proposed algorithm with the traditional greedy and RR algorithms both using numerical simulations and the semi-analytical framework. The semi-analytical framework uses analytical formulae, of probability density and cumulative distribution functions, as well as Monte-Carlo simulations. The detailed semi-analytical framework describing the system model is presented in section 4.3. The numerical framework of this system model is presented in section 4.4. Results from both the numerical simulations and semi-analytical analysis are compared and discussed in section 4.5 in terms of the system capacity obtained using the

proposed and traditional scheduling algorithms. Finally, conclusions are presented in section 4.6.

4.2 The system model

A multiuser downlink model (as shown in Fig. 4.1) is considered with a base station capable of serving two users simultaneously out of K users in a given time slot. The remaining $K-2$ users do not receive service. The model assumes that each mobile has a single antenna and the base station is equipped with several antennas that are capable of beamforming. Beamforming is used at the base station to steer the beams towards the desired users in order to reduce the interference that simultaneously transmitted data for the two mobile users may cause on each other. The mobile users are assumed to be randomly distributed between $-\pi/2$ and $+\pi/2$ around the base station. It is assumed that the information about the instantaneous SNR, γ_i, and the spatial distribution of the mobile users is available at the base station.

The users are selected for service based on the scheduling algorithms considered at the MAC layer.

For the analysis presented here, three scheduling policies are considered:

- The round-robin scheduling algorithm [75]: In the round-robin scheduling algorithm users are served in the sequential order of their arrival and thus randomly with respect to their SNR [75]. This scheme ensures fairness by treating users with both good (high SNR) and bad (low SNR) channels equally.

- The greedy scheduling algorithm: The greedy algorithm [38] prioritizes users according to the channel conditions characterized by their instantaneous signal to noise ratio (γ_i) and selects users m and n out of K available users as

$$m = \arg \max_{i=1,2,\ldots,K} [\gamma_i] \tag{4.1}$$

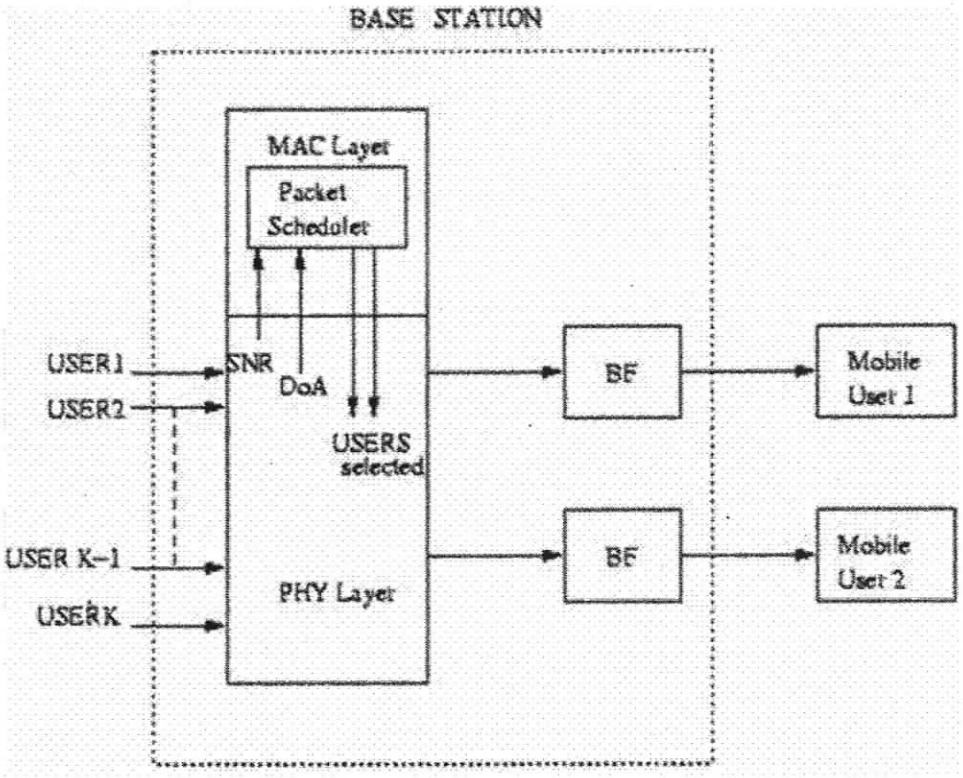


Figure 4.1: Block diagram of multiuser downlink model consisting of single base station and K mobile users

$$n = \arg \max_{i=1,2,\ldots,K, i \neq m} [\gamma_i] \tag{4.2}$$

which implies that users m and n associated with the highest and second highest instantaneous SNR are selected. This approach attempts to maximize the signal to noise ratio but tends to be unfair to the users with poor channel conditions.

- The combined SNR and angular separation based scheduling algorithm: The combined SNR and angular separation based scheduling algorithm is designed to serve at least two users simultaneously. The first user m is selected on the basis of its instantaneous signal to noise ratio as in the greedy algorithm (Eqn. (4.1)). The second user n is selected such that it is farthest away from user m in an angular sense

$$n = \arg \max_{i=1,2,\ldots,K; i \neq m} [\theta_{mn}] \tag{4.3}$$

where θ_{mn} is the angular separation between mobile users m and n. Here, the selection of user m attempts to maximize the SNR while the selection of user n attempts to reduce interference. Since the selection of user n is not based on γ_i, the scheme tends to be more fair than the greedy algorithm as the probability of users with poor channel conditions being served is higher than that for the greedy algorithm.

If s_1 and s_2 are the symbols transmitted by the base station for mobile users m (the first user) and n (the second user), respectively, then the symbols r_1 and r_2 received at mobile m and n are written as

$$r_1 = h_1 s_1 + n_1 + \Lambda_2 \tag{4.4}$$

$$r_2 = h_2 s_2 + n_2 + \Lambda_1 \tag{4.5}$$

where h_1 and h_2 are the channel responses, n_1 and n_2 the additive white Gaussian noise. Λ_2 represent the interference caused on the first mobile user by the signal destined for the second mobile user and similarly Λ_1 is the interference caused on the second mobile user by the signal destined for the first mobile user. The interference caused by simultaneously served mobiles on each other is calculated on the basis of leakage of power that depends on the angular separation between the mobiles and the beamforming approach used. The performance of the scheduling algorithms is assessed in terms of the system capacity calculated semi-analytically as well as numerically as discussed in the sections 4.3 and 4.4 respectively.

4.3 Semi-analytical model

The semi-analytical model used to calculate the capacity of the system is based on probability density functions (PDFs) of the SNR of the first and second users which in turn depends on the scheduling policy considered and the fading environment. The PDF of interference is obtained by taking into account the beamforming approach and the PDF of angular separation between mobile users. Some known preliminaries that form the basis for the semi-analytical model are presented below followed by the description of the semi-analytic system model capable of serving two users simultaneously.

4.3.1 Preliminaries

The SNR for unity transmit power and for a single user case is expressed as [67]

$$\gamma = \frac{|h|^2}{\sigma_n^2} \tag{4.6}$$

where h is the channel amplitude response whose distribution represents some fading environment and σ_n^2 is the Gaussian noise power. For example, if the channels are assumed to be Nakagami faded which represents the general fading model with

parameter d, then the PDF of the instantaneous received SNR γ is expressed as [67]

$$p_\gamma(\gamma) = \left(\frac{d}{\overline{\gamma}}\right)^d \frac{\gamma^{d-1}}{G(d)} \exp\left(\frac{-d\gamma}{\overline{\gamma}}\right) \tag{4.7}$$

where $\overline{\gamma}$ is the average SNR, $G(x) = \int_0^\infty e^{-t}t^{x-1}dt$ is the gamma function and d is the measure of fading intensity with d =1 yielding the Rayleigh faded environment and $d > 1$ and $d < 1$, respectively, corresponding to channel conditions better and worse than the Rayleigh model. Assuming h here to be Rayleigh distributed ($d = 1$), then the PDF of instantaneous SNR γ of each user is exponentially distributed [67] and given by

$$p_\gamma(\gamma) = \frac{1}{\overline{\gamma}} \exp\left(\frac{-\gamma}{\overline{\gamma}}\right) \tag{4.8}$$

and the corresponding cumulative distribution function (CDF) is

$$P_\gamma(\gamma) = 1 - \exp\left(\frac{-\gamma}{\overline{\gamma}}\right) \tag{4.9}$$

Consider the case of finding the system capacity for a single user case ($M = 1$) given a total of K users, using the round-robin and greedy scheduling algorithms. For M =1, no interference from simultaneously served users is present and the system capacity depends on the SNR, γ_1 instead of the SINR, Γ_1. For the round-robin scheduling algorithm, the PDF of instantaneous γ_1 is exponentially distributed according to Eqn. (4.8) since the users are selected without any special criteria and the system capacity for a single user case is given by [67]

$$C_{RR} = c_{RR_1} = \int_0^\infty log_2(1+\gamma)p_\gamma(\gamma)d\gamma = \frac{e^{\frac{1}{\overline{\gamma}}}}{\ln 2}\mathrm{Ei}\left(\frac{1}{\overline{\gamma}}\right) \tag{4.10}$$

where $\mathrm{Ei}(\cdot)$ is an exponential integral function [77]

$$\mathrm{Ei(x)} = \int_x^\infty t^{-1}e^{-t}dt \tag{4.11}$$

When a user is selected for service based on the greedy scheduling algorithm the PDF of the instantaneous SNR for the l^{th} best user out of K users is given by [78] (Eqn. (2.1.6))

$$p_{\gamma_{l:k}}(\gamma) = \frac{K!}{(K-l)!(l-1)!}[P_\gamma(\gamma)]^{K-l}[1-P_\gamma(\gamma)]^{l-1}p_\gamma(\gamma) \tag{4.12}$$

where $P_\gamma(\gamma)$ and $p_\gamma(\gamma)$ are the CDF and the PDF of instantaneous SNR (γ), respectively. Substituting expressions for $P_\gamma(\gamma)$ and $p_\gamma(\gamma)$ in Eqn. (4.12) and simplifying yields the PDF of instantaneous SNR of the best user (l=1) out of K users as

$$p_{\gamma_{1:k}}(\gamma) = \frac{K!}{(K-1)!\overline{\gamma}}[1-e^{-\gamma/\overline{\gamma}}]^{K-1}e^{-\gamma/\overline{\gamma}} \tag{4.13}$$

The system capacity for the single best user in the greedy algorithm is thus

$$C_G = C_{G_1} = \int_0^\infty log_2(1+\gamma)p_{\gamma_{1:K}}(\gamma)d\gamma \tag{4.14}$$

As the number of users approaches 1 ($K \rightarrow 1$), then $p_{\gamma_{1:K}}(\gamma) \rightarrow p_\gamma(\gamma)$, and the greedy scheduling algorithm leads to the same system capacity as the RR scheduling algorithm since the benefit of choosing the user with highest SNR gradually disappears.

In a multiuser environment with M ($M \geq 2$) simultaneously served users out of a total of K users, each user served experiences fading (characterized by the distribution of h_i) as well as interference from other simultaneously served users. The SINR of user i (Γ_i) for unity transmit signal power is expressed as

$$\Gamma_i = \frac{|h_i|^2}{\sigma_n^2 + \sum_{j=1,j\neq i}^{M}|h_j|^2 I_{ij}} \tag{4.15}$$

where I_{ij} ($=I_{ji}$) is a scalar that varies between 0 and 1 and represents the interference

caused by user j on user i. I_{ij} depends on the beamforming approach used and the angular separation between the users. If simultaneously served users are sufficiently far apart, they do not cause interference on each other and then $I_{ij} = 0$ for all j ($j \neq i$). In contrast, when the beams of the two users overlie directly over each other (e.g., when both users have the same angular location around the base station) then I_{ij} is 1. The commonly used array factor $F(\theta, \theta_0)$ that defines the nature of the beam for a uniformly spaced linear antenna array with L antenna elements [2] can be used to obtain I as

$$I(\theta) = F(\theta, \theta_0) = \mathbf{V^T u} \tag{4.16}$$

where $\mathbf{V^T} = [V_0, V_1, \ldots, V_{L-1}]$ is the weighting vector and the weights V_k are given by

$$V_k = A_k e^{-jk\kappa d cos\theta_0}, \;\; k = 0, 1, \ldots, L-1 \tag{4.17}$$

θ_0 is the direction in which the beam is steered (assumed equal to 0° here), κ=$2\pi/\lambda$, λ is the wavelength, d is the inter-element distance usually taken as $\lambda/2$ and A_k is the uniform amplitude equal to one for all k and $\mathbf{u}$ is the array propagation vector given by

$$\mathbf{u^T} = [1 e^{j\kappa d \cos\theta} \cdots e^{j\kappa d(L-1)\cos\theta}] \tag{4.18}$$

The array factor is thus written as [2]

$$F(\theta, \theta_0) = \sum_{k=0}^{L-1} A_k e^{j[k\kappa d(\cos\theta - \cos\theta_0)]} \tag{4.19}$$

and yields a beam pattern shown in Fig. 4.2 (both as a loss term in dB and as a ratio of power emitted in the desired direction for number of antenna elements (L) equal

to 7 and 100). Fig. 4.2 shows that there is no loss in direction θ_0 where the beam is steered (loss = 0 dB, ratio = 1) but some energy is dissipated in other directions. The beam pattern for 100 antenna elements show how a large number of antenna elements yields smaller beam width and less energy dissipation in other directions. For mathematical tractability we approximate this pattern as

$$I(\theta) = e^{-|\theta|/c} \tag{4.20}$$

where θ is the angular separation in degrees and c is a parameter that is surrogate for, and inversely proportional to, the number of antenna elements. The higher (respectively lower) the value of c, the wider (respectively narrower) the beam, and more (respectively less) energy is dissipated in other directions. Fig. 4.3 shows how Eqn. (4.20) approximates the typical beam pattern for $L = 4$. The value of $c = 73$ is obtained by minimizing the sum of the square of the differences between the original beam pattern and that obtained by Eqn. (4.20) and yields a correlation coefficient of 0.93. The beam pattern allows us to determine the leakage of power in the direction of simultaneously served co-channel users based on the angular separation between them. With the exponential beam pattern defined by Eqn. (4.20), the larger the angular separation between the users, the lower is the leakage and smaller the interference simultaneously served users cause on each another. Note that while a large number of antenna elements can be used to obtain a narrow beam that causes little interference, there is a practical limit to the number of antenna elements that can be used due to the large computational expense associated.

The performance of the multiuser system is evaluated in terms of the total system capacity (C) expressed as the sum of the capacities c_i of the individual users [67] as

$$C = \sum_{i=1}^{M} c_i \tag{4.21}$$

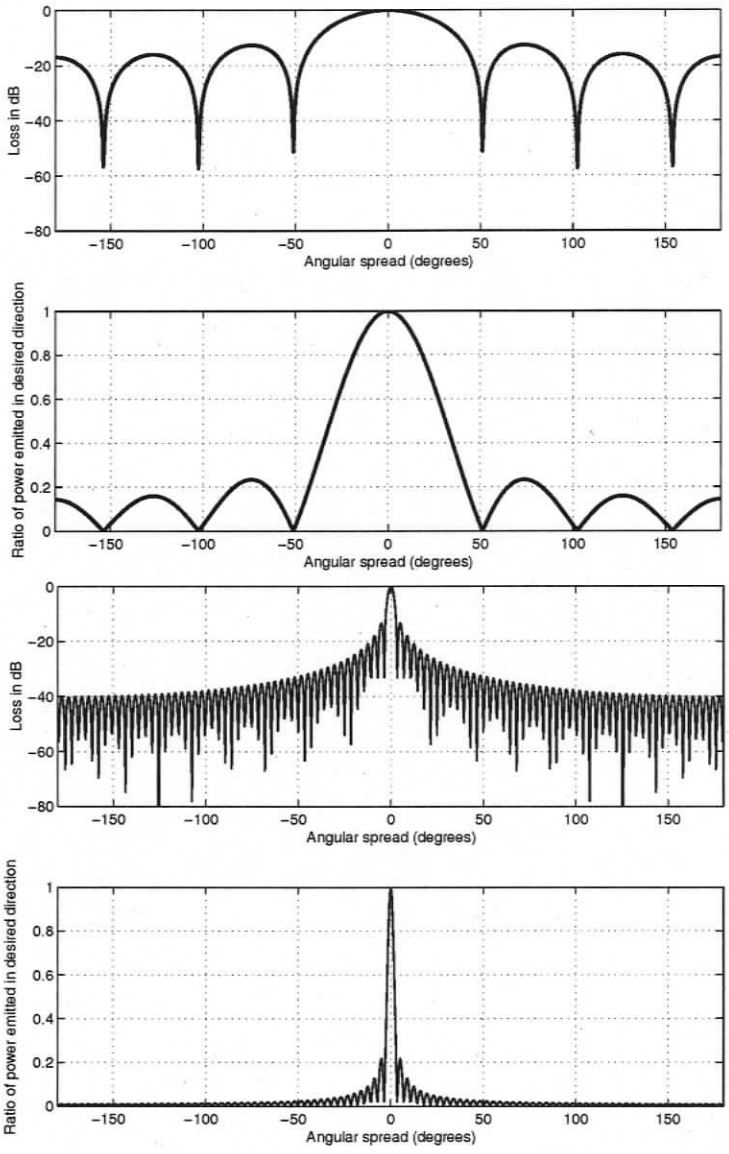


Figure 4.2: Beam pattern for antenna elements equal to 7 (top two panels) and 100 (bottom two panels).

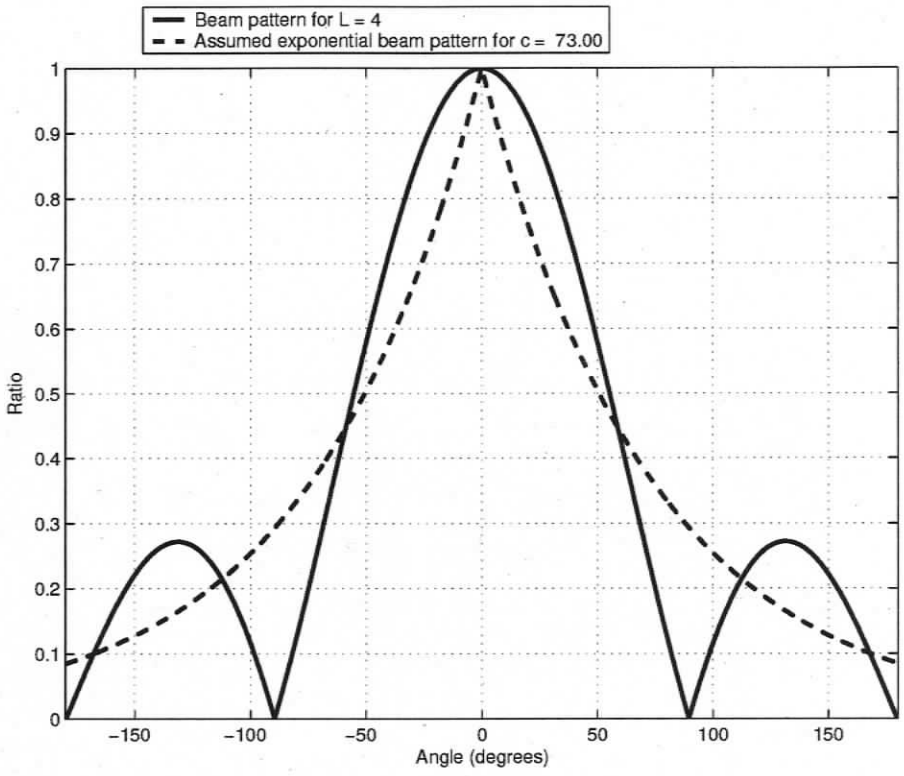


Figure 4.3: Approximation of beam pattern as a function of θ via Eqn. (4.20). Number of antenna elements $L = 4$ and parameter c in Eqn. (4.20) equals 73.

where

$$c_i = E[log_2(1+\Gamma_i)] = \int_0^\infty \log_2(1+\Gamma_i) p_\Gamma(\Gamma_i) d\Gamma_i \tag{4.22}$$

Γ_i is the instantaneous SINR and $p_\Gamma(\Gamma_i)$ is the PDF of Γ_i.

4.3.2 Capacity analysis for two simultaneously served users

When two users out of a total of K users are simultaneously served then based on Eqn. (4.15), for $M = 2$, the SINR of the first user that experiences interference from the second user is written as

$$\Gamma_1 = \frac{|h_1|^2}{\sigma_n^2 + |h_2|^2 I} = \frac{\frac{|h_1|^2}{\sigma_n^2}}{1 + \frac{|h_2|^2 I}{\sigma_n^2}} = \frac{\gamma_1}{1+\gamma_2 I} \tag{4.23}$$

where $\frac{|h_1|^2}{\sigma_n^2}$ and $\frac{|h_2|^2}{\sigma_n^2}$ are the SNRs of users 1 and 2, respectively. Γ_2, the SINR for user 2 that experiences interference from user 1 is similarly written as

$$\Gamma_2 = \frac{\gamma_2}{1+\gamma_1 I} \tag{4.24}$$

Estimation of the system capacity in the presence of interference requires the PDF of SINR, $p_{\Gamma_i}(\Gamma_i)$, for both users. The PDFs of $\frac{|h_1|^2}{\sigma_n^2}$ and $\frac{|h_2|^2}{\sigma_n^2}$ depend on the choice of the scheduling algorithm used. For example, these SNRs are exponentially distributed for the RR scheduling algorithms and distributed according to Eqn. (4.12) with appropriate values of K and l for the greedy algorithm. The PDF of the SNR of the first user for the combined SNR and angular separation based algorithm is based on Eqn. (4.12) with $l = 1$ as for the greedy algorithm. The second user, however, has an exponentially distributed SNR because it is selected on the basis of its angular location and not its SNR which are assumed to be uncorrelated to each other. The PDF of I is derived in the next two lemmas based on the angular distribution of users around the base station. The first lemma presented below derives the PDF of I, $p_I(I)$, for the case of round-robin and greedy algorithms that choose users irrespective of

their angular location. The second lemma derives the PDF of I, $p_{I_{AA}}(I_{AA})$ for the combined SNR and angular separation based algorithm proposed here that selects the second user based on its angular separation from the first user.

Lemma 1: *Given randomly distributed mobile users around a base station between* $-\pi/2$ *and* $+\pi/2$ *then the PDF of* I, $p_I(I)$, *is given as*

$$p_I(I) = \frac{2c^2 \ln I + 2\pi c}{\pi^2 I} \tag{4.25}$$

Proof: Based on [79], when two users are randomly distributed around the base station, then the angular separation (θ) in radians between them is a random variable varying between 0 and π with a triangular distribution given by

$$p_\theta(\theta) = \frac{2}{\pi} - \frac{2\theta}{\pi^2},\ 0 \leq \theta \leq \pi \tag{4.26}$$

Given $p_\theta(\theta)$ and the expression for I as a function of θ, the PDF of I, $p_I(I)$ is obtained as [79]

$$p_I(I) = \frac{p(\theta)}{-\frac{dI}{d\theta}} = \frac{\frac{2}{\pi} - \frac{2\theta}{\pi^2}}{\frac{e^{-\theta/c}}{c}} \tag{4.27}$$

Substituting $\theta = -c\ \ln(I)$ in the above equation yields

$$p_I(I) = \frac{2c^2 \ln I + 2\pi c}{\pi^2 I} \tag{4.28}$$

Lemma 2: *Given randomly distributed mobile users around a base station between* $-\pi/2$ *and* $+\pi/2$ *then* $p_{I_{AA}}(I_{AA})$, *the PDF of* I_{AA} *for the combined SNR and angular separation based algorithm is given by*

$$p_{I_{AA}}(I_{AA}) = \frac{Nc}{I}\left(\frac{-2c\ln I}{\pi} - \frac{(-2c\ln I)^2}{\pi^2}\right)^{N-1}\left(\frac{2}{\pi} + \frac{2c\ln I}{\pi^2}\right) \tag{4.29}$$

Proof: The PDF of the 1^{st} largest angular separation $\theta_{1:N}$ out of total of $N =$

$K-1$ angular separations between K users is obtained following [78] as

$$p_{\theta_{1:N}}(\theta) = \frac{N!}{(N-1)!}[P_\theta(\theta)]^{N-1} p_\theta(\theta) \tag{4.30}$$

where $p_\theta(\theta)$, the PDF of angular separation θ is given by Eqn. (4.26). The CDF $P_\theta(\theta)$ is obtained by integrating $p_\theta(\theta)$ as

$$P_\theta(\theta) = \frac{2\theta}{\pi} - \frac{\theta^2}{\pi^2}, \quad 0 \leq \theta \leq \pi \tag{4.31}$$

Substitution of the expressions for $p_\theta(\theta)$ and $P_\theta(\theta)$ in Eqn. (4.30) and simplification yields

$$p_{\theta_{1:N}}(\theta) = N\left(\frac{2\theta}{\pi} - \frac{\theta^2}{\pi^2}\right)^{N-1}\left(\frac{2}{\pi} - \frac{2\theta}{\pi^2}\right) \tag{4.32}$$

that is used to obtain $p_{I_{AA}}(I_{AA})$ as

$$p_{I_{AA}}(I_{AA}) = \frac{p_{\theta_{1:N}}(\theta)}{-\frac{dI}{d\theta}} = N\frac{(\frac{2\theta}{\pi} - \frac{\theta^2}{\pi^2})^{N-1}(\frac{2}{\pi} - \frac{2\theta}{\pi^2})}{\frac{e^{-\theta/c}}{c}} \tag{4.33}$$

Substituting $\theta = -c \ln(I)$ in Eqn. (4.33), yields Eqn. (4.29).

As expected, when N approaches 1 (i.e., when there are only 2 users, i.e., $K = 2$), $p_{I_{AA}}(I_{AA}) \rightarrow p_I(I)$ since there are no more users to choose the farthest one from. Fig. 4.4 compares $p_\theta(\theta)$ and $p_{\theta_{1:5}}(\theta)$ and shows how the selection of the user that is farthest away from the first user results in a shift of PDF of θ towards the right.

It also follows from the above analysis that all scheduling algorithms yield similar results when two users are simultaneously served from available two users. In this case the greedy algorithm does not have the advantage of selecting the best users and the combined SNR and angular separation based algorithm does not have the advantage of choosing the user that is farthest from the user with highest SNR.

Similar to Eqn. (4.23), the SINR of the first and second user for the combined SNR and angular separation based algorithm is written as

Scheduling algorithm	PDF of γ_1	PDF of γ_2	PDF of I
Round-robin	exponential, $p_{\gamma_1}(\gamma)$	exponential, $p_{\gamma_2}(\gamma)$	Eqn. (4.25), $p_I(I)$
Greedy	Eqn. (4.12), $p_{\gamma_{1:K}}(\gamma)$	Eqn. (4.12), $p_{\gamma_{2:K}}(\gamma)$	Eqn. (4.25), $p_I(I)$
Combined SNR and angular separation	Eqn. (4.12), $p_{\gamma_{1:K}}(\gamma)$	exponential, $p_{\gamma_2}(\gamma)$	Eqn. (4.29), $p_{I_{AA}}(I)$

Table 4.1: Distributions of γ_1, γ_2 and I used for calculating Γ_1 and Γ_2 for the greedy, RR and the combined SNR and angular separation based algorithms.

$$\Gamma_1 = \frac{\gamma_1}{1+\gamma_2 I_{AA}} \tag{4.34}$$

$$\Gamma_2 = \frac{\gamma_2}{1+\gamma_1 I_{AA}} \tag{4.35}$$

where γ_1 and γ_2 are the SNRs of the two simultaneously served users and I_{AA} is the interference that is now distributed according to Eqn. (4.29).

The performance of the three scheduling algorithms can now be compared in terms of the PDFs of the SINR of the two simultaneously served users, i.e., $p_\Gamma(\Gamma_i)$ for $i = 1, 2$. The approach used here is based on obtaining $p_\Gamma(\Gamma_i)$ using randomly generated values for γ_i and I based on their known distributions summarized in Table 4.1 for the three scheduling algorithms. This table is based on the results obtained in the previous two lemmas and assuming that the channels h_1 and h_2 and the angular location of mobile users are independent and uncorrelated. Using randomly generated values for γ_1, γ_2 and I for these distributions, a large number of calculated SINRs are used to estimate the PDF of Γ_i for the three scheduling algorithms.

Fig. 4.5a compares $p_\Gamma(\Gamma_1)$, the PDF of SINR of user 1 for the RR, greedy and combined SNR and angular separation based algorithm. Compared to other two algorithms, $p_\Gamma(\Gamma_1)$ is shifted to the right for the combined SNR and angular separation based algorithm which implies higher mean values of Γ_1 and subsequently higher

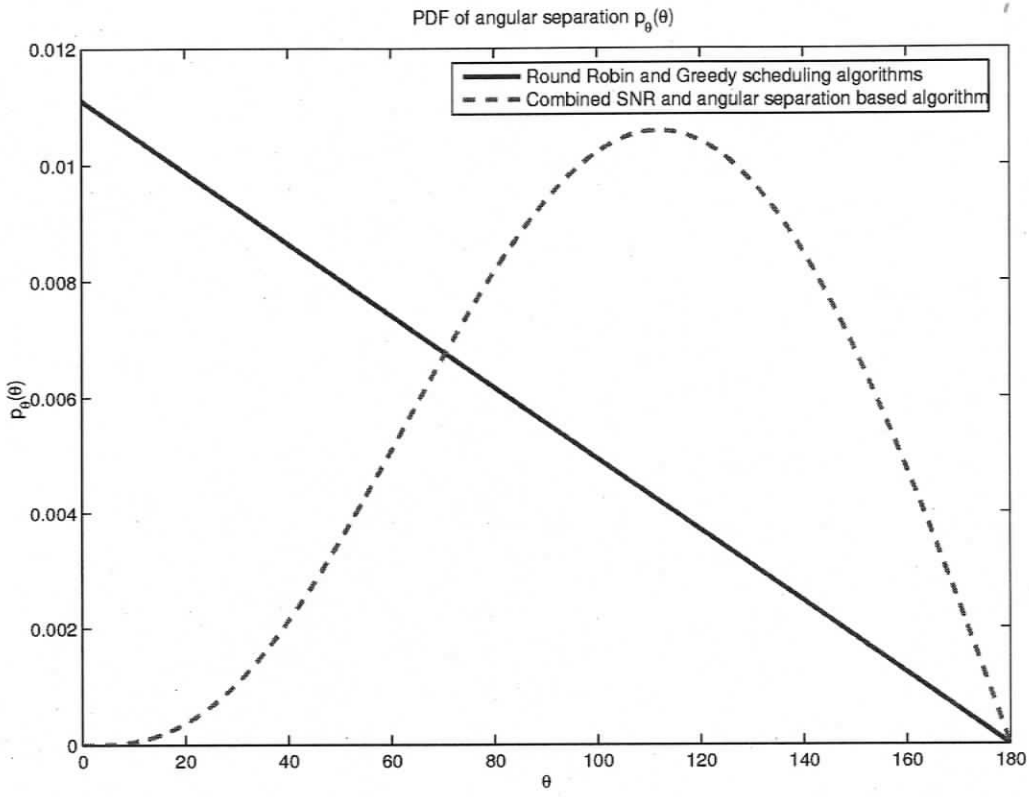


Figure 4.4: PDFs of the angular separation θ for the RR and greedy algorithms based on Eqn. (4.26) and for the combined SNR and angular separation based algorithm (based on Eqn. (4.32)). Parameter N is assumed to 5 in Eqn. (4.32).

system capacity. The reason for this is that the combined SNR and angular separation based algorithm chooses the first user with highest SNR and the second user that is farthest away from the first user so that the simultaneously served users experience less interference from each other. For user 1, the greedy algorithm performs better then the RR algorithm because it chooses a user associated with the highest value of SNR. The $p_\Gamma(\Gamma_1)$ for the RR algorithm lies on the extreme left since it chooses users randomly without any preference and does not explicitly attempt to reduce interference.

Fig. 4.5b compares $p_\Gamma(\Gamma_2)$, the PDF of SINR of the user 2, for the three scheduling algorithms. It shows that $p_\Gamma(\Gamma_2)$ is similar for all the three algorithms. The second user for the RR algorithm has low average SNR but it also experiences low average interference from another randomly chosen user. The second best user in the greedy algorithm has higher average SNR than the RR user, but it also experiences higher interference from the first greedy user. Finally, the second user in the combined SNR and angular separation based algorithm is essentially an average RR user and although it experiences interference from the user with the highest SNR this interference is reduced because of larger angular separation. The net result is that the PDFs of SINR of the second user are fairly similar for all the scheduling algorithms.

4.4 Numerical model

Numerical simulations were also carried out for a multiuser downlink model capable of serving two users simultaneously using the three scheduling algorithms discussed in the previous sections. For comparison with the semi-analytical approach the beam pattern is assumed to be exponential (Eqn. (4.20)). In the numerical model data to be transmitted is first modulated using the binary phase shift keying (BPSK) modulation. The modulated data (generated as binary symbols) are transmitted to the receivers via channel $he^{j\theta}$ where h is assumed to be Rayleigh distributed and θ is

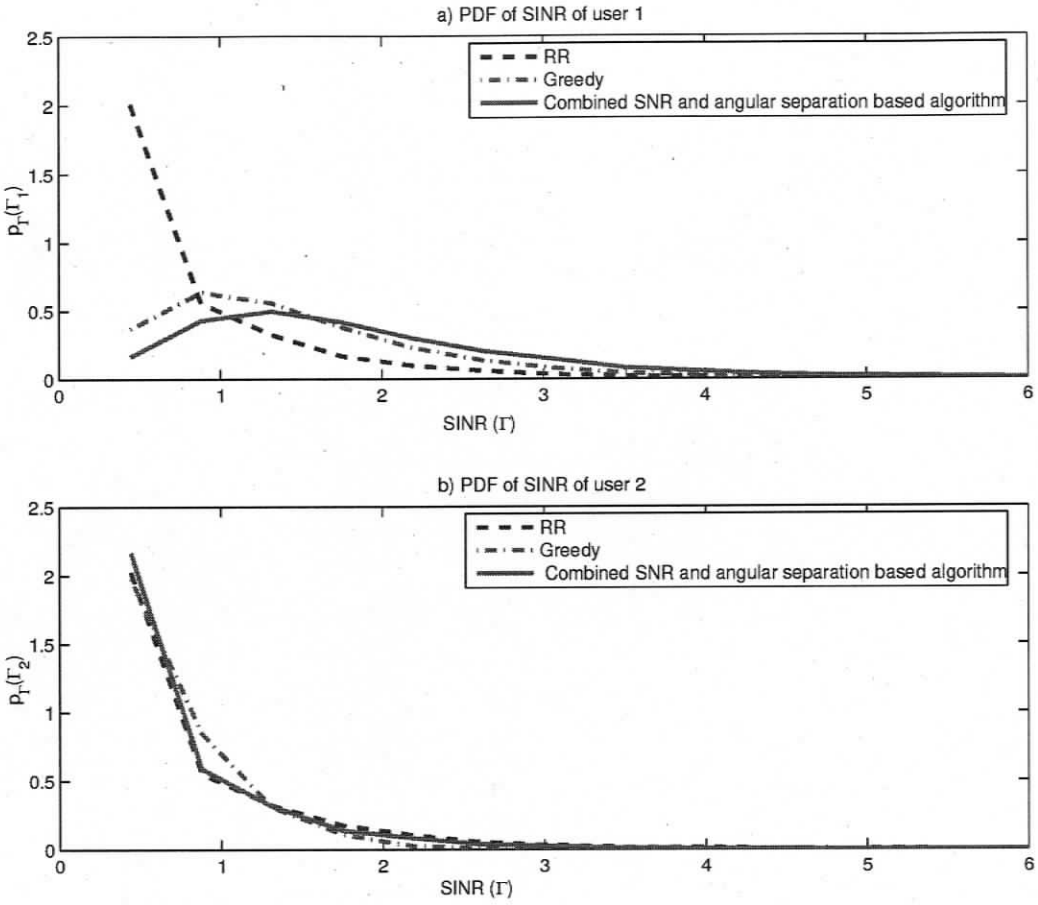


Figure 4.5: PDF of SINR for 1^{st} (panel a) and 2^{nd} (panel b) users out of 5 users ($K = 5$) for the RR, greedy and combined SNR and angular separation based algorithms.

the phase shift which is randomly distributed between 0 and π. Users are selected for service from K users according to the scheduling algorithms which in turn affects the distribution of SNR of the simultaneously served users. Simultaneously transmitted symbols from the base station to the mobile users suffer from interference from each other and the magnitude of this interference depends on the angular separation between the users. At the receiver, an inverse channel operation is performed followed by demodulation to retrieve the original symbols. Finally, the transmitted and received symbols are compared to calculate the BER.

The performance of the scheduling algorithms is assessed in terms of their system capacity that is calculated on the basis of simulated BER. First an average SINR ($\overline{\Gamma}$) is calculated for the two simultaneously served users by inverting the following equation [36]

$$BER = \alpha Q(\sqrt{\beta\Gamma}) \tag{4.36}$$

where α and β are constants that depend on the signal constellation and Q is the Q-function defined as $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\mu^2/2}d\mu$ [36] (pg. 593). The SINR values are then used to obtain their PDFs for each scheduling algorithm. The PDFs of SINR of each scheduling algorithm from both numerical and semi-analytical model are then numerically integrated to find the total system capacity as

$$C = \sum_{j=1}^{N} \log_2(1+\Gamma_j)p_\Gamma(\Gamma_j) \tag{4.37}$$

where N is the number of the bins into which the PDF of Γ is divided. For both semi-analytical and numerical analysis methods a value of $c = 73$ (that corresponds to number of antenna elements equal to 4) is used in the exponential beam (Eqn. 4.20). Also the SNR in the numerical model is set to 0 dB corresponding to $\overline{\gamma}$ =1 in

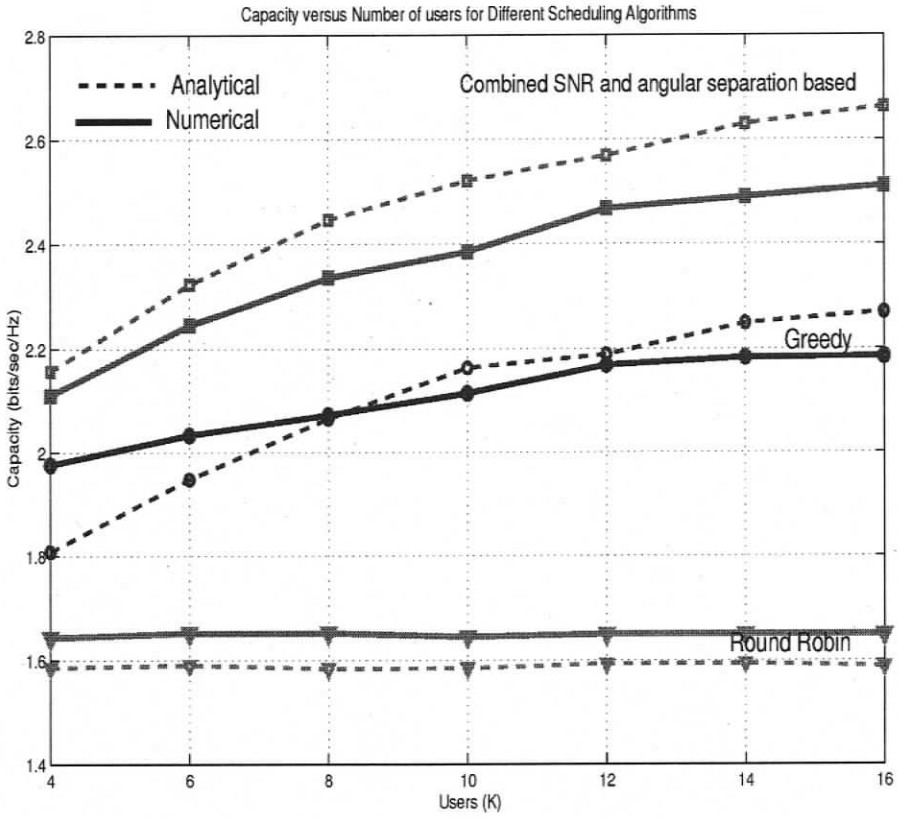


Figure 4.6: Capacity of two out of K users obtained analytically (Ana) and numerically (Num) for the Greedy, RR and Combined SNR and angular separation based algorithms.

the semi-analytical approach.

The next section compares the performance of the three scheduling algorithms in terms of capacity using results from both the semi-analytic and numerical approaches.

4.5 Analysis of simulation results

The performance of the scheduling algorithms is compared for a varying number of total mobile users K. In addition the results from the semi-analytical model and the numerical analysis are also compared to each other.

Results for the case of serving 2 out of K users

Fig. 4.6 compares the results from the numerical and semi-analytical models for

all the three scheduling algorithms for varying K and M =2. Both the numerical and the semi-analytical models exhibits similar behaviour. In both the models the RR scheduling algorithm exhibits constant system capacity because of unpreferential treatment of users and therefore despite an increase in the total number of users the system capacity does not increase. System capacity for the greedy algorithm increases (in both the semi-analytical and numerical models) with an increase in the number of users because of an increase in the probability of obtaining users with higher values of SNRs. The rate of increase in the system capacity with an increasing number of users gradually decreases because γ_1 and γ_2 are exponentially distributed and an increase in the number of users leads to diminishing returns in terms of obtaining users with higher SNRs. System capacity for the combined SNR and angular separation based algorithm is higher then the RR and greedy algorithms in both the semi-analytical and numerical models due to the reduced interference caused by simultaneously served users on one another. Although the combined SNR and angular separation based algorithm selects the second user on the basis of angular separation and not based on the SNR as in the greedy algorithm, it still exhibits overall better performance due to the effect of reduced interference.

The values of system capacities for the numerical and semi-analytical model are slightly different. This is related to the fact that in the numerical model the SINR used to compute system capacity is obtained from the simulated BER that was obtained on the basis of actual packet transmission through the channel experiencing slow Rayleigh fading and in the presence of interference from the simultaneously served users. The relation between BER and SINR used in the numerical model (Eqn. 4.36) is based on the presence of AWGN and thus, is only an approximation of the channel conditions.

The performance of the scheduling algorithms also depends on the width of the beam at the PHY layer, as discussed in [80], that determines the magnitude of the

interference. As expected, smaller beam widths lead to less interference by simultaneously transmitted data for mobile users on each other. Fig. 4.7 compares the results for large and small values of parameter c associated with small and large number of antenna elements at the PHY layer, respectively, from the semi-analytical model. Fig. 4.7 shows that for small values of c $(= 35)$ (corresponding to 8 antenna elements) the system capacity for all three scheduling algorithms is higher compared to large values of c $(= 73)$ (corresponding to 4 antenna elements). This is because a large number of antenna elements results into a narrower beam compared to a small number of antenna elements and thus large number of antenna elements are more effective in reducing the interference. In addition the difference between the greedy and the combined SNR and angular separation based algorithms reduces when there is less interference due to the narrower beam width. Clearly, if the beam width is infinitely small then users will not create interference on each other and there is no advantage in choosing users that are far apart. In this case the greedy algorithm will outperform the RR and the combined SNR and angular separation based algorithm in terms of system capacity. Small beam widths, which can be obtained using a large number of antenna elements are, however, not practically feasible due to their higher computational cost and complexity.

4.6 Conclusions

A multiuser downlink system model for data transmission is considered where two out of K users are served in a given time slot which are chosen on the basis of three scheduling algorithms (RR, greedy and combined SNR and angular separation based algorithms) at the MAC layer. Results from both the numerical and semi-analytical model indicate that taking into consideration both the angular separation between the users and their channel state information yields significant improvement in the system capacity as is the case for the combined SNR and angular separation

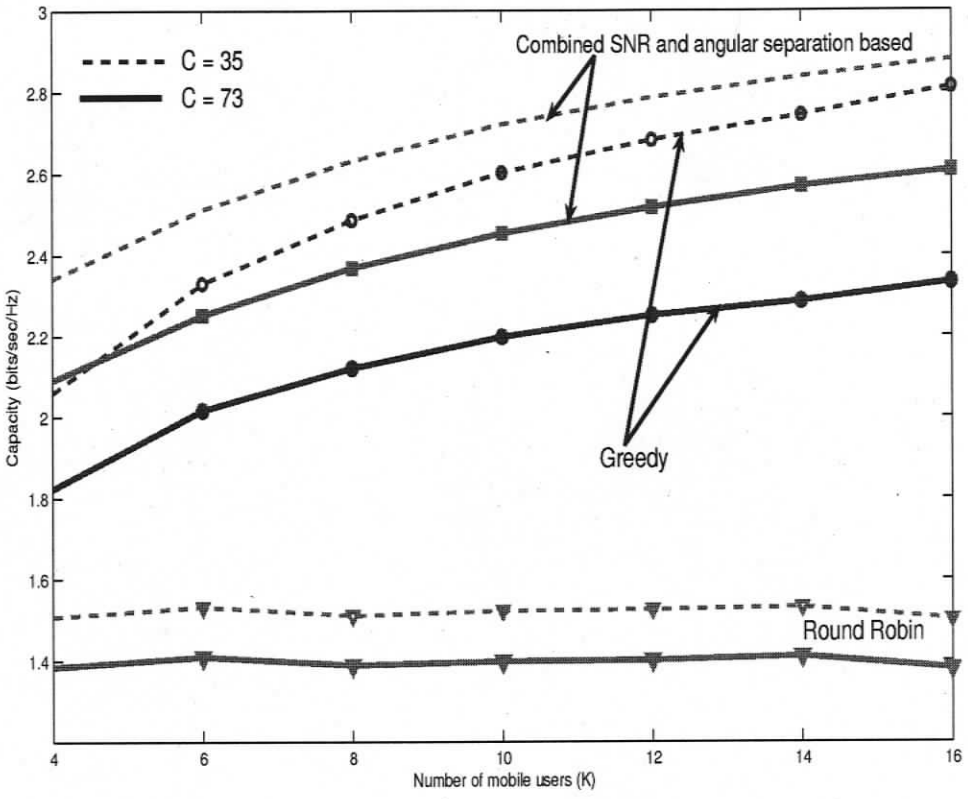


Figure 4.7: Effect of parameter c on the capacity of two out of K users obtained analytically (Ana) for different scheduling algorithms.

based scheduling algorithm that is shown to outperform both the greedy and the RR algorithms. Simulation results from the numerical and the semi-analytical model are also shown to be consistent. The improvement in system capacity obtained by additionally considering angular separation between mobile users, however, requires computation of the DoA of the users. A new computationally efficient DoA estimation technique [70] is now available that can be used in the proposed framework with relatively small additional computational expense.

Chapter 5

Performance evaluation of scheduling algorithms within a queuing framework

5.1 Introduction

In Chapter 4 a new scheduling algorithm, the combined signal to noise ratio (SNR) and angular separation based algorithm, is presented and compared with other traditional algorithms such as the greedy and the round-robin algorithms in terms of system capacity. It is shown both semi-analytically and numerically that the combined SNR and angular separation based scheduling algorithm achieves higher average system capacity when compared to the greedy and the round-robin scheduling algorithms. This improvement in system performance is achieved by taking both the channel conditions in terms of SNR and angular separation of users in terms of direction of arrival (DoA) into consideration while selecting users for service. In Chapter 4 it is assumed that in each time slot, two or more packets are available and the base station always served two packets in each time slot due to the constant availability of channel. However, a channel may not always be available and the channel state can change from good to bad within a short time. A channel is considered to be in a good state and available if it meets a minimum specified SNR threshold requirement which in turn depends on the system data rate that is required to be supported [81]. Hence,

while designing scheduling algorithms aimed at satisfying the quality of service (QoS) requirements, it becomes necessary to consider the effect of channel availability. QoS is usually assessed in terms of signal to interference and noise ratio (SINR) above some minimum threshold and/or delay below some specified acceptable value. Meeting one or both QoS requirements in wireless networks poses design challenges due to the time varying nature of the channels.

Limited bandwidth, time-varying fading channels, and resource competition among multiple users provide QoS challenges for high-speed bursty data traffic over wireless fading channels. Research focussing on improving SINR at the physical (PHY) layer usually concentrates either on reducing the effect of channel variability by using techniques such as multiple antenna systems (MIMO) [11]-[13], or minimizing the effect of interference using techniques such as beamforming [14]-[17] or a combination of both [57]-[63]. Similarly, the research focussing on meeting QoS at the MAC layer mainly focusses on reducing the delay [82].

With the development of cross-layer designs [19], it has been realized that the system performance can be improved significantly by exchange of information between layers and the scheduler, in this context, becomes a focal point to achieve cross-layer optimization [6]. Thus, new scheduling algorithms that use the PHY layer information to schedule users at the medium access control (MAC) layer such that the overall SINR is improved have been proposed. Some of these algorithms exploit channel variability to schedule a single user in a given time slot following the greedy approach leading to multiuser diversity gain. By the joint consideration of PHY and MAC layer issues, cross-layer designs not only improve system performance in terms of SINR [75]-[43] but also in terms of delay [44]-[46].

To gain further insights into the behaviour of scheduling algorithms the effect of channel availability and traffic conditions should also be considered. Tassiulas et al. [83] designed a scheduling algorithm for a wireless system consisting of multiple

queues and a single server. The arrivals at the queues were assumed be Bernoulli distributed with wireless channels varying between "ON" and "OFF" states. The authors found that serving users associated with the longest queue stabilizes the system provided their channels are in "ON" state (available). Tsibonis et al. [8] also designed a scheduling policy for a wireless system with time-varying connectivity, for general arrival rates. It was found that the scheduling policies proposed for maximizing throughput under various assumptions on the arrival rates may fail when channel availability is taken into account.

In this chapter a system model that also takes into account channel availability is presented to evaluate the performance of scheduling algorithms that serves two users simultaneously in terms of mean SINR and delay within a queueing framework. The packet arrivals and departures dictate the number of head of line (HoL) packets available in the queues from which packets are chosen for service depending on the scheduling decisions. For the analysis presented in this chapter three different scheduling algorithms are considered i.e., the greedy, the round-robin and the combined SNR and angular separation based algorithm [84] for both the numerical and semi-analytical analyses. Channel availability is modelled using a simple Markovian on/off model.

This chapter is organized as follows. The system model used for SINR analysis is presented in section 5.2. A semi-analytical model used to calculate SINR of a cross-layer downlink model based on channel availability and the three different scheduling algorithms used in both the numerical and semi-analytical models are presented in section 5.3. The semi-analytical framework uses analytical formulae, of probability density and cumulative distribution functions, as well as Monte-Carlo simulations. The numerical model is presented in section 5.4. Simulation results are presented in section 5.5 and conclusions in section 5.6.

5.2 System model

A multiuser cross-layer downlink model is considered where a single base station is capable of serving up to a maximum of two users simultaneously that are selected out of N available queues (as shown in Fig. 5.1). The packets destined for different mobile users are buffered temporarily inside the queues, each reserved for its respective mobile users, at the base station. The packets at the head of line (HoL) of each queue are scheduled based on different scheduling policies and they leave the queues based on channel availability and scheduling decisions. The three scheduling policies considered (the round-robin, the greedy and the combined SNR and angular separation based scheduling algorithms) are discussed later. Channel availability is modelled using a simple Markovian on/off model that leads to exponentially distributed time intervals between instants of channel availability. If packets are available to be sent when channels are available then the inter-departure times between packets also become exponentially distributed. Inter-arrival times between packets are modelled to be exponentially distributed as well. The buffer capacity is assumed to be infinite and the individual queues are thus modelled as M/M/1 queues where M/M/1 following Kendall's notation indicates a system in which arrivals are a Poisson distributed, service time is exponentially distributed and there is one server. It is also assumed that the maximum number of packets that can leave the system is two and thus the remaining packets are held in queues. It is also assumed that the base station has both the channel state information (CSI) and DoA, information available at the PHY layer based on the incoming pilot signals in the uplink. Each of the arriving packets is associated with a corresponding channel via which it is serviced and the angular location of users to which packets are destined is assumed to be randomly distributed between -90^o and $+90^o$ around the base station. The total available power at the base station is divided equally amongst all mobile users.

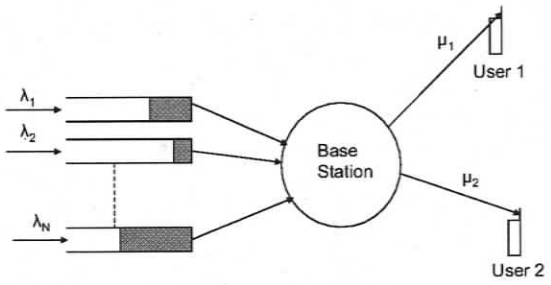


Figure 5.1: Block diagram of multiuser downlink model consisting of M/M/1 queue systems reserved for N mobile users.

5.3 Semi-analytical model

The functionality at the base station is separated into two components, the PHY and the MAC layer and these are discussed in the following sub-sections.

5.3.1 PHY layer

At the PHY layer the channel is assumed to be Rayleigh faded with instantaneous SNR (γ) distributed exponentially [67] and expressed as

$$p_\gamma(\gamma) = \frac{1}{\overline{\gamma}} \exp(\frac{-\gamma}{\overline{\gamma}}) \quad \text{where} \quad \gamma = \frac{|h^2|}{\sigma_{gn}^2} \tag{5.1}$$

parameter $\overline{\gamma}$ is the average SNR, h is the channel amplitude response and σ_{gn}^2 is the variance of the noise. When two users m and n are served simultaneously then they cause interference on each other and their SINR (Γ_i) is expressed as [80]

$$\Gamma_m = \frac{|h_m|^2}{\sigma_{gn}^2 + |h_n|^2 I_{nm}(\theta)} \quad \text{and} \quad \Gamma_n = \frac{|h_n|^2}{\sigma_{gn}^2 + |h_m|^2 I_{mn}(\theta)} \tag{5.2}$$

where h_m and h_n are the amplitude responses of the channels via which the users are served, and $I(\theta)$ is the interference scalar that varies between 0 and 1, depending on the angular separation (θ) between simultaneously served users and determines how much interference users cast on each other. We assume that $I_{mn}(\theta) = I_{nm}(\theta) = I(\theta)$ which implies that both users cast the same relative interference on each other although, of course, the absolute amount of interference depends on their channel amplitude responses (h_i). Eqn. (5.2) for user m may be simplified as

$$\Gamma_m = \frac{\frac{|h_m|^2}{\sigma_{gn}^2}}{1 + \frac{|h_n|^2 I_{nm}}{\sigma_{gn}^2}} = \frac{\gamma_m}{1 + \gamma_n I_{nm}} \tag{5.3}$$

where γ_m is the instantaneous SNR for user m, and similarly for user n. When users do not cast interference on each other, then $I = 0$ and $\Gamma_i = \gamma_i$ as mentioned in Chapter 4.

The interference scalar is estimated on the basis of a beam that is used for beamforming when the two users are served simultaneously since the beam pattern allows us to determine the fraction of power emitted in the direction of simultaneously served co-channel user based on the angular separation between the users. For a uniformly spaced linear antenna arrays with L antenna elements the beam pattern is given by

$$I(\theta) = V^T v \tag{5.4}$$

where V is the weighting vector and v is the array propagation vector that contains the information of the angle of arrival of the signal [2] given as

$$v = [1 \quad e^{j\kappa d sin\theta} \cdots \quad e^{j\kappa(L-1)dsin\theta}] \tag{5.5}$$

where $\kappa = 2\pi/\lambda$, λ is the wavelength, L is number of antenna elements and d in the inter element distance usually taken as $\lambda/2$.

It is shown in Chapter 4 that the above beam pattern can be approximated by

$$I(\theta) = e^{-|\theta|/c} \tag{5.6}$$

where θ is the angular separation and c is a parameter that is surrogate for, and inversely proportional to, the number of antenna elements. The beam pattern allows us to determine the leakage of power in the direction of simultaneously served co-channel users based on the angular separation between them. With the exponential beam pattern defined by Eqn. (5.6), the larger the angular separation between the users, the lower is the leakage and smaller the interference simultaneously served users cause on each another.

The PDF of the angular separation θ for a randomly distributed mobile user is given by a triangular distribution [80]

$$p(\theta) = \frac{2}{\pi} - \frac{2\theta}{\pi^2} \quad 0 \leq \theta \leq \pi \tag{5.7}$$

The PDFs of SINR of users m and n and interference I, depend on the scheduling algorithms used at the MAC layer, as discussed later.

5.3.2 MAC layer model

At the MAC layer, up to N packets destined for different mobile users are stored in the queues and they leave the queues based on channel and packet availability and the scheduling algorithm used. In this section the impact of channel availability and scheduling algorithms on the SINR of users is analyzed.

Channel availability model

At the MAC layer, the queues are modelled as M/M/1 queues with exponentially distributed inter-arrival and -departure times and infinite buffer capacity and thus no packet loss. The maximum number of packets that can depart in a given time

step is two, which implies that the maximum departure rate per queue cannot exceed 2/N packets/time step, where N is the total number of queues. The departure rate from a queue cannot exceed one (varying between 0 and 1), since only a maximum of one packet can depart in a given time step, and thus also represents the probability of packet departure. This probability of departure of a packet from a queue $(2/N)$ is, however, also affected by the unavailability of channel. The probability of channel availability, charaterized by its SNR (γ) value above some threshold (γ_{th}), is given by

$$P(\gamma > \gamma_{th}) = \int_{\gamma_{th}}^{\infty} p_{\gamma}(\gamma) d\gamma \tag{5.8}$$

where $p_{\gamma}(\gamma)$ is given by Eqn. (5.1). This yields the probability of a packet departing a queue (or the departure rate) as $x = \frac{2}{N}[P(\gamma > \gamma_{th})]$ and the probability that no packets will depart is $(1 - x)$. For the stability of the queues, the average arrival rate per queue (λ packets/time step) must be be less than the average departure rate (x), i.e., $\rho = \frac{\lambda}{x}$ should be less 1 and is determined as $\lambda = x \times \rho$ for a given ρ and specified γ^{th}. The behavior of queues, and in particular their queue lengths which are of interest here, is determined by ρ, the ratio of arrival to departure rates. For a single M/M/1 queue system, the queue length (q) is exponentially distributed and its PDF is given by [85]

$$p_q(q) = \rho^q(1 - \rho) \tag{5.9}$$

In Eqn. (5.9) the probability of the queue being empty $(p_q(q)|_{q=0})$ is $(1 - \rho)$ and thus the probability of the queue being not empty $(p_q(q)|_{q\neq 0})$ is ρ. For a system of N queues, each one behaving as an M/M/1 queue, it is possible to analytically obtain the distribution of the number of head of line (HoL) packets. The number of HoL packets (K) varies from zero (when all queues are empty) to N (when all queues are full). It is from the HoL packets that a scheduling algorithm chooses packets for

service. For a system of N queues the probability $p_K(K)$ that the number of HoL packets equals K is expressed in a generalized form as

$$p_K(K) = a_K(\rho)^K(1-\rho)^{N-K} \quad 0 \leq K \leq N \tag{5.10}$$

The coefficients a_K, $K = 0, 1, \ldots, N$ are symmetrical i.e., $a_0 = a_N$, $a_1 = a_{N-1}$ and given by $a_K = {}^NC_K$. For example, for N =4 and K =1, $a_1 = 4$ corresponding to four possible ways when a single queue has a HoL packet and the remaining three are empty (0001,0010,0100,1000).

Similar to Eqn. (5.10), the probability of n packets leaving, provided K HoL packets are available, can be written as

$$p_{n,K}(n) = a_{n,K}\, x^n(1-x)^{K-n} \quad 0 \leq n \leq K \tag{5.11}$$

where x is the probability of departure of a packet from a queue and coefficients $a_{n,K}$ are given by ${}^KC_n$. Since the packets available at HoL, K, vary, the probability of n packets being served ($\eta(n)$) is given by

$$\eta(n) = \sum_{K=n}^{N} p_K(K)p_{n,K}(n) \quad 0 \leq n \leq N \tag{5.12}$$

A maximum of two packets are served and therefore all $\eta(n)$ for $n \geq 2$ are assumed to contribute to the probability of two packets served. Ignoring the instances of zero departures, the fraction of instances when one (ζ_1) and two (ζ_2) packets are departing is given by

$$\zeta_1 = \frac{\eta(1)}{\sum_{j=1}^{N}\eta(j)} \quad \text{and} \quad \zeta_2 = \frac{\sum_{j=2}^{N}\eta(j)}{\sum_{j=1}^{N}\eta(j)} \tag{5.13}$$

In this framework, ζ_2 depends on ρ and N and in general, a high value of ρ leads to higher values of ζ_2. The selection of simultaneously served users m and n is based

on the scheduling policies that affect the SINR of the users as discussed below. Users are simultaneously served for only a ζ_2 fraction of time. The remaining fraction of time (ζ_1) only a single user is served.

SINR calculation for the scheduling algorithms

The scheduling algorithms considered within this queuing framework are the round-robin, the greedy and the combined SNR and angular separation based scheduling algorithms.

1. Round-robin (RR) scheduling algorithm [75]: In case of the RR scheduling algorithm the users are selected randomly irrespective of their channel state information or angular separation and the PDF of the instantaneous SNR, γ_i, is given by Eqn. (5.1). Also given I as a function of θ (Eqn. (5.6)) and PDF of θ (Eqn. (5.7)), the PDF of I, $p_I(I)$, for randomly distributed mobile users around the base station is derived as shown in Chapter 4 as

$$p_I(I) = \frac{2c^2 \ln I + 2\pi c}{\pi^2 I} \tag{5.14}$$

   Given the PDFs of the SNR of users 1 and 2 (Eqn. (5.1)), (γ_i), (that depend on the fading model) and interference I, it is possible to obtain multiple realizations of Γ_1 and Γ_2 (Eqn. (5.3)) and a large number of values of Γ_1 and Γ_2 are used to calculate the average SINR ($\overline{\Gamma}$) taking into account that two users are served only a ζ_2 fraction of time.

2. Greedy scheduling algorithm [38]: The greedy scheduling algorithm selects users based on their channel's instantaneous SNRs, γ_i. When only one user is being serviced, the user with highest γ_i from the available K users at the HoL is selected. When two users are to be serviced then the users with the highest and second highest values of γ_i are selected. Given the PDF of instantaneous

SNR (Eqn. (5.1)), the PDF of the l^{th} best user out of total of K users with available HoL packets following Chapter 4 is obtained as

$$p_{\gamma_{l:K}}(\gamma) = \frac{K!}{(K-l)!(l-1)!}[P_\gamma(\gamma)]^{K-l}[1-P_\gamma(\gamma)]^{l-1}p_\gamma(\gamma) \tag{5.15}$$

However, K is not a constant and in particular depends on ρ. If $\rho \approx 1$ then there would always be HoL packets available in all queues to choose from and $K = N$. For $\rho < 1$, however, K can be less than N. We use the PDF of number of HoL packets (Eqn. (5.10)) to obtain the resulting PDF of instantaneous SNR of the best and the second best user when K varies from 0 to N depending on ρ. In essence, we weigh the PDF $p_{\gamma_{l:K}}(\gamma)$ with $p_K(K)$. The resulting PDF of the first and second best user is thus

$$p_{\gamma_{1:K,K=1,N}}(\gamma_1) = \sum_{K=1}^{N} p_{\gamma_{1:K}}(\gamma_1)\frac{p_K(K)}{\sum_{K_1=1}^{N} p_K(K_1)} \tag{5.16}$$

$$p_{\gamma_{2:K,K=2,N}}(\gamma_2) = \sum_{K=2}^{N} p_{\gamma_{2:K}}(\gamma_2)\frac{p_K(K)}{\sum_{K_1=2}^{N} p_K(K_1)} \tag{5.17}$$

Since the two users are selected on the basis of their CSI irrespective of their angular location, the PDF of I, i.e., $p_I(I)$, given by (5.14) remains valid. Given $p_{\gamma_{1:K,K=1,N}}(\gamma_1)$ (Eqn. 5.16), $p_{\gamma_{2:K,K=2,N}}(\gamma_2)$ (Eqn. 5.17) and $p_I(I)$ the average values of Γ_1 and Γ_2 are obtained in a manner similar to that for the RR scheduling algorithm.

3. Combined SNR and angular separation based scheduling algorithm: The combined SNR and angular separation based scheduling algorithm selects the first user for service in a given time slot in a manner similar to that in a greedy algorithm i.e., the user associated with the highest value of instantaneous SNR whose PDF is given by $p_{\gamma_{1:K,K=1,N}}(\gamma_1)$ (Eqn. 5.16). The second user, however, is

selected such that it has the maximum angular separation with respect to the first user. Since the second user is selected on the basis of its angular location irrespective of its CSI and because γ_i and the angular location are uncorrelated, the PDF of SNR of the second user is the same as that for the RR algorithm (Eqn. (5.1)). The selection of the second user is aimed at minimizing the interference between the simultaneously served users and therefore the PDF of I, $p_I(I)$, is modified. The PDF of the 1^{st} largest angular separation θ out of total of $M = K - 1$ angular separations is obtained following Chapter 4 as

$$p_{\theta_{1:M}}(\theta) = \frac{M!}{(M-1)!}\left[P_\theta(\theta)\right]^{M-1} p_\theta(\theta) \tag{5.18}$$

where $P_\theta(\theta) = \frac{2\theta}{\pi} - \frac{\theta^2}{\pi^2}$ is the cumulative distribution function of θ. The $p_{\theta_{1:M}}$ is thus given by

$$p_{\theta_{1:M}}(\theta) = \frac{M!}{(M-1)!}\left[\frac{2\theta}{\pi} - \frac{\theta^2}{\pi^2}\right]^{(M-1)}\left[\frac{2}{\pi} - \frac{2\theta}{\pi^2}\right] \tag{5.19}$$

which is used to obtain $p_{I_3}(I_3)$, the PDF of interference scalar for the combined SNR and angular separation based scheduling algorithm, as expressed in Chapter 4 as

$$p_{I_3}(I_3) = \frac{Mc}{I_3}\left(\frac{-2c\ln I_3}{\pi} - \frac{(-2c\ln I_3)^2}{\pi^2}\right)^{M-1}\left(\frac{2}{\pi} + \frac{2c\ln I_3}{\pi^2}\right) \tag{5.20}$$

Given $p_{\gamma_{1:K,K=1,N}}(\gamma_1)$, $p_\gamma(\gamma_2)$ and $p_{I_3}(I_3)$, the average values of Γ_1 and Γ_2 can be obtained numerically.

5.4 Numerical model

The system model discussed in section 5.2 was also simulated numerically and similar to the semi-analytical framework it was assumed that in a given time slot a maxi-

mum of two users are served simultaneously out of K users available at the HoL of N queues. The packets destined for the remaining users were held in the queues. The packets destined for different users at the MAC layer depart through the PHY layer based on channel availability and scheduling algorithm used. The channel availability is modelled using a simple Markovian on/off model that leads to exponentially distributed time intervals between instants of channel availability and the individual queues essentially behave as M/M/1 queues. The three scheduling algorithms discussed in section 5.3 were considered for the numerical model as well. The channel is assumed to be Rayleigh faded and it is also assumed that the base station has beamforming capability and similar to the analytical model the beam pattern was approximated to be exponential. The value of c used was 73 for the exponential beam which corresponds to $L = 4$ in a uniformly spaced linear antenna array as discussed earlier. The noise power σ_{gn}^2 was assumed to be 1. The channel states associated with each packet, as they wait in their respective queues, are changed randomly. The numerical model was also used to calculate the delay suffered by the users as they wait in the queues for service. Finally, the buffer size in the numerical model was changed from infinite to a fixed capacity B to take into account the effect of scheduling algorithms on packet loss.

5.5 Simulation results

Using the described system model, the results from the numerical and semi-analytical models are compared first. The semi-analytical model is then used to assess the performance of the different scheduling algorithms in terms of average SINR and to study the impact of traffic conditions on the performance of the scheduling algorithms. The average SINR ($\overline{\Gamma}$) is based on the SNR when a single user is served as well as the SINRs of users that are served simultaneously and thus is given by

$$\overline{\Gamma} = \zeta_1 \overline{\gamma_1} + \frac{\zeta_2(\overline{\Gamma_1} + \overline{\Gamma_2})}{2} \tag{5.21}$$

where $\overline{\gamma_1}$ is the average SNR when a single user is served, $\overline{\Gamma_1}$ and $\overline{\Gamma_2}$ are the average SINRs of two users served simultaneously and ζ_1 and ζ_2 are the fractions of times when a single and two users are served as mentioned earlier. Finally, the numerical model is used to assess the effect of scheduling algorithms on delay suffered by packets using a delay metric d_{95} that represents the delay in the number of time steps such that the probability exceeds 5% i.e.,

$$\text{prob}\,[d > d_{95}] \approx 0.05 \tag{5.22}$$

The numerical model was also used to assess the effect of scheduling algorithms on packet loss when the buffer capacity is fixed.

5.5.1 Comparison of semi-analytical and numerical models

Figure 5.2 compares the performance of the scheduling algorithms in terms of average SINR for the number of queues (N) ranging from 4 to 10 using both the numerical and the semi-analytical models. For both the numerical and semi-analytical model similar behaviour is exhibited in terms of average SINR for the three scheduling algorithms considered. Average SINR is calculated based on instances when either one or two users are served. As can be seen, the average SINR for the RR scheduling algorithm remains constant irrespective of the total number of users because of its unpreferential treatment of users. The average SINR for other scheduling algorithms, however, increases with an increase in N because there is more choice in selecting a packet from the available HoL packets. The combined SNR and angular separation based algorithm yields higher average SINR than the greedy algorithm because it also attempts to reduce interference by taking into account the angular location of

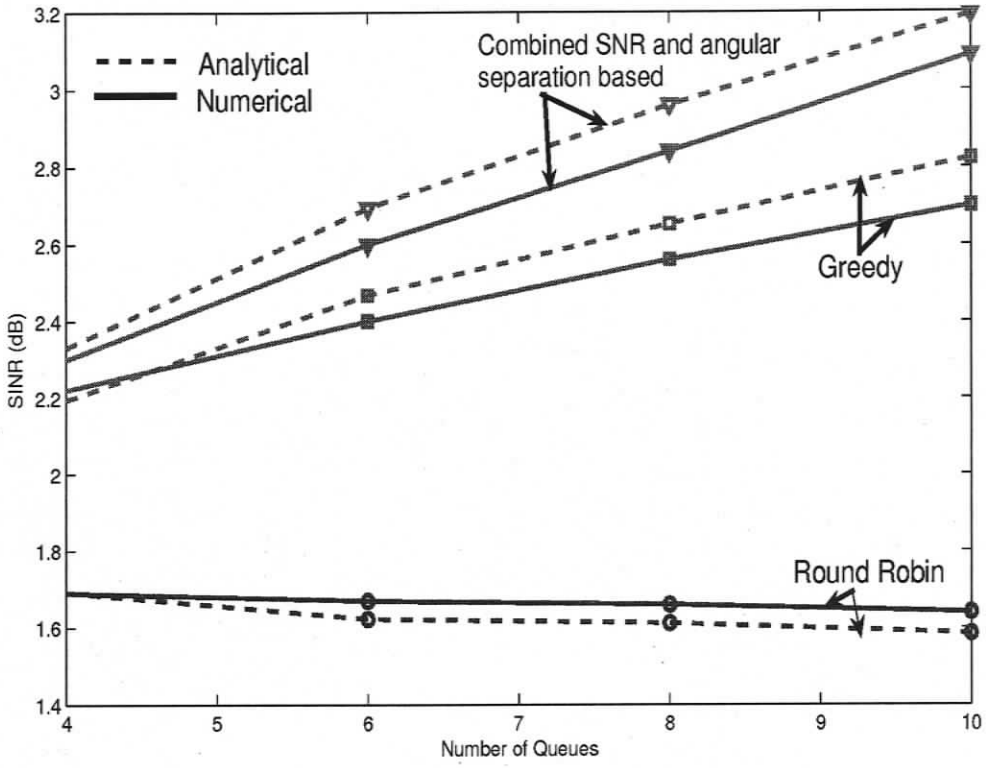


Figure 5.2: Effect of number of queues on average SINR for different scheduling algorithms. Results are from both the numerical and the semi-analytical models.

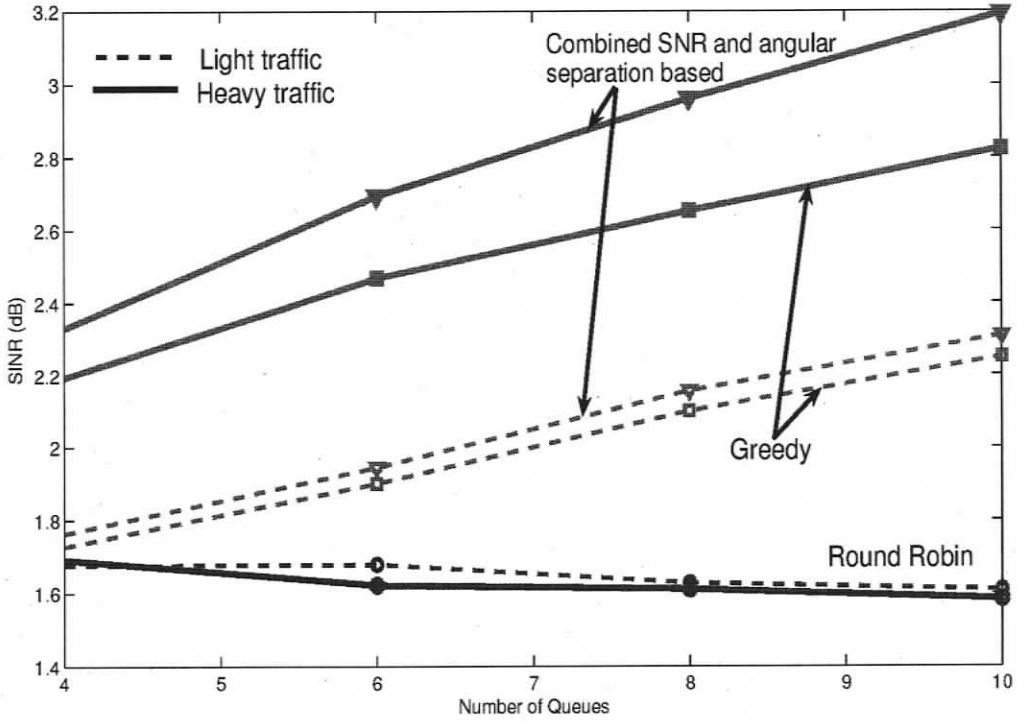


Figure 5.3: Effect of heavy (solid) and light (dotted) traffic conditions (as explained in text) on average SINR for the three scheduling algorithms. Parameter values are $c = 73$, $\sigma_{gn}^2 = 1$ and buffer size (B) = infinite.

mobiles around the base station when two users are served simultaneously. Since the fraction of time when two users are served simultaneously (ζ_2) increases with N, the relative advantage of the combined SNR and angular separation based algorithm over the greedy algorithm also increases with an increase in N.

5.5.2 Effect of traffic load

Fig. 5.3 shows the effect of traffic conditions on average SINR based on three scheduling algorithms obtained using the semi-analytical model. For the simulations presented in this section two different scenarios corresponding to light and heavy traffic conditions were considered. Here light and heavy traffic conditions refer to a lower and a higher number of packets entering the system while they depart at the same

rate. The results show that for both these scenarios the combined SNR and angular separation based algorithm yields higher $\overline{\Gamma}$ than the greedy and RR scheduling algorithms. An increase in traffic load leads to decreased SINR for the RR algorithm because of an increase in fraction of instances when two users are simultaneously served (ζ_2) which leads to more interference. The greedy and the combined SNR and angular separation based algorithms, yield higher average SINR for heavy traffic scenario due to the increased choice in selecting users because higher traffic load also leads to more HoL packets. Fig. 5.3 also shows that the difference between scheduling algorithms is larger for increasing N and for heavy traffic conditions than for light traffic conditions. This implies that the benefits of scheduling algorithms that choose users preferentially according to some criteria are realized only for heavy traffic conditions when more choice is available in selecting packets for service.

5.5.3 Comparison of packet loss rates

In addition to the average SINR, the delay and packet loss are QoS criteria that determine the overall performance of scheduling algorithms. Although high SINR is certainly preferable it may be unacceptable if it is associated with high packet delay and packet loss when buffer size at the queues is fixed. The packet loss rates for the three scheduling algorithms are compared using numerical simulations for $B = 10$. A fixed buffer size does not have any significant impact on the average SINR obtained using three different scheduling algorithms and the results are similar to those obtained with infinite buffer capacity (Figure 5.2). This is because $\overline{\Gamma}$ is governed by SNR and the angular location of packets available at the HoL. Limited buffer capacity, however, does cause packet loss because when the packets waiting at the HoL are not served due to channel unavailability or their low priority based on scheduling algorithms, the queues get filled up thereby leading to loss of packets. In Figure 5.4 a comparison of the three scheduling algorithms show that the round-robin scheduling algorithm yields the lowest packet loss among the three algorithms. This

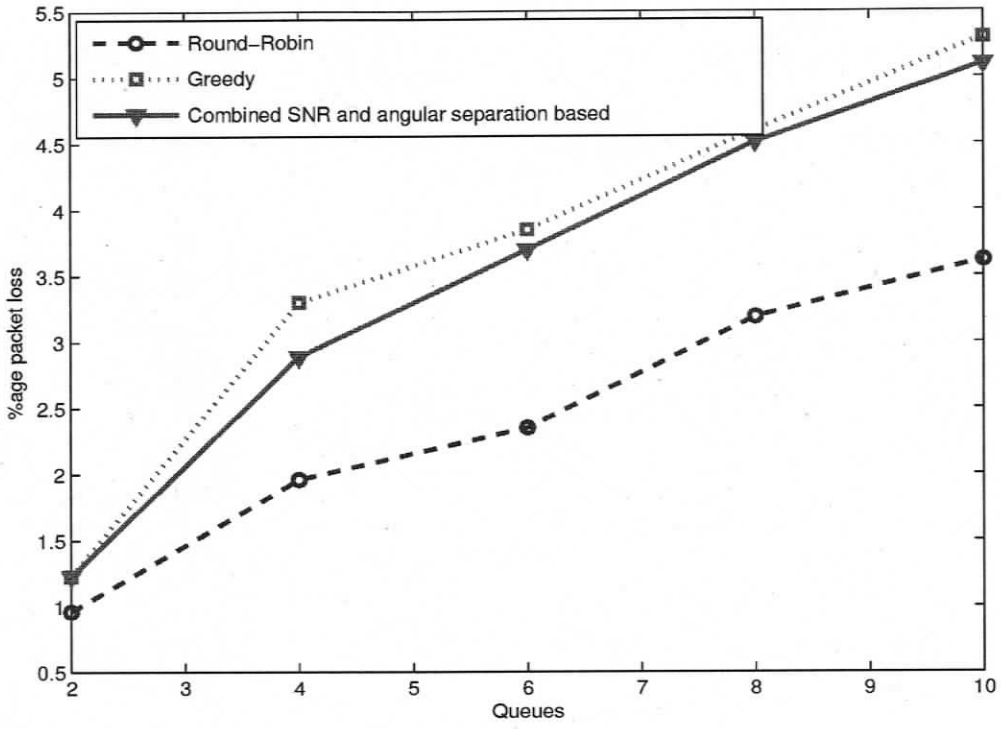


Figure 5.4: Effect of buffer size B (=10) on packet loss for different scheduling algorithms.

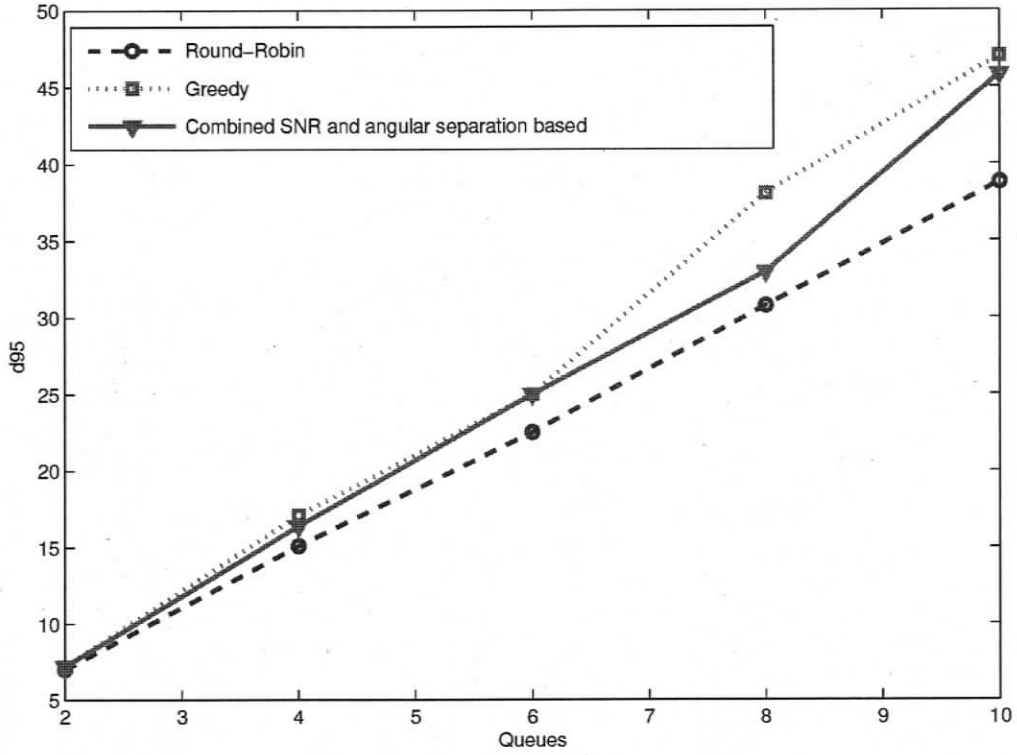


Figure 5.5: Effect of scheduling algorithms on delay.

is due to the unpreferential treatment of packets associated with different SNRs that does not lead to long queues which eventually would cause packet loss. Both the greedy and the combined SNR and angular separation based algorithms that choose packets preferentially yield similar and somewhat higher packets loss rates then the round-robin algorithm.

5.5.4 Comparison of delay

Figure 5.5 compares the performance of the scheduling algorithms in terms of the delay metric d_{95}. As expected, the round-robin scheduling algorithm yields less delay compared to the greedy and the combined SNR and angular separation based scheduling algorithms due to its unpreferential treatment of users. The greedy and

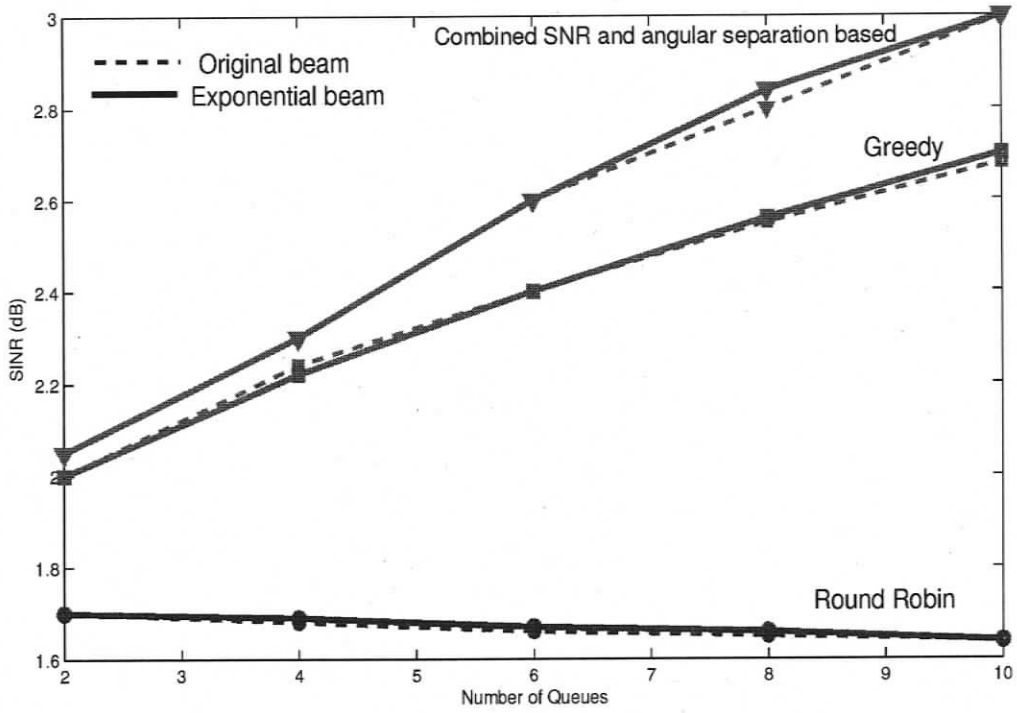


Figure 5.6: Effect of beam pattern on average SINR for different scheduling algorithms based on numerical model.

the combined SNR and angular separation scheduling algorithms yield similar delays. The delay for all three scheduling algorithms increases with an increase in number of queues due to increased number of users waiting for the service at the HoL.

5.5.5 Comparing exponential and actual beam patterns

Since the beam pattern used in the semi-analytical model was an approximation of the actual beam pattern, it would be interesting to evaluate the accuracy of this approximation. The performance of the scheduling algorithms in terms of average SINR for number of queues (N) ranging from 4 to 10 using both the exponential and the actual beam patterns is shown in Figure 5.6. Figure 5.6 shows that the results obtained using the exponential beam pattern are very similar to those obtained using the actual beam pattern indicating that the approximation of the actual beam pattern

obtained using $L = 4$ antenna elements with that of an exponential beam pattern with $c = 73$ is reasonable.

5.6 Conclusions

The performance of the three different scheduling algorithms described in Chapter 4 has been evaluated using semi-analytical and numerical models for a cross-layered system within a queuing framework. The system model includes queues modelled as M/M/1 with exponentially distributed inter-arrival and inter-departure times (corresponding to the times when the channel is available) traffic load which depends on both arrivals and departures of packets in the system and the availability of CSI and DoA information at the physical layer. Three different scheduling algorithms were considered, the greedy, round-robin and the combined SNR and angular separation based algorithm. The results indicate that for low traffic, there is no significant difference in the performance between the three algorithms because the fraction of time two users are served simultaneously is low and this leads to a small difference in the SINR for the greedy and combined SNR and angular separation based scheduling algorithms. For high traffic conditions, the combined SNR and angular separation based scheduling algorithm gives the highest SINR implying that taking into account the angular location of mobile users can yield better performance. The round-robin scheduling algorithm that selects users unpreferrentially has the advantage of somewhat lower delay and packet loss compared to the other two algorithms but at the cost of a lower SINR. The results show that designing a system model which takes into account channel availability and traffic conditions provides further insights into the behaviour of the scheduling algorithms.

Chapter 6

A modified beamformer to solve the hidden beam problem

6.1 Introduction

As mentioned in Chapter 2, cross-layered systems offer a promising solution to support both real and non-real time services following either the bottom-up or the top-down approach. Bottom-up approaches of cross-layer design (CLD) primarily attempt to design algorithms or protocols at the higher layers (e.g., medium access control (MAC)) based on the resources and information available at the lower layers (e.g., physical (PHY)) to improve system performance. The top-down approaches, on the other hand, try to find the best way to transmit data reliably to mobile users selected for service at the higher layers using techniques such as space, time or frequency diversity and beamforming at the PHY layer. The advantages of using the space-time diversity and/or beamforming techniques are discussed in detail in Chapter 3. The development of beamforming techniques [15]-[16] and their use in wireless systems offers better directivity and substantial power savings compared to omni-directional antennas. This has led to research focussing on the development of several beamforming algorithms [3]. The use of beamforming is not only limited to mobile phones but has also made its way into the area of wireless local area networks

(LANs). This has also led to re-designing the MAC protocols such as ALOHA and carrier sense medium access (CSMA) [86]-[87] based on using directional antennas at PHY layer. MAC protocols are a set of rules that allow the users to share a common channel for the fair and orderly transfer of data. Issues raised by implementation of beamforming at the PHY layer in the systems using medium access control protocols such as CSMA have recently gained attention due to the increased popularity of CSMA based systems including Ethernet, Wi-Fi and the Bluetooth technologies. In systems using CSMA protocols, users wishing to transmit have to first listen to sense the channel (medium) to check for any activity on the channel. If the channel is sensed as "idle" then they transmit, else they continue listening. When directional antennas are used in systems based on the CSMA protocol, they present their own share of issues such as the deafness, the hidden terminal problem, and the hidden beam problem [49]. Deafness occurs when the transmitter fails to communicate to its intended receiver, because the receiver's antenna is oriented in a different direction. Due to deafness a transmitter's repeated attempts to send a "request to send" (RTS) signal to a destination are not acknowledged with a "clear to send" (CTS) signal. The directional hidden terminal problem occurs when a transmitter is unable to sense an ongoing uplink transmission of another transmitter because of its antenna orientation and initiates its own uplink transmission thus causing a collision. Both deafness and hidden terminal are pure uplink (mobile to base station) problems. When a base station is transmitting data to a desired mobile terminal using a directional beam, with maximum energy radiated in the direction of this user and nulls or side lobes formed towards other directions, the mobile terminals lying in deep nulls are not able to sense the busy channel due to low power. If these mobile terminals start transmitting packet collisions occur. This problem is known as the hidden beam problem and leads to performance degradation. The hidden beam is an uplink/downlink problem where an uplink packet is lost due to ongoing downlink transmission.

While the hidden terminal problem and deafness have been investigated in earlier studies [50], [88], the hidden beam problem has only been addressed recently ([51], [52]). In this chapter we investigate the cross-layer design problem of the hidden beam. In particular, we address the hidden beam problem at the PHY layer such that the performance of a system using CSMA protocol at the MAC layer can be improved.

Tarokh et al. [51], [52] proposed two approaches to solve the hidden beam problem. In their first approach they direct additional power towards the users that cannot sense the medium using a complementary beam [51]. The data bearing directional beams steered towards the desired users are combined with the appropriate power level broadcast signal directed towards other users thus enabling them to sense the medium. Their second approach forms a complementary beam by expanding the beamforming weight matrix and is referred to as subspace complementary beamforming [52]. First a complementary beam is generated using additional dummy data streams such that the total beam pattern becomes omni-directional. This complementary beam causes interference to the desired mobile terminals. In order to eliminate interference a projection technique is applied such that the complementary beam forms nulls towards the desired users. Finally, the complementary beam level is adjusted through scaling such that the received power in the complementary region becomes equal to or greater than some minimum threshold required to sense the medium.

The hidden beam problem can also be addressed by altering the beamforming weight vector to yield a new beam pattern that radiates some specified minimum power in the direction of passive users allowing them to sense the medium. Tarokh et al. [52] argue that subspace complementary beamforming technique offers more flexibility than altering the weight vectors to generate a new beam pattern. However, while the approach presented in [52] offers more flexibility in terms of raising the min-

imum power level to a desired level it is also associated with increased computational complexity.

In this chapter, a methodology for altering the beamforming weight vector of an original beam, to reduce the hidden beam problem is presented. The beamforming weight vector is modified such that the minimum power level of the users lying in the deep nulls is raised enabling them to sense the medium. The design methodology followed to achieve the desired beam along with its limitations is discussed. The effect of including the modified beam developed here on the throughput of slotted non-persistent CSMA based systems is also investigated.

This chapter is organized as follows. A new methodology to address the hidden beam problem is presented in section 6.2 and the CSMA system model considered is presented in section 6.3. Results are presented in section 6.4 that discusses the improvement in throughput obtained by using the proposed methodology and finally conclusions are presented in section 6.5.

6.2 Proposed methodology to solve the hidden beam problem

For a uniformly spaced linear antenna array with L antenna elements the array factor $F(\theta, \theta_0)$ is expressed as

$$F(\theta, \theta_0) = \mathbf{V^T u} \tag{6.1}$$

where $\mathbf{V^T} = [V_0, V_1, \ldots, V_{L-1}]$ is the weighting vector and the weights V_k are given by

$$V_k = A_k e^{-jk\kappa d \cos\theta_0}, \quad k = 0, 1, \ldots, L-1 \tag{6.2}$$

where θ_0 is the direction in which the beam is steered, κ=$2\pi/\lambda$, λ is the wavelength, d is the inter-element distance usually taken as $\lambda/2$ and A_k is the uniform amplitude equal to one for all k and $\mathbf{u}$ is the array propagation vector given by

$$\mathbf{u}^T = [1\ e^{j\kappa d\cos\theta} \cdots e^{j\kappa d(L-1)\cos\theta}] \tag{6.3}$$

The array factor is thus written as [2]

$$F(\theta,\theta_0) = \sum_{k=0}^{L-1} A_k e^{j(k\kappa d(\cos\theta - \cos\theta_0))} \tag{6.4}$$

and is expressed in closed form [3] as

$$F(\theta,\theta_0) = \frac{e^{jL\psi}-1}{e^{j\psi}-1} \tag{6.5}$$

where $\psi = \kappa d(\cos\theta - \cos\theta_0)$. The normalized power gain of the array is then expressed in dB as [3]

$$P_{dB}(\theta)(dB) = 10\ \log\left[\frac{|F(\theta,\theta_0)|^2}{(max_\theta|F(\theta,\theta_0)|)^2}\bigg|\ \theta_0 = \hat{\theta}_0\right] \tag{6.6}$$

where $\hat{\theta}_0$ is assumed here to be 90^0. The maximum response of the array factor $F(\theta,\theta_0)$ in Eqn. (6.4) results at an angle θ_0 and the beam is steered in this direction with little or no energy in other directions leading to the hidden beam problem.

The hidden beam problem is addressed by altering the weighting vector (Eqn. (6.2)) to reduce the depth of the nulls. The modified beam pattern $F_m(\theta,\theta_0,\phi)$ is obtained as an average of two original beams shifted by $+\phi$ and $-\phi$ degrees, which is now also a function of ϕ, i.e.,

$$F_m(\theta,\theta_0,\phi) = \frac{1}{2}\left(F(\theta,\theta_0+\phi) + F(\theta,\theta_0-\phi)\right) \tag{6.7}$$

where

$$F(\theta, \theta_0 + \phi) = \sum_{k=0}^{L-1} A_k e^{j(k\kappa d(\cos\theta - \cos(\theta_0 + \phi)))} \tag{6.8}$$

and

$$F(\theta, \theta_0 - \phi) = \sum_{k=0}^{L-1} A_k e^{j(k\kappa d(\cos\theta - \cos(\theta_0 - \phi)))} \tag{6.9}$$

Substituting Eqns. (6.8) and (6.9) in (6.7) yields an expression for $F_m(\theta, \theta_0, \phi)$ given by

$$F(\theta, \theta_0, \phi) = \frac{1}{2}\left(\sum_{k=0}^{L-1} A_k e^{j(k\kappa d(\cos\theta - \cos(\theta_0 - \phi)))} + A_k e^{j(k\kappa d(\cos\theta - \cos(\theta_0 + \phi)))}\right) \tag{6.10}$$

Using trigonometric identities and assuming $A_k = 1$,

$$F(\theta, \theta_0, \phi) = \frac{1}{2}\left(\sum_{k=0}^{L-1} e^{j(k\kappa d\cos\theta)}(e^{-jkdk(\cos\theta_0\cos\phi - \sin\theta_0\sin\phi)}) + e^{j(k\kappa d\cos\theta)}(e^{-jkdk(\cos\theta_0\cos\phi + \sin\theta_0\sin\phi)})\right) \tag{6.11}$$

$$F(\theta, \theta_0, \phi) = \frac{1}{2}\left(\sum_{k=0}^{L-1} e^{j(k\kappa d\cos\theta)}(e^{-jkkd\cos\theta_0\cos\phi})(e^{jk\kappa d\sin\theta_0\sin\phi} + e^{-jkkd\sin\theta_0\sin\phi})\right) \tag{6.12}$$

$$F_m(\theta, \theta_0, \phi) = \sum_{k=0}^{L-1} A_k e^{j(k\kappa d(\cos\theta - \cos\theta_0\cos\phi))}\cos(\kappa dk\sin\theta_0\sin\phi) \tag{6.13}$$

In Eqn. (6.13) V_k of Eqn. (6.2) is now modified to

$V_k = A_k e^{-jk\kappa d\cos\theta_0\cos\phi}\cos(\kappa dk\sin\theta_0\sin\phi)$.

Similar to Eqn.(6.5), Eqn.(6.13) is written in closed form as

$$F_m(\theta, \theta_0, \phi) = \frac{1}{2}\left[\frac{e^{jL(\xi+\beta)} - 1}{e^{j(\xi+\beta)} - 1} + \frac{e^{jL(\xi-\beta)} - 1}{e^{j(\xi-\beta)} - 1}\right] \tag{6.14}$$

where $\xi = \kappa d(\cos\theta - \cos\theta_0 \cos\phi)$ and $\beta = \kappa d(\sin\theta_0 \sin\phi)$. Note that as $\phi \rightarrow 0$, then $F_m(\theta, \theta_0, \phi) \rightarrow F(\theta, \theta_0)$ and the original beam is retrieved.

Similar to Eqn. (6.6), the normalized array gain for the modified beam $P_m(\theta, \phi)$, is given in dB as

$$P_{m,dB}(\theta, \phi)(dB) = 10 \log \left[\frac{|F_m(\theta, \theta_0, \phi)|^2}{(\max_\theta |F_m(\theta, \theta_0, \phi)|^2)} \middle| \theta_0 = \hat{\theta}_0 \right] \tag{6.15}$$

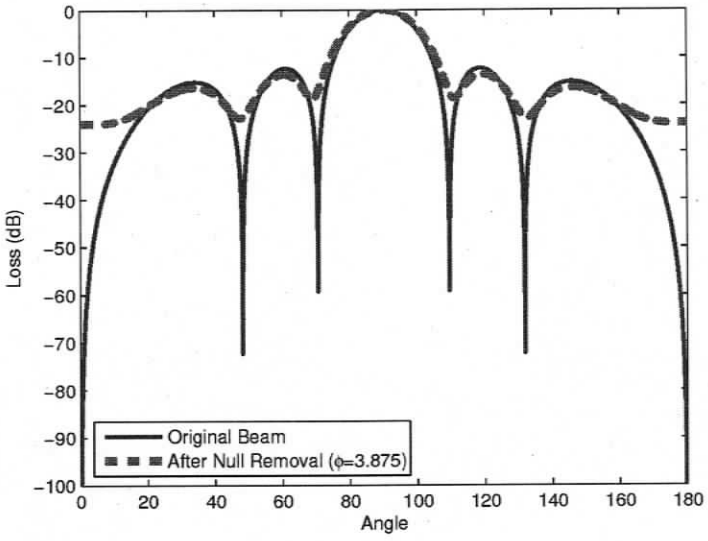
6.2.1 Raising the null depth level using ϕ

Figures 6.1a and 6.2a show how an increase in ϕ from 0° to 3.875° for $L = 6$ and from 0° to 2.96° for $L = 9$ raises the maximum null depth from less than -70 dB to around -24 dB. Figure 6.4 and Figure 6.5 plot this maximum null depth against ϕ and show how a gradual increase in ϕ reduces the maximum depth of the nulls by raising the minimum power level $P_{m,min}(\theta, \phi)$ for $L = 6$ and 9, respectively.

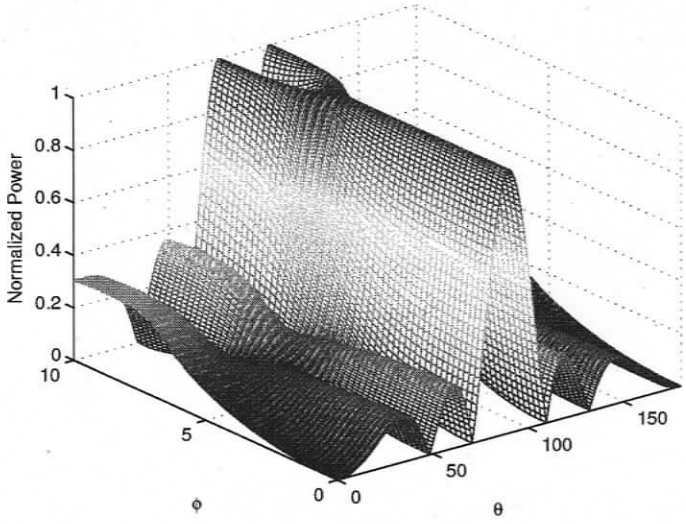
While an increase in ϕ reduces the null depths, there is an upper threshold, referred to as ϕ_{max}, after which the null depth starts increasing. The value of the ϕ_{max} decreases with L, e.g., for $L = 6$, the value of ϕ_{max} is 5.2° while for $L = 9$ it is 3.4°. Beyond ϕ_{max} not only the null depth increases but also the main lobe doubles shifting the null locations as shown in Figure 6.1, lower panel for $L = 6$ and Figure 6.2, lower panel for $L = 9$ respectively, which is undesirable.

In Eqns. (6.13) - (6.15) the parameter ϕ is unknown and the objective is to find the ϕ that raises the minimum power level $P_{m,min,dB}(\theta, \phi)$ to certain minimum desired level of $P_{des,dB}$, that would enable mobile terminals lying in the deep nulls to sense the medium. Obtaining the value of ϕ that would yield a desired $P_{m,min,dB}(\theta, \phi)$, however, is not straightforward and the approach we followed is explained below.

Given L, $\hat{\theta}_0$ and the minimum power threshold, we first find the locations of the nulls (θ_n) of the original beam ($\phi = 0$) as shown in the appendix,

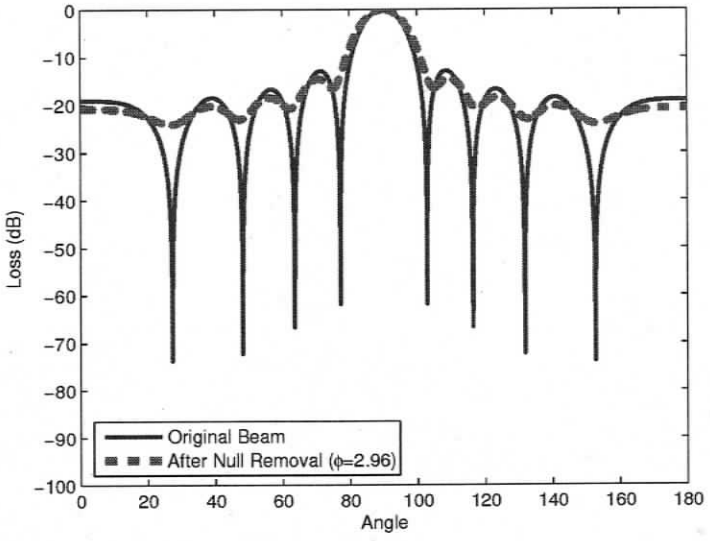


(a)

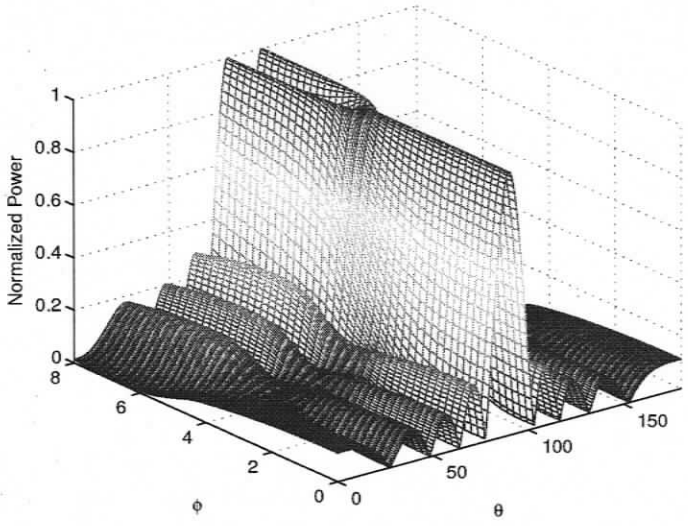


(b)

Figure 6.1: The effect of ϕ (in degrees) on reducing the null depths to -24 dB in the modified beamforming technique for $L = 6$ (panel a). The surface plot of the normalized power as a function of ϕ and θ (both in degrees) (panel b)



(a)



(b)

Figure 6.2: The effect of ϕ (in degrees) on reducing the null depths to -24 dB in the modified beamforming technique for $L = 9$ (panel a). The surface plot of the normalized power as a function of ϕ and θ (both in degrees) (panel b).

$$\theta_n = \cos^{-1}\left(\pm\frac{2n}{L}\right) \tag{6.16}$$

where n varies from $-L/2$ to $+L/2$, (n$\neq$ 0) for even L and $-(L-1)/2$ to $+(L-1)/2$, (n$\neq$ 0) for odd L. The magnitude of power at each nulls $P_{m,min}(\theta_n, \phi = 0)$ is also found.

For each null of the original beam, ϕ is increased in steps from zero ($\phi_{i+1} = \phi_i + \Delta\phi$, initial $\Delta\phi = 1^0$) and the location of that null in the modified beam is found that exists in between but not necessarily at the centre of the location of the nulls of the shifted beams which are given by

$$\theta_n = \cos^{-1}\left(\pm\frac{2n}{L} \pm \sin\phi\right) \tag{6.17}$$

Any basic line search method may be used to find the location of the null of the modified beam because there is a single minimum (the null) in the range defined by the nulls of the two shifted beams. The magnitude of the power at the nulls for both phase shifts ϕ_i and ϕ_{i+1} is found as $P_{m,min}(\theta_n, \phi_i)$ and $P_{m,min}(\theta_n, \phi_{i+1})$. If the desired power $P_{des,dB}$ lies in between $P_{m,min}(\theta_n, \phi_i)$ and $P_{m,min}(\theta_n, \phi_{i+1})$ then linear approximation is used until a ϕ corresponding to $P_{des,dB}$ within a certain tolerance is found using the algorithm shown in Figure 6.3, else the next iteration is performed.

The algorithm also ensures that at each iteration $dP_{m,min}(\theta_n, \phi)/d\phi$, the slope of the curve in Figures 6.4 and 6.5, is increasing such that the search for ϕ does not go past ϕ_{max}, else the step size ($\Delta\phi$) is halved and the iteration is performed again. When $\Delta\phi$ is sufficiently small then iterations stop and ϕ approaches ϕ_{max} that yields the maximum possible reduction in the null depth. Since ϕ is found for each null, its highest value amongst all nulls is taken as the desired ϕ that satisfies the minimum power level specifications.

The removal of deep nulls in the modified beam implies that the fraction of angular

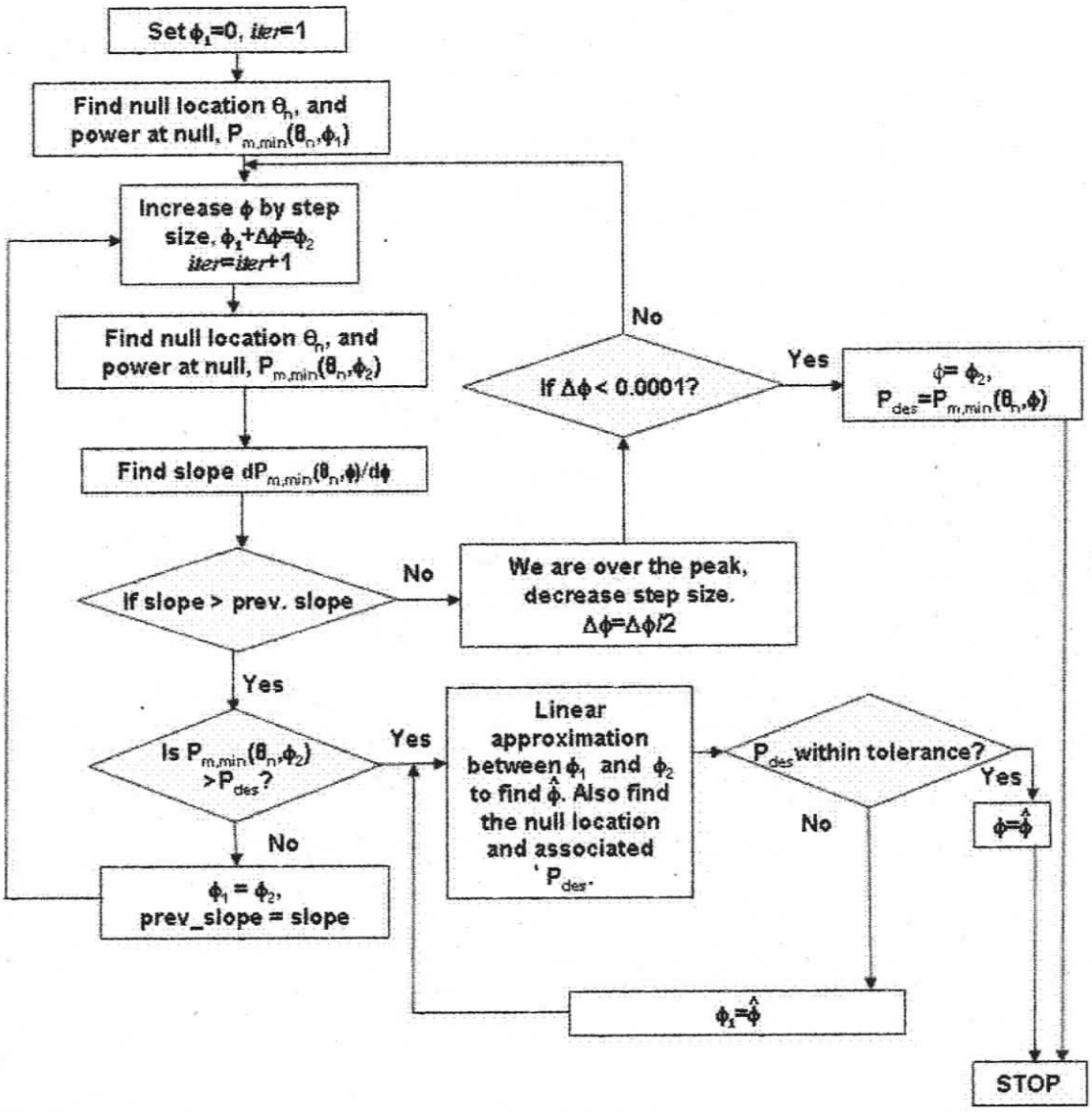


Figure 6.3: Flowchart illustrating the steps required to calculate ϕ for reducing null depths in the modified beamforming technique.

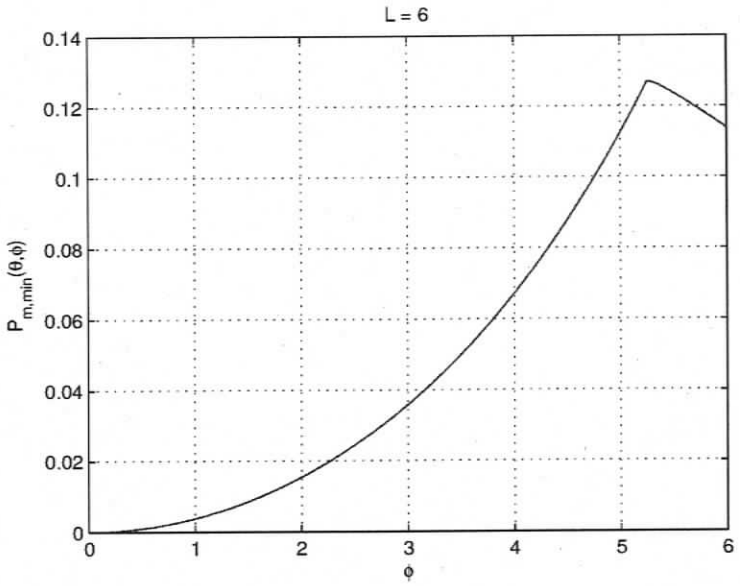


Figure 6.4: The effect of phase shift ϕ (in degrees) on the maximum depth of the nulls, i.e., the power $P_{m,min}(\theta, \phi)$, for $L = 6$. An increase in ϕ leads to an increase in power $P_{m,min}(\theta, \phi)$ until ϕ_{max} after which the power starts decreasing.

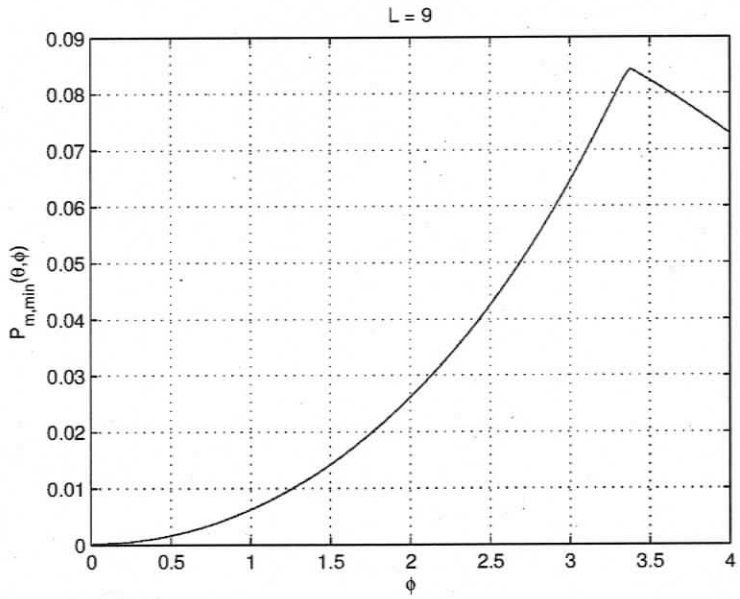


Figure 6.5: The effect of phase shift ϕ (in degrees) on the maximum depth of the nulls, i.e., the power $P_{m,min}(\theta, \phi)$, for $L = 9$. An increase in ϕ leads to an increase in power $P_{m,min}(\theta, \phi)$ until ϕ_{max} after which the power starts decreasing.

space below a specified power threshold also decreases. If the angular distribution of mobile terminals around the base station is random then the fraction of time users lie in nulls (η) given a specified power threshold is the same as the fraction of angular space below that specified power threshold.

6.3 Performance assessment of the modifed beam

To compare the structures of the modified and the original beams we compare their half power beam widths (HPBWs) and directivities. The performance of the modified beam is also assessed in terms of system throughput when implemented within a slotted non-persistent CSMA system.

6.3.1 Comparison of the basis beam parameters

The reduced depth of nulls in the modified beam is obtained at the cost of some widening of the main lobe. To quantify the differences between the modified and the original beams we compare their half-power beam widths (HPBW) and directivities. HPBW or the 3 dB beamwidth is the angular width measured between the points on the main lobe that are 3 dB below the peak of the main lobe for both the original and modified beams using their respective power curves $P_{dB}(\theta)$ and $P_{m,dB}(\theta, \phi)$. Directivity is the array gain against isotropic noise and obtained for original beam as

$$D = \left\{ \frac{1}{2} \int_0^{\pi} P(\theta) \sin\theta d\theta \right\}^{-1} \tag{6.18}$$

The directivity of the modified beam D_m is obtained in a similar way using $P_m(\theta, \phi)$. Both the directivity and HPBWs of the original and modified beams are compared in section 6.4.

6.3.2 Throughput analysis for a slotted non-persistent CSMA system

Consider a wireless system in which several mobile users wish to communicate with a single base station in an uplink. These users rely on carrier sensing before accessing

the medium and can communicate with the base station only when the medium is idle. The MAC protocol used to sense the medium is slotted non-persistent CSMA and only one user communicates with the base station in a given time slot, where a slot is a time unit required to fully transmit one packet. The base station is assumed to either operate in uplink or downlink mode during any given time slot and both operations cannot be performed simultaneously. It is also assumed that each user generates constant length packets independently and the collective traffic model of these users is considered to be Poisson distributed with average rate of λ packets/slot. Collision occurs during uplink transmissions when two users sensing the medium at any time slot sense it to be idle and thus start transmitting simultaneously. The traffic offered in an uplink system consists of a combination of both the newly generated packets and previously collided packets from the users and this increases the mean offered rate (packets/slot) denoted by G where $G \geq \lambda$,

The throughput of a slotted non-persistent CSMA system S_C, that represents packets that are successfully transmitted in a time slot, can be obtained by considering the time when the channel is utilized without collision (U), when the channel is busy (B) (defined as the time between arrival of a packet and until the channel becomes available again) and idle (I) times [89] as

$$S_C = \frac{\bar{U}}{\bar{B} + \bar{I}} \tag{6.19}$$

where the overbar represents the expected values of U, B and I. Eqn. (6.19) is solved in terms of offered load G for the case of zero propagation delay as [89]

$$S_C = G p_s = G \frac{1}{1+G} \tag{6.20}$$

where $p_s = \frac{1}{1+G}$ is the probability of success and throughput S_C (packets/slot) lies between 0 and 1, since only a maximum of one packet can be transmitted during a

time slot. For slotted non-persistent CSMA systems, when using an omni-directional antenna, packets are not transmitted when a downlink is in progress because mobile terminals can sense the downlink transmission. When a directional antenna is used for downlink transmission the throughput is affected by the hidden beam problem and in addition to the collisions due to simultaneous uplink transmission, packets are also lost when mobile terminals cannot sense an ongoing downlink transmission.

Consider that uplink and downlink transmissions occur f_u and f_d fractions of the time respectively. In absence of the hidden beam problem the throughput of the uplink transmissions during f_u fraction of time is S_C. If the hidden beam problem occurs η fraction of time during the downlink transmissions then ηf_d fraction of the total time mobile terminals sense no downlink transmission and the fraction of time suitable for uplink transmissions as seen by mobile terminals becomes $f_u + \eta f_d$.

Since we are interested in only the average throughput of uplink transmissions we consider the effect of the hidden beam over $f_u + \eta f_d$ fraction of time and normalize it with respect to f_u as

$$\frac{f_u + \eta f_d}{f_u} = 1 + \epsilon \tag{6.21}$$

where $\epsilon = \frac{\eta f_d}{f_u}$ is the duration when the hidden beam problem occurs. Mobile terminals see additional ϵ duration of time suitable for uplink transmissions and this increases the busy and idle times. However all the packets transmitted during the increased duration are lost and thus the utilization time without collision remains the same. Equivalently, the probability of success for a slotted non-persistent CSMA system suffering with the hidden beam problem is $p_{s,hb} = \frac{1}{(1+G)(1+\epsilon)}$ which yields the corresponding throughput as

$$S_{C,hb} = \frac{G}{(1+G)(1+\epsilon)} \tag{6.22}$$

L	3	6	9	12	15	18
ϕ_{max}	12	5.2	3.5	2.51	2	1.75
P_{max}(dB)	-12.9	-18.15	-21.8	-23.7	-25.8	-25.8

Table 6.1: Maximum power depth level that can be achieved for different number of antenna elements (L) using corresponding values of ϕ_{max}.

As expected, $S_{C,hb} \rightarrow S_C$ as $\eta \rightarrow 0$ and $\epsilon \rightarrow 0$. The throughput $S_{C,hb}$ depends on the uplink to downlink ratio $f_u : f_d$. If the $f_u : f_d$ ratio is 1:0, implying no downlink transmissions at all, then $\epsilon = 0$ and the throughput is not affected by the hidden beam problem at all. In contrast, as the fraction of time when downlink transmissions occur increases then the $f_u : f_d$ ratio decreases and the effect of the hidden beam problem increases. The effect of the hidden beam problem is only experienced when $f_d > 0$.

6.4 Results

In this section the results obtained using the modified beam in terms of maximum power level achieved and the comparison of the original beamformer with the modified beamformer in terms of directivity and HPBW for different antenna elements L are presented. The improvement in throughput achieved by implementing the modified beam in non-persistent CSMA system is also discussed.

6.4.1 Reduction in null depths

Table 6.1 highlights the maximum power level that can be achieved using different antenna elements L along with the corresponding values of ϕ_{max}. As can be seen as L increass the value of ϕ_{max} decreases and the maximum power level at the deepest null that can be achieved also decreases. This is due to the fact that larger number of antenna elements (L) yields narrower width of the main and the side lobes. As the angular space between the lobes decreases, the amount of phase shift that can be used to decrease null depth also reduces yielding less reduction in null depths.

	L=6	L=6	L=9	L=9
Beam Parameter	**org.**	**mod.**	**org.**	**mod.**
Directivity	6	5.7	9	8
HPBW	17.2^o	18.2^o	11.4^o	12.7^o
ϕ	0	3.875^o	0	2.96^o
Deepest null	-100 dB	-24 dB	-75 dB	-24 dB

Table 6.2: Comparison of basic beam properties for the original (org.) and the modified (mod.) beams.

6.4.2 Comparison of the modified beam with the original beam

In this subsection numerical results that compare the behaviour of the modified beam with that of the original one in terms of HPBW and directivity are presented. The modified beams satisfying different minimum power thresholds were obtained using the algorithm described in section 6.2 to find the ϕ required to satisfy the desired minimum power threshold.

Table 6.2 compares the HPBWs and directivities of original and modified beams for $L = 6$ and 9 and a required power threshold of $P_{des,dB}$ = -24 dB. The value of ϕ for the modified beam obtained using the algorithm described in section two and the power values at the deepest nulls of both beams are also shown. Table 6.2 shows that the modified beam yields significant increases in the power at the deepest null at the cost of relatively small deterioration in HPBW and directivity. Figures 6.6 and 6.7 compare the normalized power of the original and modified beams for $L =$ 6 and 9 respectively and show the widening of the main lobe of the beam using the proposed technique. Figures 6.8 and 6.9 compare the directive gain using original and modified beams and show the loss of directivity in the modified beam compared to the original one for $L = 6$ and 9 respectively.

Figures 6.10 and 6.11 compare the HPBW and directivity of the original beam with that of the modified beam designed for varying minimum power thresholds and $L = 3$, 6, 9 and 12. For the modified beam, the value of ϕ each satisfying minimum

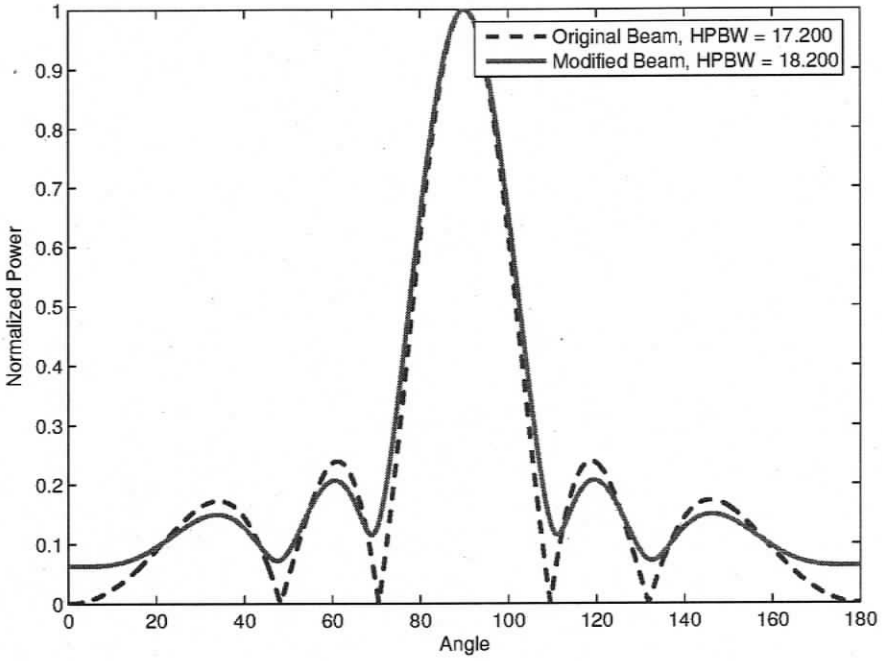


Figure 6.6: Comparison of the modified and original beam in terms of HPBW for $L = 6$ and power threshold = -24 dB

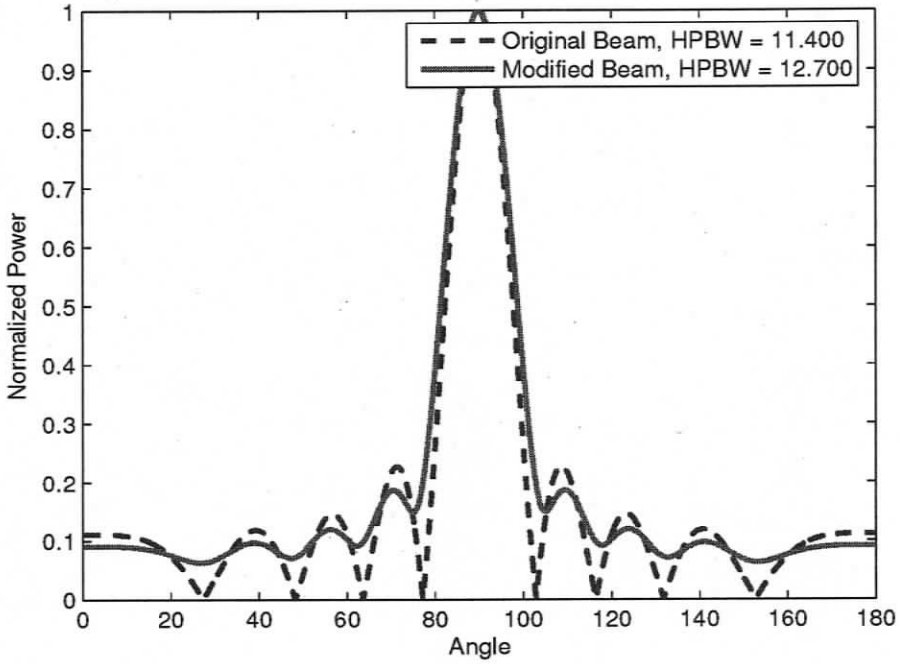


Figure 6.7: Comparison of the modified and original beam in terms of HPBW for $L = 9$ and power threshold = -24 dB

L	3	6	9	12
%age power savings (org.)	59.1	78.8	83.5	83.5
%age power savings (mod.)	58.9	77.1	82.0	83.0

Table 6.3: Percentage power savings that can be achieved for different number of antenna elements (L) using the original and modified beams.

power thresholds varying from -40 to -24 dB was obtained. As the required minimum power threshold is raised, the value of ϕ obtained increases and the directivity and HPBW experience slight degradation compared to the original beam. In Figure 6.11, the directivity of the original beam, as expected, remains constant at 3 for $L = 3$, 6 for $L = 6$, 9 for $L = 9$ and 12 for $L = 12$, while the directivity of the modified beam decreases with an increase in minimum required power threshold for all these cases. Similarly, in Figure 6.10, the HPBW of the original beam remains constant for all L while the modified beam shows an increase in HPBW with increased minimum power threshold requirement. The decrease in directivity and the increase in HPBW, however, are relatively small compared to the increase in minimum power achieved by the modified beam. As expected, an increase in number of antenna elements from $L = 3$ to $L = 12$ leads to increased directivity and reduced HPBW for both the modified and the original beams. The modified beam is a result of the phase shifted versions of the original beam and therefore it is expected to experience reduction in directive gain. This loss in directive gain is more prominent for increased number of antenna elements that yield higher directive gains. It was also found that the modified beam reduces the hidden beam problem significantly with little loss in power savings associated with the original beam as shown in Table 6.3 for different number of antenna elements.

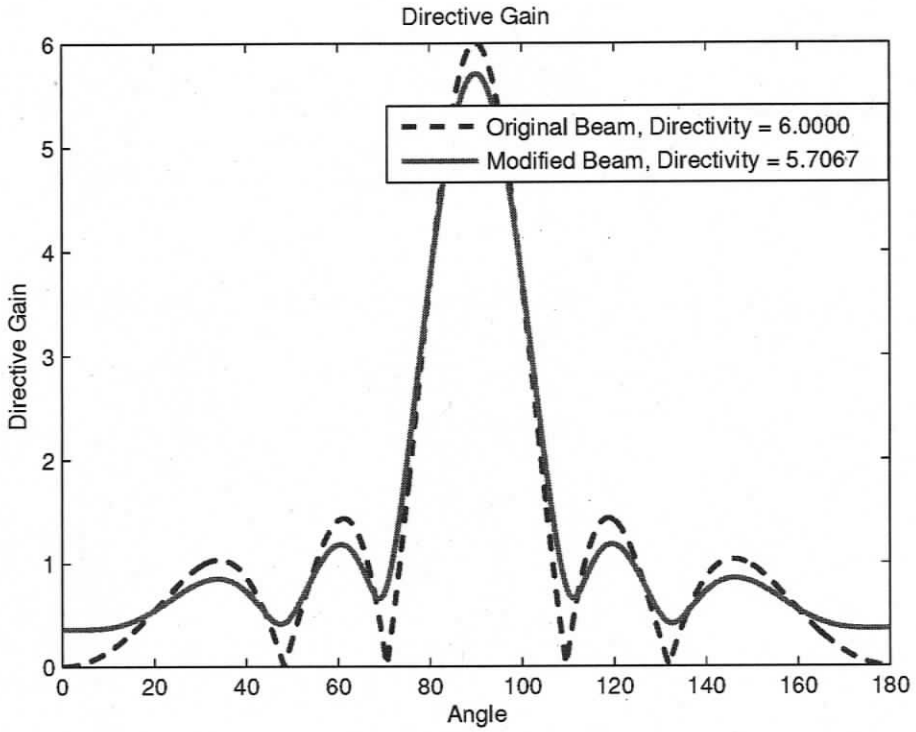


Figure 6.8: Comparison of the modified and original beam in terms of directive gain for $L = 6$ and power threshold = -24 dB

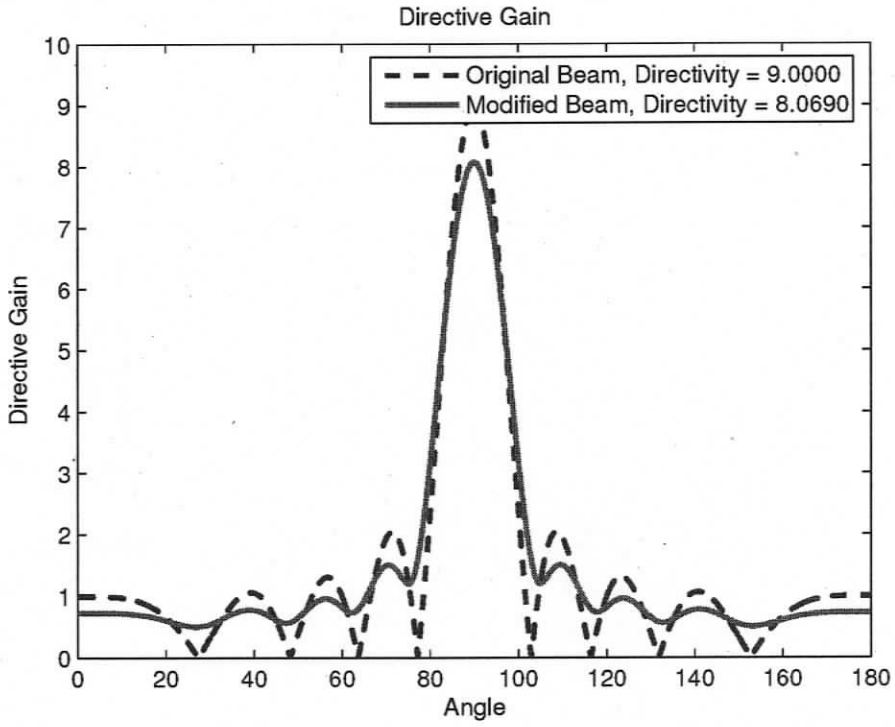


Figure 6.9: Comparison of the modified and original beam in terms of directive gain for $L = 9$ and power threshold = -24 dB

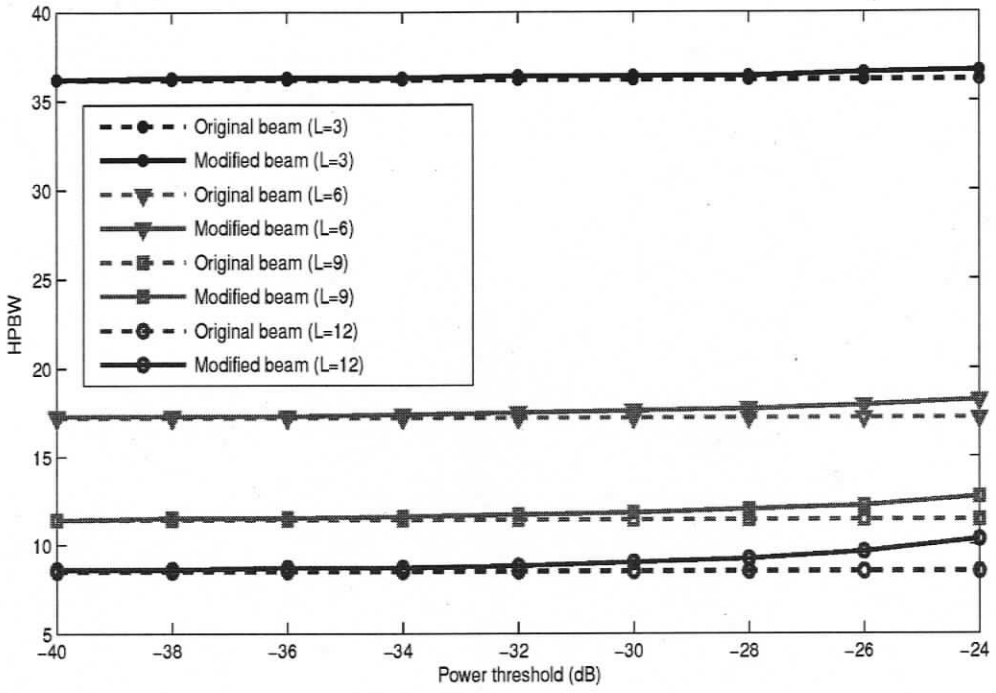


Figure 6.10: Comparison of original beam with modified beam in terms of HPBW for different power thresholds for $L = 3$, 6, 9, and 12.

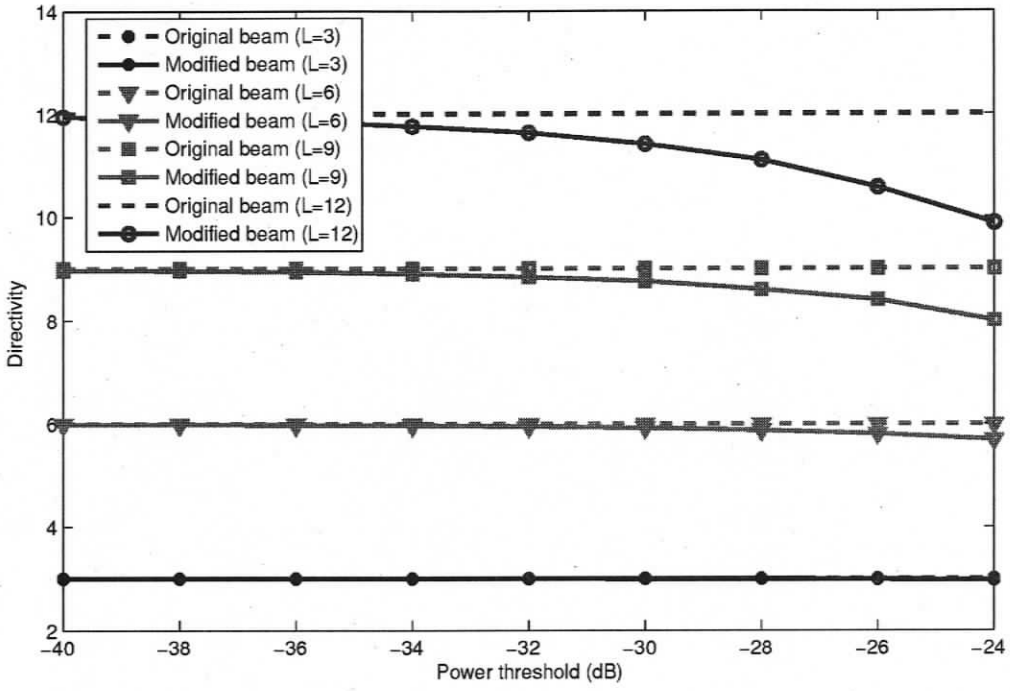


Figure 6.11: Comparison of original beam with modified beam in terms of Directivity for different power thresholds for $L = 3$, 6, 9, and 12.

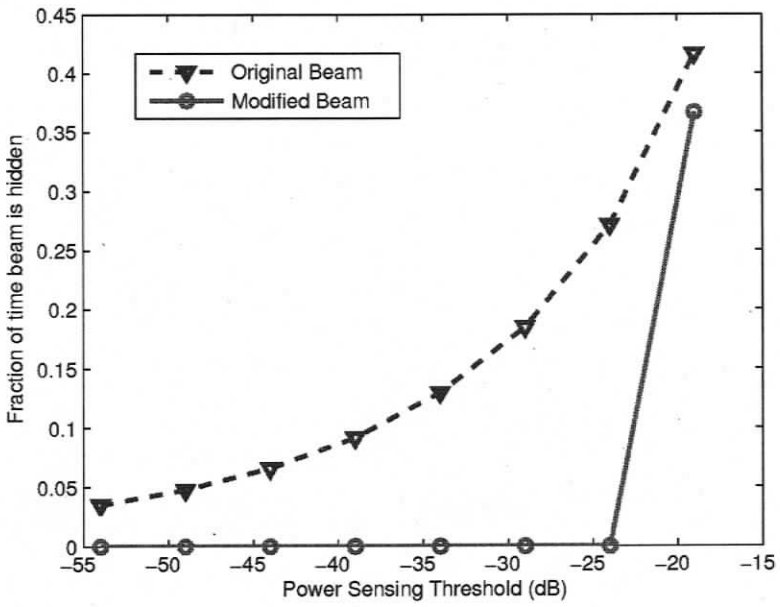


Figure 6.12: Comparison of original beam with modified beam in terms of η for different power thresholds for $L = 6$ and $\phi = 3.875^o$ (corresponding to highest power threshold of -24 dB)

6.4.3 Implementation of the modified beam in a non-persistent CSMA system

The removal of deep nulls in the modified beam implies that the fraction of angular space below a specified power threshold also decreases as mentioned above.

Figure 6.12 compares the original and the modified ($\phi = 3.875^o$) beams for $L = 6$ in terms of η for a specified power threshold that can be sensed by mobile terminals varying from -52 to -20 dB. As expected, the original beam leads to larger values of η implying increased fraction of time when the beam is hidden. The value of η for the modified beam is zero for power threshold upto -24 dB for $\phi = 3.875°$ but non-zero for a power threshold of -20 dB. However, as ϕ is increased to 5.2°, the power threshold of -18 dB can be achieved for $L = 6$ using the modified beam.

Figure 6.13 shows the throughput of the non-persistent CSMA system estimated

using Eqn. (6.22) for both the original and the modified beams for $L = 6$. The f_u:f_d ratio is assumed to be 1:1. Three different power thresholds of -30, -24 and -20 dB are considered. The fraction of time the beam is hidden, corresponding to these thresholds is found to be 0.17, 0.27 and 0.38 respectively for the original beam. For the modified beam the fraction of time beam is hidden is zero for all these thresholds and the resulting throughput is the same as that of a slotted non-persistent CSMA system with an omni-directional antenna. For comparison, the throughput of ALOHA given by $S = Ge^{-G}$ and slotted ALOHA systems expressed as $S = Ge^{-2G}$ are also shown. For the original beam, the hidden beam problem leads to a decrease in throughput and this decrease is intensified as the power sensing threshold increases. The use of the modified beam in a CSMA system thus yields power savings of a directional beam and the performance of an omni-directional beam.

6.5 Conclusions

A new beamforming technique that addresses the hidden beam problem in slotted non-persistent CSMA systems by modifying the weight vectors of a uniform linear antenna arrays is proposed. This is an example of a top down approach of a cross-layer design. The modified weight vector yields reduced null depths and lower fraction of time the beam is hidden. For a power sensing threshold of -24 dB and $L = 3$ to 12, the methodology used is successfully able to yield weighting vectors so that the desired power thresholds at all nulls are satisfied. This implies complete elimination of the hidden beam problem, exhibiting throughput for the slotted non-persistent CSMA that is the same as that obtained using omni-directional antenna and the power savings that are associated with the use of a directional antenna. The reduction in the null depth is, however, associated with a small increase in HPBW and some loss in directivity of the modified beam.

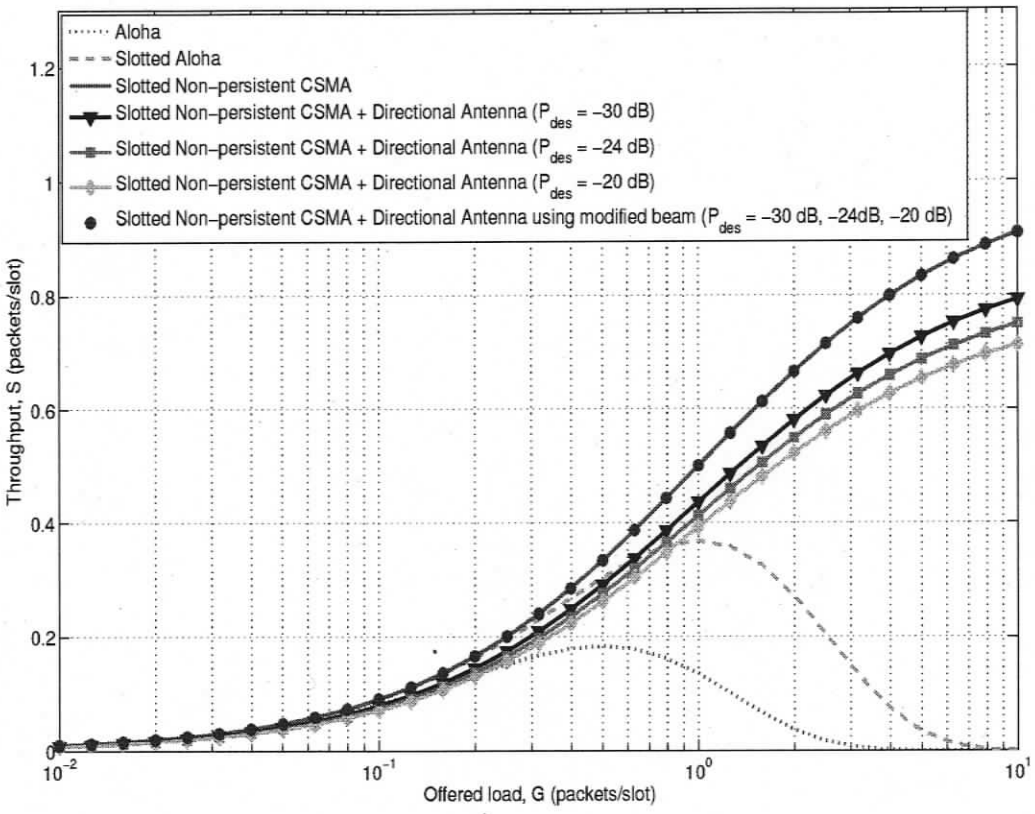


Figure 6.13: Comparison of original and modified beams in terms of throughput for non-persistent CSMA systems for minimal power sensing thresholds of -30 ($\phi = 2.80^o$), -24 ($\phi = 3.875^o$) and -20 dB ($\phi = 4.756^o$) for L=6.

Chapter 7

Conclusions and future work

The main objective of this dissertation was to investigate the effect of techniques that improve the performance of wireless networks. As mentioned earlier, wireless communications suffer from fading and interference. Therefore to mitigate the effect of fading and interference in order to improve the performance of wireless systems, a combination of antennas and antenna arrays is used at the physical (PHY) layer and is discussed in Chapter 3. The performance of wireless communications can also be improved within a cross-layered framework by designing scheduling algorithms at the medium access control (MAC) layer that take into account the PHY layer parameters. One such scheduling technique is proposed and discussed in detail in Chapter 4. Wireless channels are not always constant and can change from good to bad state (characterized by their signal to noise ratio below some threshold) instantly. Therefore to gain further insight into the performance of scheduling algorithms, the effect of channel availability was also taken into account. The effect of channel availability on the performance of and comparison between scheduling algorithms are presented in Chapter 5. Current wireless standards can support multiple applications including voice, video, data transmission and require both fixed and random channel assignment techniques depending on the application. However, when such systems are operating in packet switching mode using random assignment techniques based

on carrier sensing, the implementation of antenna arrays at the PHY layer can lead to severe performance degradation due to the hidden beam, deafness and hidden node problems. A new methodology to address the hidden beam problem, caused by the use of directional antennas, is presented in Chapter 6.

This chapter summarizes the main results and presents the conclusions drawn in the preceding chapters and suggestions are made for possible future work.

7.1 Conclusions

7.1.1 Role of antenna/antenna arrays

In Chapter 3, a combination of variable receive antennas in a STBC scheme and a minimum variance distortionless response beamforming algorithm with variable number of antenna array elements is used to investigate the relative merits and demerits of antennas and antenna array elements in an uplink environment. Simulation results showed that in a high noise and fading environment, diversity achieved by using space-time block coding is more efficient in reducing BER compared to beamforming while in the presence of high interference environment beamforming performs efficiently in terms of reducing BER. In addition, it was investigated how the role of antenna and antenna arrays is affected in varying noise and fading environments and by the number of interferring users and their angular locations. For practical applications, the number of antennas/antenna array elements that can be deployed at the PHY layer is more likely to be determined by the balance between the improvement in capacity (or reduction in BER) and other implementation issues including cost, available space and computational complexity.

7.1.2 Design of the combined SNR and angular separation based scheduling algorithm

In Chapter 4, a multiuser downlink system model for data transmission is considered where two out of K users are served in a given time slot which are chosen on the

basis of three scheduling algorithms (the RR, the greedy and the combined SNR and angular separation based algorithms) at the MAC layer. Results from both numerical and semi-analytical models indicated that taking into consideration both the angular separation between the users and their channel state information yields improvement in the system capacity. This is the case for the combined SNR and angular separation based scheduling algorithm that is shown to outperform both the greedy and the RR algorithms. Simulation results from the numerical and the semi-analytical model are also shown to be consistent. The improvement in system capacity obtained by additionally considering angular separation between mobile users, however, requires computation of the DoA of the users. A new computationally efficient DoA estimation technique [70] is now available that can be used in the proposed framework with relatively small additional computational expense.

7.1.3 Effect of channel availability on the performance of scheduling algorithms

In Chapter 5, the performance of three different scheduling algorithms (presented in Chapter 4) is evaluated using semi-analytical and numerical models for a cross-layered system within a queuing framework. The system model includes queues modelled as M/M/1 with exponentially distributed inter-arrival and inter-departure times (corresponding to the times when the channel is available), traffic load which depends on both arrivals and departures of packets in the system and the availability of CSI and DoA information at the physical layer. Results indicate that for low traffic, there is no significant difference in the performance between the three algorithms. For low traffic conditions, the fraction of time two users are served simultaneously is low and this leads to the little difference in the SINR for the greedy and combined SNR and angular separation based scheduling algorithms. For high traffic conditions, the combined SNR and angular separation based scheduling algorithm gives the highest SINR due to increased choice in selecting users for service, implying that taking into

account the angular location of mobile users can yield better performance. The round-robin scheduling algorithm that selects users unpreferrentially has the advantage of somewhat lower delay and packet loss compared to the other two algorithms but at the cost of lower SINR. The results show that designing a system model that takes into account channel availability and traffic conditions provides further insights into the behaviour of the scheduling algorithms.

7.1.4 A modified beamformer to solve hidden beam problem in wireless systems

A new beamforming technique that solves the hidden beam problem in slotted non-persistent CSMA systems by modifying the weight vectors of the uniform linear antenna arrays is presented in Chapter 6. The proposed beam yields lower probability of mobile terminals lying below a given power sensing threshold, and thus exhibits higher throughput. For power sensing thresholds of -24 dB and for $L = 3$ to 12, the proposed beam was successfully able to adjust the weighting vector so that the desired power thresholds at all nulls were satisfied. This implies complete elimination of the hidden beam problem, exhibiting throughput for the slotted non-persistent CSMA that was the same as that obtained using an omni-directional antenna and the power savings that are associated with the use of a directional antenna. The reduction in the null depth is, however, associated with a small increase in the HPBW and some loss in directivity of the proposed beam.

7.2 Future work

The work presented in this dissertation provides useful insights in addressing two important cross-layer issues i.e., serving multiple users by taking into account the effect of interference they cast on one another and addressing the hidden beam problem in CSMA systems that use beamforming at the PHY layer. There is also potential for further extension of this work and possible future research work is briefly mentioned

below.

7.2.1 Multiuser scheduling

In Chapter 4, the new scheduling algorithm designed to serve two users simultaneously can be extended to serve multiple users. The semi-analytical framework can also be extended for serving multiple users simultaneously. However, an increase in the number of simultaneously served users also results in an increase in interference that may require efficient algorithms for selecting these users.

7.2.2 Impact of traffic conditions

In Chapter 5, traffic was assumed to be Poisson distributed with inter-arrival times being exponentially distributed and an M/M/1 queue model was considered to account for channel availability. The system model presented in Chapter 5 can be generalized to use the G/G/1 model where G refers to general arrivals and departures. Further, the traffic pattern can be changed from Poisson to more realistic self-similar traffic to study its effect on the performance of the scheduling algorithms within a cross-layer system.

7.2.3 Addressing the hidden terminal and deafness problems

The modified beamformer presented in Chapter 6 may also be used to address the hidden terminal and deafness problem discussed in Chapter 6. Implementation of the proposed technique to solve these additional problems created by directional antennas when used in CSMA systems would constitute an interesting problem.

Bibliography

[1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, 1st ed. Cambridge University press, 2003.

[2] J. Litva and K. Titus, *Digital Beamforming in Wireless Communications*, 1st ed. Norwood, MA, USA: Artech House, Inc., 1996.

[3] H. VanTrees, *Optimum Array Processing (detection, estimation, and modulation theory, Part IV)*, 1st ed. Wiley Interscience, 2002.

[4] P. V. Rooyen, M. Lotter, and D. Wyk, *Space-Time Processing for CDMA Mobile Communications*, 1st ed. Springer, 2000.

[5] T. Rappaport, A. Annamalai, R. Buehrer, and W. Tranter, "Wireless communications: past events and a future perspective," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 148–161, May 2002.

[6] R. Laroia, S. Uppala, and J. Li, "Designing a mobile broadband wireless access network," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 20–28, Sept. 2004.

[7] V. Lau and Y.-K. Kwok, *Channel-Adaptive Technologies and Cross-Layer designs for Wireless Systems with Multiple Antennas: Theory and Applications*, 1st ed. Wiley Series in Telecommunications and Signal Processing, 2006.

[8] V. Tsibonis, L. Georgiadis, and L. Tassiulas, "Exploiting wireless channel state information for throughput maximization," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2566–2582, Nov. 2004.

[9] W. Stallings, *Data and Computer Communications*, 7th ed. Prentice hall, 2006.

[10] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112–119, Dec. 2005.

[11] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct 1998.

[12] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, Mar 1999.

[13] ——, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, Jul 1999.

[14] A. Paulraj and C. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49–83, Nov 1997.

[15] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, Jul 1996.

[16] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, April 1988.

[17] A. J. Paulraj, D. Gesbert, and C. Papadias, "Smart antennas for mobile communications," *Encyclopedia for Electrical Engineering*, vol. 35, no. 10, pp. 1365–1375, 1987.

[18] V. T. Raisinghani and S. Iyer, "Cross layer design optimizations in wireless protocol stack," *Elsevier Computer Communications Journal*, vol. 27, pp. 720–724, 2004.

[19] S. Shakkottai, T. Rappaport, and P. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine*, vol. 41, no. 10, pp. 74–80, Oct 2003.

[20] G. Xylomenos and G. C. Polyzos, "Quality of service support over multi-service wireless internet links," *International Journal of Computer and Telecommunication Networking*, vol. 37, no. 5, pp. 601–615, 2001.

[21] G. Dimic, N. Sidiropoulos, and R. Zhang, "Medium access control - physical cross-layer design," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 40–50, Sept. 2004.

[22] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 74–85, Jan. 2004.

[23] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746–1755, Sept. 2004.

[24] B. D. Noble, M. Satyanarayanam, D. Narayanan, J. E. Tilton, J. Flinn, and K. Walker, "Agile application aware adaptation for mobility," *ACM SIGOPS Operating Systems Review archive*, vol. 31, no. 5, pp. 276–287, 1997.

[25] TCP performance in wireless access with adaptive modulation and coding. Silicon. [Online]. Available: http://whitepapers.silicon.com/0,39024759,60303407p,00.htm

[26] L. Iannone, R. Khalili, K. Salamatian, and S. Fdida, "Cross-layer routing in wireless mesh networks," *1st International Symposium on Wireless Communication Systems*, pp. 319–323, Sept. 2004.

[27] A. Goldsmith, *Wireless Communications*, 1st ed. Cambridge University press, 2005.

[28] Wikipedia, the free encyclopedia. [Online]. Available: http://en.wikipedia.org/wiki

[29] Statistical characteristics of wireless network traffic and its impact on adhoc network performance. [Online]. Available: http://www.scs.org/getDoc

[30] E.-S. Jung and N. H. Vaidya, "A power control MAC protocol for ad hoc networks," *Wireless Networking Journal*, vol. 11, no. 1-2, pp. 55–66, 2005.

[31] J. Monks, V. Bharghavan, and W.-M. Hwu, "A power controlled multiple access protocol for wireless packet networks," *Proceedings of Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 1, pp. 219–228, 2001.

[32] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: a media access protocol for wireless LAN's," in *Proceedings of the conference on Communications architectures, protocols and applications.* ACM, 1994, pp. 212–225.

[33] P. Karn, "MACA: a new channel access method for packet radio," in *Proceedings of the conference on Communications architectures, protocols and applications*, 1990.

[34] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," *IEEE International Conference on Communications*, vol. 1, pp. 331–335, Jun 1995.

[35] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi, "CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70–77, Jul 2000.

[36] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 1st ed. Prentice hall, 1996.

[37] Forward link multiuser diversity through proportional fair scheduling. Bell Labs.

[38] M. Airy, S. Shakkattai, and J. Heath, R.W., "Spatially greedy scheduling in multi-user MIMO wireless systems," *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 982–986, Nov. 2003.

[39] D. Aktas and H. El Gamal, "Multiuser scheduling for multiple antenna systems," *IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, vol. 1, pp. 502–505, Aug. 2003.

[40] L. Dong, T. Li, and Y.-F. Huang, "Opportunistic transmission scheduling for multiuser MIMO systems," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. V–65–8, April 2003.

[41] R. Gozali, R. Buehrer, and B. Woerner, "The impact of multiuser diversity on space-time block coding," *IEEE Communications Letters*, vol. 7, no. 5, pp. 213–215, May 2003.

[42] T. Park, O.-S. Shin, and K. B. Lee, "Proportional fair scheduling for wireless communication with multiple transmit and receive antennas," *IEEE 58th Vehicular Technology Conference*, vol. 3, pp. 1573–1577, Oct. 2003.

[43] L. Berger, T. Kolding, J. Ramiro-Moreno, P. Ameigeiras, L. Schumacher, and P. Mogensen, "Interaction of transmit diversity and proportional fair schedul-

ing," *The 57th IEEE Semiannual Vehicular Technology Conference*, vol. 4, pp. 2423–2427, April 2003.

[44] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, Feb 2001.

[45] R. Srinivasan and J. Baras, "Understanding the trade-off between multiuser diversity gain and delay - an analytical approach," *IEEE 59th Vehicular Technology Conference*, vol. 5, pp. 2543–2547, May 2004.

[46] P. Liu, R. Berry, and M. Honig, "Delay-sensitive packet scheduling in wireless networks," *IEEE Wireless Communications and Networking*, vol. 3, pp. 1627–1632, March 2003.

[47] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," *IEEE Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 1084–1094, March-3 April 2003.

[48] C.-S. Hwang, K. Seong, and J. Cioffi, "Opportunistic p-persistent CSMA in wireless networks," *IEEE International Conference on Communications*, vol. 1, pp. 183–188, June 2006.

[49] B. Jose, H. Yin, P. Mehrotra, and E. Casas, "MAC layer issues and challenges of using smart antennas with 802.11," *IEEE 58th Vehicular Technology Conference*, vol. 5, pp. 3169–3173, Oct. 2003.

[50] A. Prabhu and S. Das, "Addressing deafness and hidden terninal problem in directional antenna based wireless multi-hop networks," *2nd International Conference on Communication Systems Software and Middleware*, pp. 1–6, Jan. 2007.

[51] V. Tarokh, Y.-S. Choi, and S. Alamouti, "Complementary beamforming," *IEEE 58th Vehicular Technology Conference*, vol. 5, pp. 3136–3140, Oct. 2003.

[52] Y.-S. Choi, S. Alamouti, and V. Tarokh, "Complementary beamforming: new approaches," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 41–50, Jan. 2006.

[53] V. Kawadia and P. Kumar, "A cautionary perspective on cross-layer design," *IEEE Wireless Communications*, vol. 12, no. 1, pp. 3–11, Feb. 2005.

[54] C. Barrett, A. Marathe, M. V. Marathe, and M. Drozda, "Characterizing the interaction between routing and MAC protocols in ad hoc networks," in *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*. ACM, 2002, pp. 92–103.

[55] S. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. Calderbank, "Great expectations: the value of spatial diversity in wireless networks," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 219–270, Feb 2004.

[56] S. Blostein and H. Leib, "Multiple antenna systems: their role and impact in future wireless access," *IEEE Communications Magazine*, vol. 41, no. 7, pp. 94–101, July 2003.

[57] R. Morelos-Zaragoza and M. Ghavami, "Combined beamforming and space-time block coding with a sparse array antenna," *The 5th International Symposium on Wireless Personal Multimedia Communications*, vol. 2, pp. 432–434, Oct. 2002.

[58] M. Katz and J. Ylitalo, "Extension of space-time coding to beamforming WCDMA base stations," *Proceedings of IEEE 51st Vehicular Technology Conference*, vol. 2, pp. 1230–1234, 2000.

[59] Z. Lei, F. Chin, and Y.-C. Liang, "Combined beamforming with space-time block coding for wireless downlink transmission," *Proceedings of IEEE 56th Vehicular Technology Conference*, vol. 4, pp. 2145–2148, 2002.

[60] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Combined array processing and space-time coding," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1121–1128, May 1999.

[61] L. lin Wang, S. xun Wang, X. ying Sun, and F. ye Hu, "Combined beamforming and space-time block coding for wireless communications," *IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 607–611, Sept. 2003.

[62] X. N. Tran, T. Taniguchi, and Y. Karasawa, "Adaptive beamforming for multiuser space-time block coded systems," *IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications*, pp. 745–749, Aug.-2 Sept. 2004.

[63] C. Sun, N. Karmakar, K. S. Lim, and A. Feng, "Combining beamforming with alamouti scheme for multiuser MIMO communications," *IEEE 60th Vehicular Technology Conference*, vol. 2, pp. 1415–1419, Sept. 2004.

[64] G. Bauch and J. Hagenauer, "Smart versus dumb antennas-capacities and FEC performance," *IEEE Communications Letters*, vol. 6, no. 2, pp. 55–57, Feb 2002.

[65] C. van Rensburg and B. Friedlander, "Transmit diversity for arrays in correlated rayleigh fading," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 6, pp. 1726–1734, Nov. 2004.

[66] B. Friedlander and S. Scherzer, "Beamforming versus transmit diversity in the downlink of a cellular communications system," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1023–1034, July 2004.

[67] J. Proakis, *Digital Communications*, 4th ed. McGraw Hill, 2001.

[68] R. T. Compton, *Adaptive Antennas: Concepts and Performance*, 1st ed. Prentice Hall, 1988.

[69] J. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels [mobile radio]," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, Nov 1991.

[70] N. Wang, P. Agathoklis, and A. Antoniou, "A new DOA estimation technique based on subarray beamforming," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3279–3290, Sept. 2006.

[71] L. C. Godara, "Applications of antenna arrays to mobile communications, part i: Performance improvement, feasibility, and system considerations," *Proceedings of the IEEE*, vol. 85, no. 7, pp. 1029–1030, Jul 1997.

[72] A. Seeger, M. Sikora, and W. Utschick, "Combined beamforming and scheduling for high speed downlink packet access," *IEEE Global Telecommunications Conference*, vol. 1, pp. 50–54, Dec. 2003.

[73] R. Pabst, J. Ellenbeck, M. Schinnenburg, and C. Hoymann, "System level performance of cellular WiMAX IEEE 802.16 with SDMA-enhanced medium access," *IEEE Wireless Communications and Networking Conference*, pp. 1820–1825, March 2007.

[74] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," *Proceedings of 25th IEEE International Conference on Computer Communications*, pp. 1–12, April 2006.

[75] O.-S. Shin and K. Bok, "Antenna-assisted round robin scheduling for MIMO cellular systems," *IEEE Communications Letters*, vol. 7, no. 3, pp. 109–111, March 2003.

[76] A. Kogiantis and L. Ozarow, "Downlink best-effort packet data with multiple antennas," *IEEE International Conference on Communications*, vol. 1, pp. 715–719, May 2003.

[77] M. R. Spiegel, *Schaum's Mathematical Handbook of Formulas and Tables*, 2nd ed. McGraw-Hill, 1998.

[78] H. A. David, *Order Statistics*, 2nd ed. Wiley-Interscience, 1981.

[79] P. P. Jr., *Probability, Random Variables, and Random Signal Principles*, 4th ed. McGraw-Hill Science/Engineering/Math, 2000.

[80] D. Arora and P. Agathoklis, "Capacity analysis for a multiuser cross-layer downlink model in the presence of fading and interference," *IET Communications*, vol. 1, no. 3, pp. 440–446, June 2007.

[81] Understanding wireless LAN performance trade offs. [Online]. Available: www.commsdesign.com

[82] F. Gebali, *Analysis of Computer and Communication Networks*, 1st ed. Springer, 2008.

[83] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466–478, Mar 1993.

[84] D. Arora and P. Agathoklis, "Multiuser scheduling for downlink in multi-antenna wireless systems," *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 1718–1721, May 2005.

[85] L. Kleinrock, *Queueing Systems. Volume 1: Theory*, 1st ed. Wiley-Interscience, 1975.

[86] C. Sakr and T. Todd, "Carrier-sense protocols for packet-switched smart antenna basestations," *Proceedings of International Conference on Network Protocols*, pp. 45–52, Oct 1997.

[87] T. ElBatt, T. Anderson, and B. Ryu, "Performance evaluation of multiple access protocols for ad hoc networks using directional antennas," *IEEE Wireless Communications and Networking conference*, vol. 2, pp. 982–987, March 2003.

[88] M. Sekido, M. Takata, M. Bandai, and T. Watanabe, "A directional hidden terminal problem in ad hoc network mac protocols with smart antennas and its solutions," *IEEE Global Telecommunications Conference*, vol. 5, pp. 2579–2583, Nov.-2 Dec. 2005.

[89] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part i–carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Transactions on Communications*, vol. 23, no. 12, Dec 1975.

Appendix A

Finding the location of the nulls

To find the nulls of the conventional beam (θ_n), the closed form expression of Eqn. (6.5) is simplified and equated to zero and is written as follows

$$F(\theta, \theta_0) = \frac{\sin(L\psi/2)}{L\sin(\psi/2)} = 0 \tag{A.1}$$

$$L\psi/2 = \pm n\pi \tag{A.2}$$

which on simplification yields the location of the nulls as shown below

$$L(\kappa d(\cos\theta_n - \cos\theta_0))/2 = \pm n\pi \tag{A.3}$$

$$L(\kappa d\cos\theta_n)/2 = \pm n\pi \text{ for } \theta_0 = 90^o \tag{A.4}$$

For $\kappa = 2\pi/\lambda$ and d $= \lambda/2$

$$\theta_n = cos^{-1}[\pm\frac{2n}{L}] \tag{A.5}$$

where $n = -L/2$ to $+L/2$ for even L and $n = -(L-1)/2$ to $+(L-1)/2$ for odd L, and $(n \neq 0)$.

Appendix B

Glossary

1. 2G: 2G (or 2-G) is short form for second-generation wireless telephone technology in which radio signals are digital. It can be divided into TDMA-based and CDMA-based standards depending on the type of multiplexing used. IS-54 and IS-136 are second-generation (2G) mobile phone systems, known as digital AMPS (D-AMPS).

2. 3G: 3G is the third generation of mobile phone standards and technology, superseding 2G. It is based on the international telecommunication union (ITU) family of standards under the international mobile telecommunications programme, "IMT-2000".

3. Additive white Gaussian noise (AWGN): AWGN channel model is one in which the only impairment is the linear addition of white noise with a constant spectral density and a Gaussian distribution of amplitude. AWGN is commonly used to simulate background noise of the channel under study.

4. AMPS: It is a first-generation cellular technology that uses separate frequencies, or "channels", for each conversation. It therefore requires considerable bandwidth for a large number of users and is no longer used.

5. Asymmetric Digital Subscriber Line (ADSL): ADSL is a form of DSL, a data communications technology that enables faster data transmission over copper telephone lines than a conventional voiceband modem can provide.

6. Automatic Repeat-request (ARQ): ARQ is an error control mechanism for data transmission.

7. B3G: B3G systems (often referred to as fourth generation or 4G) wireless communication systems will consist of a set of various networks using IP as a common protocol so as to access information anywhere, anytime, with a seamless connection to a wide range of information and services with various QoS requirements and at data bit rates of tens and hundreds of mbps.

8. Beamforming: Beamforming is a signal processing technique used for directional signal transmission or reception.

9. BER: Bit error ratio or bit error rate (BER) is the ratio of the number of bits incorrectly received to the total number of bits sent during a specified time interval.

10. Bluetooth: Bluetooth is an industrial specification for wireless personal area networks (PANs). Bluetooth provides a way to connect and exchange information between devices such as mobile phones, laptops, personal computers, printers, GPS receivers, digital cameras, and video game consoles over a secure, globally unlicensed short-range radio frequency.

11. BPSK: BPSK is the simplest form of PSK. It uses two phases which are separated by 180 and so can also be termed 2-PSK.

12. Carrier Sense Multiple Access (CSMA): CSMA is a probabilistic Media Access Control (MAC) protocol in which a node verifies the absence of other traffic before transmitting on a shared physical medium, such as an electrical bus, or a band of electromagnetic spectrum.

13. Carrier Sense Multiple Access With Collision Detection (CSMA/CD): CSMA/CD in computer networking, is a network control protocol in which a carrier sensing scheme is used and a transmitting data station that detects another signal while transmitting a frame, stops transmitting that frame, transmits a jam signal, and

then waits for a random time interval (known as "backoff delay" and determined using the truncated binary exponential backoff algorithm) before trying to send that frame again. CSMA/CD is a modification of pure Carrier Sense Multiple Access (CSMA).

14. Code division multiple access (CDMA): CDMA is a channel access method utilized by various radio communication technologies. CDMA employs spread-spectrum technology and a special coding scheme (where each transmitter is assigned a code) to allow multiple users to be multiplexed over the same physical channel.

15. CDMA2000: CDMA2000 is a hybrid 2.5G / 3G technology of mobile telecommunications standards that use CDMA, a multiple access scheme for digital radio, to send voice, data, and signalling data (such as a dialed telephone number) between mobile phones and cell sites.

16. Collision: In telecommunication, the collision occurs when two or more demands are made simultaneously on equipment that can handle only one at any given instant and in a computer, the situation that occurs when an attempt is made to store simultaneously two different data items at a given memory address that can hold only one of the items.

17. Channel state information (CSI): CSI is information about the current value of H, a mathematical value which represents a signal channel. It forms part of the signal model in Wireless communications.

18. Digital AMPS (D-AMPS): IS-54 and IS-136 are second-generation (2G) mobile phone systems, known as Digital AMPS. It is used throughout the America, particularly in the United States and Canada. D-AMPS is considered end-of-life, and existing networks are in the process of being replaced by GSM/GPRS and CDMA2000 technologies.

19. dB: The decibel (dB) is a logarithmic unit of measurement that expresses the magnitude of a physical quantity (usually power or intensity) relative to a specified or

implied reference level. Since it expresses a ratio of two (same unit) quantities, it is a dimensionless unit.

20. Digital Audio Broadcasting (DAB): DAB also known as Eureka 147, is a digital radio technology for broadcasting radio stations, used in several countries, particularly in Europe.

21. Enhanced Data rates for GSM Evolution (EDGE): EDGE is a digital mobile phone technology that allows increased data transmission rates and improved data transmission reliability. EDGE is generally classified as 2.75G and has been introduced into GSM networks around the world since 2003,

22. Evolution-Data Optimized: Evolution-Data Optimized or Evolution-Data only, abbreviated as EV-DO or EVDO and often EV, is a telecommunications standard for the wireless transmission of data through radio signals, typically for broadband Internet access. It uses multiplexing techniques including code division multiple access (CDMA) as well as time division multiple access (TDMA) to maximize both individual user's throughput and the overall system throughput.

23. Frequency Division Multiple Access (FDMA): FDMA is an access technology that is used by radio systems to share the radio spectrum.

24. General Packet Radio Service (GPRS): GPRS is a packet oriented mobile data service available to users of global system for mobile communications (GSM) and IS-136 mobile phones. It provides data rates from 56 up to 114 kbit/s.

25. GSM: Global System for Mobile communications is the most popular standard for mobile phones in the world operating in four different frequency ranges. Most GSM networks operate in the 900 MHz or 1800 MHz bands. Some countries in the America (including Canada and the United States) use the 850 MHz and 1900 MHz bands.

26. HDR: HDR is a Qualcomm wireless technology that provides high-speed, high-capacity IP packet data services that carriers can deploy in metropolitan areas. It

can provide up to a 2.4-Mbit/sec data transfer rate in a 1.25-MHz channel for use by stationary users, portable devices, and mobile services.

27. High-Speed Downlink Packet Access (HSDPA): HSDPA is a 3G (third generation) mobile telephony communications protocol in the high-speed packet access (HSPA) family, which allows networks based on universal mobile telecommunications system (UMTS) to have higher data transfer speeds and capacity. Current HSDPA deployments support down-link speeds of 1.8, 3.6, 7.2 and 14.4 Mbit/s.

28. IEEE 802.11: It is a set of standards for wireless local area network (WLAN) computer communication, developed by the IEEE LAN/MAN standards committee (IEEE 802) in the 5 GHz and 2.4 GHz public spectrum bands.

29. IEEE 802.11a: The 802.11a standard uses the same core protocol as the original standard, operates in 5 GHz band with a maximum raw data rate of 54 Mbit/s, which yields realistic net achievable throughput in the mid-20 Mbit/s.

30. 802.11b: 802.11b has a maximum raw data rate of 11 Mbit/s and uses the same media access method defined in the original standard. 802.11b products appeared on the market in early 2000, and is used in wireless LAN technology.

31. IEEE 802.11g: IEEE 802.11g works in the 2.4 GHz band (like 802.11b) but operates at a maximum raw data rate of 54 Mbit/s, or about 19 Mbit/s net throughput. 802.11g hardware is fully backwards compatible with 802.11b hardware.

32. IEEE 802.16: The IEEE 802.16 working group on broadband wireless access standards, which was established by IEEE standards board in 1999, aims to prepare formal specifications for the global deployment of broadband wireless metropolitan area networks.

33. Interim Standard 95 (IS-95): It is the first CDMA-based digital cellular standard pioneered by Qualcomm. The brand name for IS-95 is cdmaOne. It is a 2Gmobile telecommunications standard that uses CDMA, a multiple access scheme for digital radio, to send voice, data and signaling data between mobile telephones and cell sites.

34. ISI: In telecommunication, intersymbol interference (ISI) is a form of distortion of a signal in which one symbol interferes with subsequent symbols.

35. LAN: A local-area network (LAN) is a computer network covering a small geographic area, like a home, office, or group of buildings e.g. a school.

36. MAC: The media access control (MAC) data communication protocol sub-layer, also known as the medium access control, is a sublayer of the data link layer specified in the seven-layer OSI model (layer 2). It provides addressing and channel access control mechanisms that make it possible for several terminals or network nodes to communicate within a multipoint network, typically a local area network (LAN) or metropolitan area network (MAN).

37. MIMO: In radio, multiple-input and multiple-output, or MIMO (pronounced mee-moh or my-moh), is the use of multiple antennas at both the transmitter and receiver to improve communication performance.

38. Mobitex: Mobitex is an OSI based open standard, national public access wireless packet-switched data network. Mobitex puts great emphasis on safety and reliability with its use by military, police, firefighters and ambulance services. It is also used by the first model of Research in Motion's BlackBerry, and PDAs such as the Palm VII.

39. Multiple Access with Collision Avoidance for Wireless (MACAW): It is a slotted medium access control (MAC) protocol widely used in Ad-hoc networks.

40. Metropolitan area networks: Also known as MANs are large computer networks usually spanning a city. They typically use wireless infrastructure or Optical fiber connections to link their sites.

41. Network-assisted diversity multiple access (NDMA): It is a novel random medium access method, which provides impressive throughput efficiency for multiplexing variable-bit-rate data sources.

42. OSI: The open systems interconnection basic reference model (OSI Reference Model or OSI Model for short) is a layered, abstract description for communications and computer network protocol design.

43. Phase-shift keying (PSK): PSK is a digital modulation scheme that conveys data by changing, or modulating, the phase of a reference signal (the carrier wave).

44. Physical layer (PHY): It is level one in the seven-level OSI model of computer networking as well as in the five-layer TCP/IP reference model providing only the means of transmitting raw bits over a physical link connecting network nodes.

45. Quality of service (QoS): QoS is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow.

46. Radio frequency (RF): RF is a frequency or rate of oscillation within the range of about 3 Hz to 300 GHz. This range corresponds to frequency of alternating current electrical signals used to produce and detect radio waves.

47. Signal-to-noise ratio: SNR or S/N is defined as the ratio of a signal power to the noise power corrupting the signal.

48. Signal to Interference plus Noise Ratio (SINR): SINR is the ratio of the received strength of the desired signal to the received strength of undesired signals (noise and interference).

49. Space-time block coding (STBC): It is a technique used in wireless communications to transmit multiple copies of a data stream across a number of antennas and to exploit the various received versions of the data to improve the reliability of data-transfer.

50. Time division multiple access (TDMA): TDMA is a channel access method for shared medium (usually radio) networks. It allows several users to share the same frequency channel by dividing the signal into different timeslots.

51. Universal Mobile Telecommunications System (UMTS): UMTS is one of the third-generation (3G) cell phone technologies also sometimes marketed as 3GSM, emphasizing the combination of the 3G nature of the technology and the GSM standard which it was designed to succeed.

52. Voice over Internet Protocol (VoIP): VoIP is a protocol optimized for the transmission of voice through the Internet or other packet switched networks. VoIP is often used abstractly to refer to the actual transmission of voice (rather than the protocol implementing it).

53. Wide Area Network (WAN) is a computer network that covers a broad area (i.e., any network whose communications links cross metropolitan, regional, or national boundaries)

54. Wi-Fi: Wi-fi is the common name for a popular wireless technology used in home networks, mobile phones, video games and more. Wi-Fi is supported by nearly every modern personal computer operating system and most advanced game consoles.

55. WiMAX: WiMAX is the worldwide interoperability for microwave access, is a telecommunications technology aimed at providing wireless data over long distances in a variety of ways, from point-to-point links to full mobile cellular type access. It is based on the IEEE 802.16 standard, which is also called Wireless MAN.