University of Alberta

MIMO Relays for Increased Coverage and Capacity in Broadband Cellular Systems

by

Kevin Robert Jacobson

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Communications

Department of Electrical and Computer Engineering

Fall 2010
Edmonton, Alberta

Examining Committee

Witold A. Krzymień, Electrical and Computer Engineering

Abraham O. Fapojuwo, Electrical and Computer Engineering, University of Calgary

Mike MacGregor, Computing Science

Hai Jiang, Electrical and Computer Engineering

Yindi Jing, Electrical and Computer Engineering

To Mom, Dad, and Claudia
and Elvis and Ella.

Abstract

A significant challenge for fourth generation cellular systems is the reliable delivery of high speed (up to 1 gigabit per second) data to mobile or nomadic users throughout a cluttered urban environment. The wireless channel is a difficult channel over which to achieve high rate reliable communications. The wireless channel suffers many impairments such as small-scale multipath fading, shadowing, high path loss, co-channel interference, and Doppler shift due to mobility of the terminals and mobility in the propagation environment. Since radio spectrum is a scarce resource it is necessary to build cellular networks with high spectral efficiency. Two promising methods to solve this problem are multihop (MH) relaying and multiple-input multiple-output (MIMO) antenna techniques. The most difficult mobile users to serve reliably are those close to cell edges and those shadowed by large objects such as buildings. With MH relaying, a number of simple and inexpensive wireless relays are deployed throughout the cell to relay transmissions around obstacles and to reduce the path loss to distant mobile users. Also, MH relaying enables the deployment of small subcells throughout the cell, increasing the system's area averaged spectral efficiency. Various MIMO techniques can be used in scattering channels to increase capacity and reliability of data links in a wireless network. MH relaying and MIMO are key inclusions in emerging cellular standards such as IEEE 802.16 and LTE-Advanced, so it is necessary to study how these may be used jointly in a cellular environment.

We look at various techniques available in MH relaying and MIMO, and assess the benefits and difficulties of these techniques when used in cellular systems. We put together a realistic cellular system model, with typical cellular topologies and well-accepted propagation models, and assess the performance of a multihop MIMO system. We find that there are tradeoffs in using these techniques jointly since they provide gains by somewhat conflicting methods. MH relaying lowers path loss and mitigates scattering in the channel, while MIMO benefits from significant scattering. As a result, it is necessary to understand how to design a MH-MIMO network carefully in order to maximize the net benefit.

Acknowledgements

I would like to express my appreciation to my supervisor, Dr. Witold A. Krzymień, for his financial support, thorough document reviews, and expert guidance throughout my research program. I would like to thank my candidacy examination committee, Dr. Ivan Fair, Dr. Mike MacGregor, Dr. Hai Jiang, and committee chair Dr. Kambiz Moez, for their valuable input and guidance. I would also like to thank my final examination committee members, Dr. Abraham O. Fapojuwo, Dr. Mike MacGregor, Dr. Hai Jiang, and Dr. Yindi Jing and committee chair Dr. Ashwin Iyer for their valuable input on my thesis.

The research for this thesis was generously funded by Telecommunication Research Labs (TRLabs), the Natural Sciences and Engineering Research Council (NSERC) of Canada, the informatics Circle of Research Excellence (iCORE), the Rohit Sharma professorship, the University of Alberta, and the Engineers Canada/Manulife Financial Scholarship. I am grateful for their support.

I have appreciated the many academic and casual discussions and encounters with the students I have met at the University of Alberta, in particular, Dr. Boon-Chin Lim, Dr. Shreeram Sigdel, Dr. Mohsen Eslami, and Robert Elliot.

I would like to thank my parents, Lesia and Ken Jacobson for guiding me through stages of life, and patiently supporting me in all of my varied interests, whether academic, scientific, artistic or otherwise.

Last but not least I have greatly appreciated the unfailing support from my wife, Claudia, who has stood by and encouraged me for many years.

Contents

List of Tables

List of Figures

List of Abbreviations

List of Abbreviations

Acronym	Meaning
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
AASE	Area averaged spectral efficiency
AF	Amplify and forward
AMC	Adaptive modulation and coding
ARQ	Automatic repeat request
ARS	Ad hoc relay station
BE	Best effort
BER	Bit error rate
BPSK	Binary phase shift keying
b/s	Bits per second
BS	Base Station
CB	Coordinated beamforming
CCDF	Complementary cumulative distribution function
CDF	Cumulative distribution function
CDMA	Code division multiple access
CIR	Carrier to interference ratio
CoMP	Coordinated multipoint transmission/reception
CR	Cooperative relaying
CS	Coordinated scheduling
cm	Centimeters
CSI	Channel state information
dB	decibel
dBm	decibel relative to 1 milliwatt
DDF	Dynamic decode and forward
DF	Decode and forward
DL	Downlink
DMT	Diversity-multiplexing tradeoff
DPC	Dirty paper coding
DSR	Digital selective relaying
FEC	Forward error correction
FFT	Fast Fourier transform
FTP	File transfer protocol
Gb/s	Gigabits per second
GHz	Gigahertz
GPC	Generalized power control

Continued on Next Page...

Acronym	Meaning
HiPERMAN	High performance metropolitan area network
HTTP	Hypertext transfer protocol
Hz	Hertz
IEEE	Institute of Electrical and Electronics Engineers
iid	Independent and identically distributed
IMT	International mobile telecommunications
ITU-R	International Telecommunication Union - Radiocommunication Sector
JT	Joint transmission
km/h	Kilometers per hour
LOS	Line-of-sight
LTE	Long term evolution
MAC	Medium access control (when referring to protocol layer)
Mb/s	Megabits per second
MCN	Multihop cellular network
MCS	Modulation and coding scheme
MH	Multihop
MH-MIMO	Multihop multiple-input multiple-output
MHz	Megahertz
MIMO	Multiple-input multiple-output
MPEG	Motion picture experts group
MR	Multihop relaying (IEEE 802.16j)
MRC	Maximal ratio combining
MS	Mobile station
ms	Millsecond
MU	Multiple user
MU-MIMO	Multiple user multiple-input multiple-output
μs	Microsecond
NAF	Non-orthogonal amplify and forward
NLOS	Non-line-of-sight
nrtPS	Non-real time polling service
OAF	Orthogonal amplify and forward
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
PDF	Probability density function
PER	Packet error rate
PHY	Physical (layer)
QAM	Quadrature amplitude modulation
QoS	Quality of service
RF	Radio frequency
RS	Relay station
RTG	Receive/transmit transition gap
rtPS	Real time polling service
SAS	Single antenna selection
SC	Selective combining
SC	Simulcast
SINR	Signal to interference plus noise ratio
SNR	Signal to noise ratio
SR	Spatial reuse
SS	Subscriber station (IEEE 802.16-2004)

Continued on Next Page...

List of Abbreviations – Continued

Acronym	**Meaning**
SSRTG	Subscriber station receive/transmit transition gap
SSTTG	Subscriber station transmit/receive transition gap
ST	Selective transmit
TDD	Time division duplexing
TDMA	Time division multiple access
TTG	Transmit/receive transition gap
UGS	Unsolicited grant service
UL	Uplink
VAA	Virtual antenna array
VoIP	Voice over Internet Protocol
WLAN	Wireless local area network
ZMCSCG	Zero mean circularly symmetric complex Gaussian

Mathematical Notation

Notation	Meaning
a	Scalar
$\vec{a}$	Vector
$\mathbf{A}$	Matrix
$\min(a, b)$	Minimum value of a and b
$\max(a, b)$	Maximum value of a and b
$\mathbf{A}^T$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian (conjugate transpose) of matrix $\mathbf{A}$
$\mathcal{E}[\star]_A$	Expectation operation with respect to variable A
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbf{I}_N$	The $N \times N$ identity matrix
$\det\{\mathbf{A}\}$	Determinant of the matrix $\mathbf{A}$
$\|\mathbf{A}\|_F$	Frobenius norm of the matrix $\mathbf{A}$
$(m)_n$	Pochhammer symbol
$tr_j(\mathbf{A})$	j^{th} elementary symmetric function of $\mathbf{A}$
$(x)^+$	Maximum of 0 and x
$\lceil x \rceil$	Smallest integer greater than or equal to x

Chapter 1

Introduction

1.1 Cellular System Evolution

The first cellular systems appeared in the late 1970s and provided only voice service. As systems developed, packet data transmission capability was added, but only low rates were possible. Clearly the demand for mobile communications services in Canada and worldwide has increased dramatically, with about a 70% penetration rate in Canada and 91% penetration rate in the US. The penetration of smartphones with data capability, such as the Blackberry and iPhone, doubled to about 21% in Canada in 2008 and the market is quickly growing. As a result, demand for high speed mobile data has grown significantly and will accelerate in coming years.

Current cellular systems, called third generation (3G) systems, have begun to address the demands of data users but the fourth generation, 4G, will provide a significant improvement. The International Telecommunication Union's Radiocommunication Sector (ITU-R) has developed a set of requirements and evaluation criteria for 4G systems in the International Mobile Telecommunications-Advanced (IMT-Advanced) documentation [13–16]. 4G cellular networks are expected to provide peak data rates of up to 1 Gb/s, with high reliability over a wide area. Bit rates of up to 100 Mb/s should be maintained at speeds up to 120 km/h. This represents a difficult technical challenge. First of all, radio spectrum is scarce. There is only a limited radio spectrum available and it must be used efficiently. According to a recent presentation by Alcatel-Lucent [17], a cell capacity limit is looming. Even with aggressive spectrum reallocation, it is expected that a capacity limit of 350 Mb/s/cell will be reached by 2016. Second, the wireless communications channel has numerous impairments. Multipath propagation of radio waves causes the channel to fade randomly as multiple scattered, reflected and diffracted signals arrive at the receiver. Movement of the user handset and scatterers causes Doppler shift of the carrier frequency. These random effects cause *small scale signal fading*. Large objects such as buildings and geographical features cause larger scale random fading due to *shadowing*. The signal also experiences significant *path loss* due

to large distances between the transmitter and receiver. And finally, signal reception suffers from co-channel interference from other user signals since the cellular system must reuse spectrum. Mobile users at the edge of the cell suffer especially from poor service because they are the furthest from any base station, and there may be pockets of poor coverage in a cell due to shadowing.

There are a large number of advanced techniques available and under development to address these difficulties. Two general approaches of key relevance to the viability of future broadband spectrally efficient cellular systems, various elements of which are under development by numerous research groups are: multiple antenna techniques and multihop relaying. Two proposals have been submitted in October, 2009 to the ITU-R, one by the Institute of Electrical and Electronics Engineers (IEEE) 802.16 standards working group [18] and one by the Third Generation Partnership Project (3GPP) [19]. Multiple antenna techniques and multihop relaying are key elements in both of these proposals. The benefits of multihop relaying have been recognized early in the development of 802.16 [20], which resulted in [21] issued in 2009. These developments occurred in parallel with the research for this thesis.

Brief descriptions of multiple antenna techniques and multihop relaying are given in the remainder of this chapter, and more detailed explanations of the techniques and performance measures are given in Chapter 2.

1.2 Wireless Relaying

Wireless relaying is a fairly old technique of improving the performance of a wireless link. Simple analogue repeaters have been used in the past to receive, regenerate and retransmit analogue signals over great distances. The relaying method can be extended significantly in modern data communications and so there has been a resurgence in interest in recent years. A simple relay channel with one source, one relay and one destination has been studied from an information theory point of view in [22]. For many years, the capacity of the relay channel had only been determined for a simple relaying system, however some very recent research [23, 24] has given some bounds on capacities for the multiple relay case. In a significant development, the throughput for an ad hoc mesh network of n terminals was found [25, 26]. This gives the scaling properties of throughput for large ad hoc networks using some simple models for the propagation channel, but the results are not necessarily applicable to a real world multihop cellular network.

There are two main areas of research in relaying: cooperative relaying and multihop relaying.

Cooperative relaying provides diverse signal paths which when combined create a much more reliable data link. Due to random superposition at the receiver of multiple replicas of the transmitted signal resulting from multipath propagation, the received signal in a wireless

channel suffers from small-scale random fading which results in unreliable data reception. It can be shown [27] that the bit error rate (BER) of such a channel (a Rayleigh channel) falls off linearly with signal to noise ratio (SNR). This BER performance is far worse than in a Gaussian channel in which the BER falls off exponentially with SNR. However, if a number of independently fading signal paths are available, it is much less likely that all paths are in a fade simultaneously. Combining these multiple uncorrelated received fading signals in an intelligent manner improves the data link reliability so that it can approach the BER performance of a non-fading Gaussian channel. This technique is known as *diversity combining*. A significant amount of work has been done in cooperative relaying [28–32] which shows great performance improvement in a Rayleigh channel. However, diversity techniques are less useful in Ricean channels, which are relatively common in cellular systems [27]. In addition, other forms of diversity are used in cellular and multiple antenna systems, which reduces the benefit of cooperative relaying. Hence this research will focus primarily on non-cooperative multihop relaying.

Significant loss in signal power occurs when transmitting wireless signals over long distances. Multihop relaying can significantly reduce the loss in cellular systems by placing wireless relays in between the transmitter and receiver. Careful design of relaying in a cellular system can avoid shadowing obstacles like buildings and geographical features, which can reduce the fading component of the channel. Cell coverage can be extended and improved, and data transmission reliability can be increased. Multihop relaying can especially help improve throughput and reliability to mobile users at the cell edges or in heavily shadowed areas. The deployment of relays in a system provides the system designer degrees of freedom in relay placement and use. Since a large number of relays may be necessary to provide significant benefit, these relays should be relatively cheap compared to the base station, and thus less complex in design.

It has been observed that a cellular capacity wall of 350 Mb/s/cell [17] is on the horizon. Therefore it is necessary to use smaller cells in order to achieve a higher spectral efficiency over an area (i.e. higher b/s/Hz/km^2). One method of achieving this is to divide the larger cell, typically 1 to 2 kilometers in radius, into smaller subcells in which relay stations (RSs) serve mobile stations (MSs) closest to them. Numerous researchers have looked at the various approaches to MH relaying in cellular systems [33–38]. IMT- Advanced is a set of requirements for future broadband cellular systems set by the ITU [13–15]. Two system proposals under consideration for 4G IMT-Advanced, IEEE 802.16m [21, 39] and Long Term Evolution (LTE) Advanced [19, 40], will include relaying as options. Clearly, relaying requires more complicated system level algorithms (medium access control - MAC - layer and higher) in order to achieve good results in a network of wireless stations. Also, MH relaying requires additional system resources (time or frequency slots), and hence the

spectral efficiency (measured in b/s/Hz) may suffer unless careful system design is done. The use of spatial reuse is a key method for increasing area averaged spectral efficiency (AASE).

1.3 Multiple Antenna Techniques

In general, there are several ways in which multiple antennas at the base station and at mobile users can be used. With transmit-side classical beamforming, signals at the different antennas are weighted in order to steer the signal towards the intended receiver and steer nulls towards other users. This strengthens the signal to the intended receiver and minimizes interference to other users. A similar signal-phasing approach can be used at the receive end in order to increase the received signal level (receive array gain). Classical beamforming is relatively simple and works well in line-of-sight (LOS) channels with knowledge of the direction to the receivers. Beamforming can also be done in a non-line of sight (NLOS) environment typical in cluttered urban areas. In NLOS, no direct specular path exists between the transmitter and receiver, but there may exist a dominant eigenmode in the matrix channel at a particular time. This represents a strong eigen-channel to one particular user. Using a rank one input covariance matrix, the transmitter projects that user's signal along the dominant eigenmode. Under certain conditions [41], this strategy can achieve capacity.

Multiple-input multiple-output (MIMO) techniques use multiple antennas at transmitting and receiving ends of a data link in order to create degrees of freedom that can be exploited to improve performance of a wireless data link. There has been a significant amount of research on MIMO in the past decade. Key papers [42, 43] have shown theoretically that rich scattering in wireless channels can be exploited to obtain significant improvements in capacity and reliability of wireless data links. It was shown that the Shannon capacity of a MIMO channel scales with $\min(N_T, N_R)$ (under certain conditions) where N_T is the number of transmit antennas at the source and N_R is the number of receive antennas at the destination. It was previously believed that the capacity of a channel could only be increased by increasing the channel bandwidth or increasing the transmit power.

With a diversity approach, the degrees of freedom in the channel are used to form multiple transmission branches. As described earlier, if branches fade independently of one another, the reliability of the data link is improved since it is much less likely that all branches suffer a fade condition simultaneously.

A very promising newer technique is spatial multiplexing. With this technique, data is encoded for the different antennas in order to form numerous spatial substreams. Data is carried on each of these substreams which greatly increases the data carrying capacity of the link. An uncorrelated rich scattering channel is required for this method to work.

In a multiuser system, different user channels will be in fades at different times. With a large pool of users, the system can schedule transmissions so that data is sent to users whose channels are in good condition only rather than wasting resources on users whose channels are poor. This provides multi-user diversity gain. Again a rich scattering channel is required for this method to work. It can be shown that the multiuser sum capacity of a spatially multiplexed MIMO system can be approached by transmitting to several users at a time. In the downlink direction, complex precoding is done at the base station and the spatial layers carrying independent data streams are received at various mobile receivers. The receivers cannot cooperate since they are not colocated, and so the layers must be separated by a suitable precoding technique at the base station. The optimal precoding technique is dirty paper coding (DPC) [44], which encodes signals in such a manner that multiuser interference is cancelled. However, DPC is complex and requires non-causal knowledge of the user signals. Simpler and sub-optimal precoding methods are required for implementation in practical systems. A large number of users exist in a cellular system and they cannot be simultaneously served by the limited number of substreams available. Hence a subset of users must be chosen for service at a given time. User selection and scheduling may be accomplished by an exhaustive search, but this becomes unwieldy with a large number of users. Numerous sub-optimal user scheduling techniques have been investigated by various researchers.

Both diversity and spatial multiplexing can be used simultaneously which results in a diversity-multiplexing tradeoff [45, 46].

A very important application of MIMO to cellular networks is Coordinated Multipoint (CoMP) transmission/reception in which base stations in different cells cooperate in the delivery of data to mobiles [47, 48]. Three key methods are possible: joint transmission (JT), coordinated beamforming (CB), and coordinated scheduling (CS). With JT, user data streams are jointly encoded at the coordinating base stations for transmission to several mobiles. In this technique, the multiple base stations act as one MIMO array. This is very complicated since user data and channel state information must be available at all the coordinating base stations, and optimal processing is extremely complex. In CB [49], multiple base station jointly optimize their beamforming vectors in order to minimize inter-cell interference. In this case, the serving base station requires knowledge of the user data, user channel, and the interference channels. With CS, the transmission of data destined for various mobiles is scheduled in time and frequency in order to minimize interference. CS is is the simplest, requiring only scheduling information to be shared across coordinating base stations.

With these different approaches available to a system designer, it is necessary to understand the benefits and costs of each as implemented in a realistic cellular system. In

addition, it is necessary to study how well these techniques work in typical cellular propagation environments.

1.4 Relaying With Multiple Antennas

Since relaying and MIMO are two techniques very likely to be used in 4G systems, it is necessary to understand how they work jointly. At the outset of this research there was very little work in the literature on the topic, but more has appeared in the last few years. Both of these techniques can be complicated to model mathematically, and it is difficult to find closed form expressions for their performance. Thus there remains very little theoretical analysis on the joint application of the two approaches. Previously published work will be briefly reviewed in Chapter 2.

Relaying and MIMO both can improve transmission in the wireless channel, but they accomplish this using different means. MIMO takes advantage of the signal scattering in the channel, and it has been shown that great gains can be obtained as compared to a non-scattering channel at the same SNR. Multihop relaying improves transmission by lowering the path loss which increases the SNR. It is important to recognize that high scattering in a wireless channel also means high path loss. So correct analysis of the combination of the techniques in a given cellular system geometry requires a system model that ties together the scattering effect and the path loss as functions of distance.

1.5 Research Objectives

As mentioned earlier, 4G cellular systems will include both relaying and MIMO, but little research has been done on their combined use. It seems natural to combine MIMO and relaying techniques in order to improve the performance of a cellular system, but it is necessary to determine how well they work together and what tradeoffs exist in combining them. In addition, it is necessary to use a system model that captures the radio frequency (RF) propagation of signals in a realistic cellular system.

Our work studies a cellular system combining decode and forward (DF) MH relaying with multiple antenna techniques with the goal of achieving higher data carrying capacity simultaneously with good system coverage. The primary objective of the research is to investigate how MH relaying and MIMO techniques may be used jointly in order to provide a cost effective, highly reliable, high capacity cellular system that can be realistically implemented. The MIMO field has developed very rapidly in the past decade and so we do not expect to find significant advances in that area. Relaying had some early interest, but only some results for much simplified cases have emerged from information theory to date. We analyzed the performance of MH-MIMO using coverage, network throughput, and link

reliability for cell edge mobile users as performance measures. Overall design guidelines for combined MH-MIMO cellular networks were determined.

During the course of this research we found that determining closed form expressions for the capacity of the relay channel, and particularly for the MH-MIMO channel is a very difficult and as-yet unsolved problem. Thus we have focussed on simulating and calculating system performance using realistic cellular environments, with realistic parameters and channel models recommended in emerging standards such as 802.16 [20] and 3rd Generation Partnership Project (3GPP) [1]. In particular, the measure of success of MIMO combined with MH relaying depends greatly on the physical environment in which the system operates. We have considered typical urban scenarios, at first analyzing a one-dimensional system and then looking at two-dimensional cellular systems with both hexagonal and Manhattan topologies.

There are numerous options and variations in both MIMO and relaying techniques, so our main objectives in this research are

- Study the key aspects of MIMO and determine the most practical alternatives for a cellular system.
- Similarly, we will look at the various relaying options and their performance, and determine the most practical variation for a cellular system.
- Find out how MH relaying and MIMO may complement each other in a typical cellular system, with the goal of improving aggregate rate per cell and cell coverage.

1.6 Methodology

A major difficulty in this area of research is that closed form expressions for capacity of the relay channel have remained unknown for decades. In addition, capacity achieving techniques are currently impractical and sub-optimal MIMO techniques generally require significant processing time to find capacity-approaching solutions. So combining these two methods results in some difficulty in calculating results. As a result, the research was carried out by mathematical modelling and computer simulation. System models incorporating parameters and methods established by such groups as the Third Generation Partnership Project (3GPP) [1] and other sources have been developed and extended as necessary. When random variables are involved performance is evaluated using Monte Carlo methods. Performance of systems is characterized using normalized aggregate network data rate (b/s/Hz) and AASE (b/s/Hz/km^2).

We often use 802.16 [20] with 802.16j [21] as a base network for much of the work since this standard was the first to consider relaying. Some of our results are based on the earlier

and superseded 802.16-2004 since that was most current at the time much of this research was completed, and our work preceded full incorporation of relaying into 802.16.

1.7 Contributions

Here we summarize the main original contributions of our work.

- The development of a simple algorithm for multiple relay selection which simultaneously considers the best paths to the target mobile and the impact of interference to other mobiles [50]. This is described in Section 4.4.
- The development and use of a realistic dual-slope Ricean model for single antenna multihop relaying in cellular systems, and the study of network throughput and throughput regions [51–53]. The model is described in Chapter 3 and results are given in Chapter 4.
- We have shown that the design of relaying networks must be done with suitable consideration of the specific geography and propagation environment, and we have provided some design guidelines in Section 5.3.
- We have extended the above work to multihop MIMO [54–56] (Chapter 5).
- Using the dual slope Ricean model in fixed geometry cellular layouts, we have shown that there is a tradeoff in using multihop relaying and MIMO simultaneously. In prior work, MIMO was shown to perform much better in a richly scattering environment than in a low-scattering scattering environment at the same SNR. However, a key fact to note is that a richly scattering environment also suffers from high loss, so proper comparison in a realistic system must tie these two propagation characteristics together. It has been shown that lowering path loss due to MH relaying gives a significant benefit, and although MIMO gains suffer there remain benefits in using the two techniques jointly. These results are shown in Chapter 5.
- In order to further justify our choice of multihop relaying over cooperative relaying, we have made a comparison of cooperative and multihop relaying approaches using a dual-slope Ricean model in Section 4.6 [57].

During the course of our work, we followed the developments emerging from IEEE 802.16 and 3GPP's LTE-A. Our group was contracted by a major wireless network operator to report on developments in multihop relaying and MIMO. Five full length confidential annual reports on the topic were delivered to the operator.

1.8 Organization of the Thesis

This thesis is organized as follows. Chapter 2 gives some background on MIMO and multihop relaying systems, some discussion of the important performance measures, overview of relevant prior work, and brief discussion of the wireless standards that include MIMO and relaying. Chapter 3 outlines the system models for single antenna and multiple antenna multihop relaying. Chapter 4 discusses multihop relaying with single antennas. This chapter looks at the throughput of relaying in typical cellular system geometries. Interference avoidance and delay are also addressed. Chapter 5 discusses multihop relaying and multiple antennas, and presents some results for both one dimensional and two dimensional cellular systems. Finally, Chapter 6 presents some overall conclusions from our work, and discusses some possible future work.

Chapter 2

Background on MIMO and Relaying

In this chapter we cover some background concepts and refer to previous work in multiple-input multiple-output (MIMO) and relaying. Both methods have numerous variations, and it is necessary to determine configurations that are most suitable for a practical cellular system. We discuss the main variations and their benefits and disadvantages.

2.1 Multiple-Input Multiple-Output (MIMO) Antenna Systems

2.1.1 Overview of MIMO

Multiple-input multiple-output (MIMO) antenna systems have received a significant amount of attention due to the promise of achieving large data capacities in wireless channels [42, 43]. A major impairment in the wireless channel is multi-path propagation of the radio signal causing the spreading of pulses and random signal fades at the receiver. Multiple antenna techniques can take advantage of richly scattered signals to improve the throughput and reliability of a wireless communication link. The standard MIMO model in Fig. 2.1 [42, 43, 58] is used in our work. Signals launched from N_T transmit antennas propagate through the channel and are received by N_R receive antennas. A matrix of complex channel gains, $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$, describes the signal propagation from each transmit antenna to each receive antenna.

Elements of $\mathbf{H}$ are modelled by a random variable that captures the stochastic nature of the wireless channel. We wish to model both line of sight (LOS) and non-line of sight (NLOS) conditions, and so we express the channel matrix (normalized) as a sum of two components [58]:

$$\mathbf{H} = \sqrt{\frac{K_r}{1+K_r}}\mathbf{H}_{LOS} + \sqrt{\frac{1}{1+K_r}}\mathbf{H}_{NLOS} \tag{2.1}$$

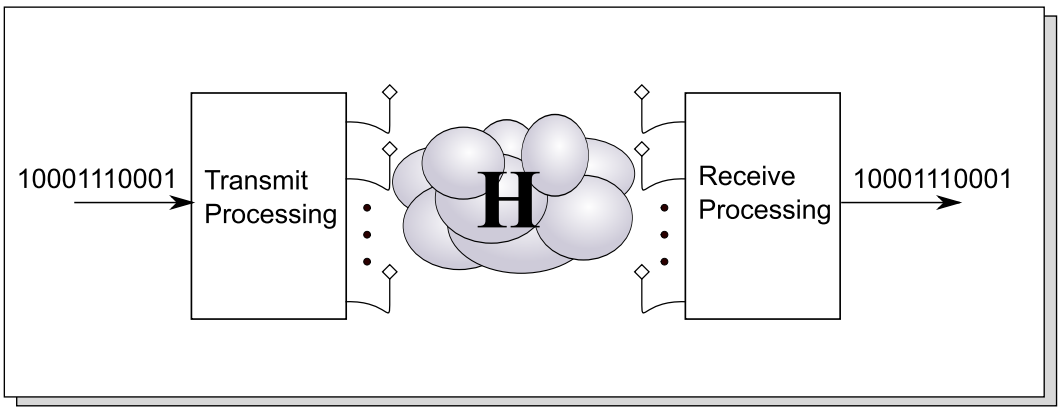


Figure 2.1: MIMO system model.

$K_r \geq 0$ is the Rice factor which is the ratio of power in the specular component to the power in the scattered component. $\mathbf{H}_{NLOS}$ is the NLOS (scattered) component, and its elements are complex Gaussian distributed with zero mean and unity variance. The squared envelope of the elements follow a Rayleigh distribution, and such a channel is referred to as a Rayleigh channel. $\mathbf{H}_{LOS}$ is the LOS (specular) component, and its elements are deterministic. When $K_r > 0$, the elements of $\mathbf{H}$ are complex Gaussian distributed with non-zero mean and unity variance, and the squared envelope of the elements follow a Ricean distribution. It can be shown [58, 59] that if the transmit and receive antennas are spaced by at least one half of a carrier wavelength (at the mobile station) or multiple wavelengths (at the base station) and the radio propagation environment ensures that scattering occurs in all directions, then the elements of $\mathbf{H}_{NLOS}$ are independent and identically distributed (iid) with a complex Gaussian distribution. $\mathbf{H}_{NLOS}$ is full rank in this case. This is generally easy to arrange for typical frequencies used in cellular systems, so for our work we assume that $\mathbf{H}_{NLOS}$ is full rank with $r_{NLOS} = \text{rank}(\mathbf{H}_{NLOS}) = \min(N_T, N_R)$.

$\mathbf{H}_{LOS}$ has maximum rank $r_{LOS} = \text{rank}(\mathbf{H}_{LOS}) = \min(N_T, N_R)$ but this only occurs when the distance between the transmitter and receiver is similar to the antenna spacing [58, 60]. For propagation distances and antenna array sizes typical of practical cellular systems, $\mathbf{H}_{LOS}$ is rank-deficient and often has rank $r_{LOS} = 1$. We will study both full rank and $r_{LOS} = 1$ cases, but for practical results we will assume $r_{LOS} = 1$.

The processing that takes place at the transmit and receive ends of the data link has been the subject of much research in recent years. This topic is not investigated in this thesis, but we draw from published results as necessary. There are number of ways to improve the performance of a wireless link with multiple antennas. The first method is *classical beamforming* in LOS environments. At the transmit end, the signals can be phased in order to steer the beam in the direction of the receiver. With knowledge of the angle of departure from the transmitter to receiver, it is arranged that the N_R signals arrive at the

receiver so that they add constructively. This increases the antenna gain in the direction of the receiver, and can minimize interference by steering nulls towards unintended receivers. With angle of arrival information at the receiver, the receiver linearly combines the received signals to provide receive array gain. Under ideal conditions, these methods achieve *power gain* equal to $N_T \times N_R$. With each doubling of N_R or N_T, a power gain of 3 dB is obtained. In practice, the power gain may be lower.

A second method is *diversity combining*. Wireless transmissions suffer from multi-path propagation, which results in a randomly fading signal at the receiver. Combining replicas of the transmitted signal arriving over different propagation paths using different combining techniques can significantly reduce the probability of outage. Coding over multiple antennas can be used to create diversity branches. If these replicas fade independently of one another, then data can be received more reliably. The channel experiences outage only when all diversity branches are simultaneously in outage. For example, if the probability of outage on any one branch is $0 < P_{out} < 1$, then the probability of outage on d diversity branches is $0 < P_{out}^d < P_{out}$. With N_T transmit antennas and N_R receive antennas, there are $N_T \times N_R$ links, and so $N_T \times N_R - 1$ links may be in a fade yet the channel is still reliable. Thus the maximum diversity order is $d = N_T \times N_R$.

With multiple antennas at both ends, an increase in *degrees of freedom* results, which can be exploited to obtain gains due to *spatial multiplexing*. Up to $K = \min(N_T, N_R)$ data streams can be multiplexed, which increases the spectral efficiency by a factor of up to K. In order to achieve this maximum, the channel matrix $\mathbf{H}$ must be full rank, with $r = \text{rank}(\mathbf{H}) = K$.

There exists a tradeoff between diversity, which reduces BER, and spatial multiplexing, which increases capacity. With diversity, degrees of freedom are used to increase the slope of the BER curve with the number of antennas, but the capacity only grows with the logarithm of the number of antennas. With spatial multiplexing, the capacity increases linearly with the number of antennas. The performance of MIMO channels will be discussed in Section 2.3.

Multiuser (MU) diversity is a major benefit of MIMO when used in a multiuser system. Due to the randomness of the channels from a source to many different users, some user channels will be in a poor state at a particular time while others will enjoy good channel conditions. So the system can choose to transmit only to those users whose channels can support a good rate at that time while ignoring poor user channels. The set of users receiving data changes as channel conditions change, so that the average aggregate network spectral efficiency can be maximized. User scheduling is used to transmit to users in good channel conditions.

2.1.2 Practical Implementation of MIMO In Cellular Systems

In theory, MIMO has a number of very desirable benefits. We are also interested in how well it can perform in a real-world application like a cellular system. There are a number of practical challenges in a cellular system that must be considered when choosing specific MIMO techniques. One challenge is the existence of co-channel interference. Cellular systems are interference limited since there are a large number of base stations (BSs) transmitting simultaneously. Efficient use of spectrum is important, so BSs in the various cells increase their transmit power until they are all mutually interfering with each other at which point any power increase will not result in any further increase in carrier to interference ratio (CIR). This means that Gaussian channel noise is irrelevant in comparison to interference levels. MIMO performance is poor at low SNRs or CIRs, and channel inversion techniques such as zero-forcing are not useful due to noise enhancement[1].

As mentioned earlier, MIMO performs best when the scattering environment is rich. In cellular systems, it is often advantageous to use high gain directional antennas or beam forming in order to improve the received signal strength in the desired direction. This results in a stronger LOS component.

The physical size of mobile stations (MSs), which may be laptop computers or hand-held devices, severely limits the number of antennas available at that end of the data link. Antennas must be spaced wide enough to make sure that the received signals are uncorrelated. With an isotropic scattering environment at the MS, antenna spacing of one half of a carrier wavelength is adequate to ensure low correlation [27]. With the carrier frequencies available in most countries (generally frequencies of 2 GHz to 5 GHz, which corresponds to wavelengths of 6 to 15 cm), it is reasonable to have about two antennas at the MS. The situation at the macrocell BS is different. BS antennas are generally mounted up high, and there are fewer scatterers. As a result, the scattered signals are received within a narrower angle. Thus the minimum antenna element separation required for uncorrelated signal reception is on the order of 10 to 30 wavelengths. Nevertheless, BSs are far less restricted in size and so there can be a larger number of uncorrelated antennas there. Relay station (RS) are intended to be smaller devices, suitable for mounting on lamp posts, sides of buildings and similar locations. They would therefore be mounted at a lower height than the BS, and the antenna spacing requirement falls between the MS and BS requirements. A practical consideration with RSs is keeping their size and complexity to a minimum to allow cost-effective deployment of a relay system. Therfore antenna arrays at the RS may be moderately sized, having perhaps about four antennas.

[1] Signal to interference plus noise ratio (SINR) is used generally throughout this thesis. However, in some specific cases there is no interference so only noise is considered and this quantity is SNR. Cellular systems are interference limited, and therefore noise in the channel is overwhelmed by the interference, in which case we will use carrier to interference ratio (CIR).

In this research we pay attention to these considerations and incorporate realistic cellular characteristics in modelling in order to obtain results directly applicable to real world implementations.

2.2 Relaying Systems

There are a number of different ways of structuring and implementing wireless relaying systems. Here we give a brief overview of the main variants, results published to date in the literature and some of the key considerations in the use of relaying in our work. More details on some variants will appear throughout the thesis.

Two major categories of wireless relaying are cooperative relaying (CR), which uses the relays in parallel paths, and multihop (MH) relaying which has the relays organized to give series hops. Two major data relaying schemes are amplify and forward (AF) in which signals received at the relay are amplified and retransmitted, and decode and forward (DF) in which the data stream is decoded and re-encoded at the relay for transmission on the next hop.

2.2.1 Cooperative Relaying

In a CR or parallel relay system, shown in Fig. 2.2, $n-1$ RSs cooperate in delivering the BS signal through to the MS. The paths are described by complex channel gains, $h_{ik}, i = 0, \ldots (n-1), k = 1, 2$. Including the direct path, this can provide up to diversity order $d = n$ depending on the nature of the channels and the coding scheme used. In a fading wireless system, any of these paths may be in a deep fade at any given time, but it is far less likely that all will fade simultaneously. Diversity gain will be discussed in Section 2.3.3. Disadvantages of this technique include increased processing and relay complexity, the need for widespread dissemination of channel state information (CSI), and redundant signal transmission which decreases spectral efficiency. Most research in cooperative relaying [30, 61, 62] uses a Rayleigh fading model, applicable to NLOS channels, and shows a significant improvement in outage probability. In the work presented in this thesis, we also look at cooperative relaying in a Ricean channel.

A variation of cooperative relaying is user cooperation, [28, 29], in which the relaying is accomplished via users' terminals in the system. This can be beneficial in a system in which there are many mobile terminals deployed throughout a field. An example of where this is most useful is a military application where all mobiles are owned by a single organization and the sharing of user terminals benefits the whole organization. However, this technique has some practical problems in a commercial cellular system. First, the depletion of users' batteries adds another constraint in selecting relaying paths. MSs are made to be as small as possible, and users expect the batteries to last as long as possible. Not having any affiliation with other users in the system, the individual mobile users would not be pleased about their

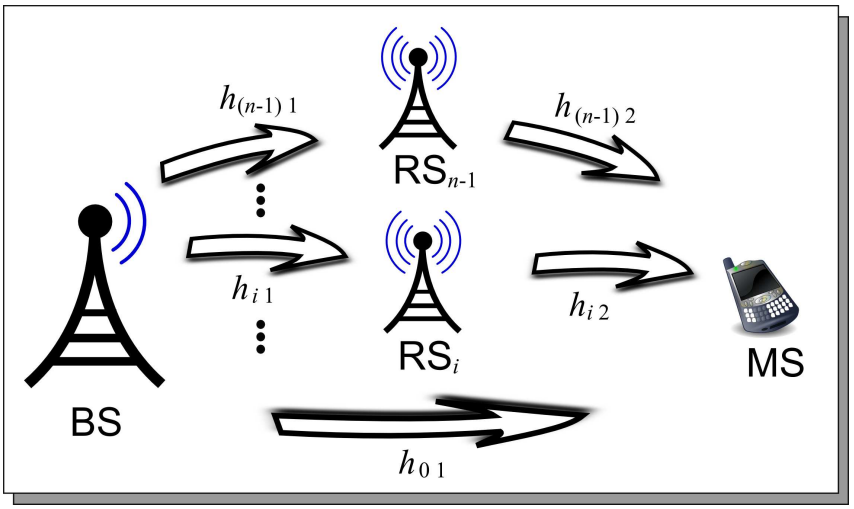


Figure 2.2: A cooperative relaying system.

batteries being depleted to assist operation of the whole system. Another disadvantage is the added randomness. Position, channel strength and lifetime of each user terminal are all unknown and random, which makes it more difficult for a system to schedule transmissions and operate reliably. Protocols in such a system would be much more complicated as a result and system design is more complicated. We believe that a better way to build a relay-assisted cellular system would be with *fixed infrastructure* relay stations which are deliberately planned, installed and operated by the cellular system provider. This would allow them to be powered by the commercial power system, and to be simpler in design which would allow them to be deployed in high enough density. In fact, this approach was eventually adopted by the 802.16j Working Group [21].

2.2.2 Multihop Relaying

MH relaying, Fig. 2.3, uses $n-1$ RSs in series between the source and destination creating n hops. The hop paths are described by complex channel gains, $h_i, i = 1, \ldots n$. Here, both non-cooperative and cooperative techniques [63, 64] are possible. With numerous relay stations in between the source and destination, not only can shadowing obstacles be avoided which drastically reduces the path loss, but the randomly fading component of the received signal can also be reduced. In this case it is necessary to use a Ricean fading model and a dual slope path loss model in order to capture the environment accurately. As fading severity is reduced, diversity combining techniques become less useful (and ultimately not needed), and the signal arrives with higher signal to interference plus noise ratio (SINR). This technique requires less widespread knowledge of channel state information (CSI) and can utilize spatial reuse (SR) methods in order to improve the AASE.

With SR, multiple RSs can transmit simultaneously using the same channel as long

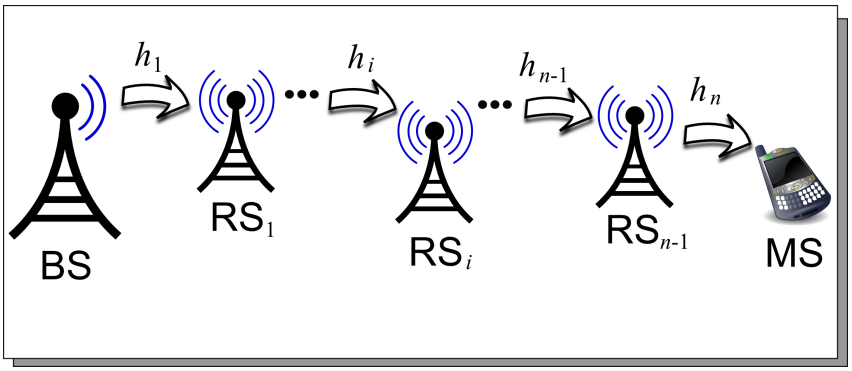


Figure 2.3: A multihop relaying system.

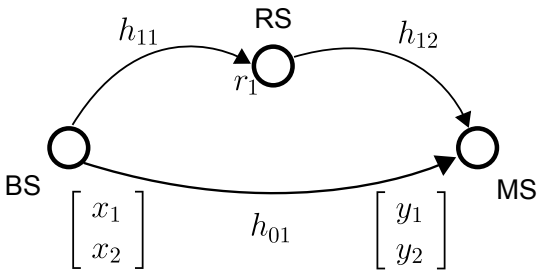


Figure 2.4: A cooperative relaying system with $n = 2$.

as they are far enough away from each other to minimize interference. When using SR scheduling, RSs can be scheduled based not only on their physical separation, but based on the state of the channels as well. With this type of spatial scheduling, relaying can provide diversity gains similar to those of MU diversity.

2.2.3 Amplify and Forward Protocols

As mentioned earlier, relays employing AF simply amplify the received radio signal and retransmit it to the next relay or final destination. To illustrate AF, let us consider downlink of a single relay CR system as shown in Fig. 2.4 with $n = 2$, [65, 66]. First consider the case in which two transmission time slots are used for each data symbol relayed to the destination MS. The source BS transmits symbol $x_1 \in \mathbb{C}$ multiplied by a scalar, $A_1 \in \mathbb{C}$, in the first time slot, and both the RS and the MS receive it. The received signal at the RS is:

$$r_1 = h_{11} A_1 x_1 + w_1 \tag{2.2}$$

The received signal at the MS is:

$$y_1 = h_{01} A_1 x_1 + v_1 \tag{2.3}$$

According to the half-duplex constraint, the RS cannot transmit and receive at the same time. Having a transmitter co-located with a receiver will cause unacceptable interference,

and current technology cannot provide the necessary RF isolation. This problem is not expected to be resolved in the near future, and so current research in relaying usually assumes half-duplex operation at the relay. During the second time slot, the BS transmits symbol x_2 multiplied by, $A_2 \in \mathbb{C}$, and simultaneously the RS transmits its received signal multiplied by B. The total signal received at the MS during the second slot is:

$$y_2 = h_{01}A_2x_2 + v_2 + h_{12}h_{11}BA_1x_1 + h_{12}Bw_1 \tag{2.4}$$

w_i and $v_i, i = 1, 2$ are received noise samples at the RS and MS respectively. In matrix form we have

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} h_{01}A_1 & 0 \\ h_{12}h_{11}BA_1 & h_{01}A_2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ h_{12}B \end{bmatrix} \mathbf{w} + \mathbf{v} \\ &= \mathbf{H}\mathbf{x} + \mathbf{n} \end{aligned} \tag{2.5}$$

where

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} h_{01}A_1 & 0 \\ h_{12}h_{11}BA_1 & h_{01}A_2 \end{bmatrix} \\ \mathbf{n} &= \begin{bmatrix} 0 \\ h_{12}B \end{bmatrix} \mathbf{w} + \mathbf{v} \end{aligned} \tag{2.6}$$

With normalization, $\mathcal{E}\{\mathbf{n}|\mathbf{H}\} = N_0\mathbf{I}_2$. This looks like a MIMO system and so we can write the mutual information as [65, 66]

$$I^{AF} = \frac{1}{2}\log_2 \det\left(\mathbf{I}_2 + \frac{1}{N_0}\mathbf{H}^H\mathbf{H}\right) \quad \text{b/s/Hz} \tag{2.7}$$

assuming that the channels are ergodic block-faded, channels are known at the receivers, and an i.i.d Gaussian codebook is used with a covariance matrix of $\mathcal{E}\{\mathbf{x}\mathbf{x}\} = \mathbf{I}_2$. A factor of $1/2$ appears due to the use of two time slots during the relaying process.

More generally, the BS can transmit a sequence of l symbols while the RS listens for the first l' symbols. The RS transmits a linear combination of those l' received signals during the remaining $(l - l')$ symbols. Any AF protocol can be described as follows.

$$\mathbf{y} = \begin{bmatrix} h_{01}\mathbf{A}_1 & 0 \\ h_{12}h_{11}\mathbf{B}\mathbf{A}_1 & h_{01}\mathbf{A}_2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ h_{12}\mathbf{B} \end{bmatrix} \mathbf{w} + \mathbf{v} \tag{2.8}$$

where $\mathbf{y}, \mathbf{x}, \mathbf{w}, \mathbf{v} \in \mathbb{C}^l$, $\mathbf{A}_1 \in \mathbb{C}^{l' \times l'}$, $\mathbf{A}_2 \in \mathbb{C}^{(l-l') \times (l-l')}$, $\mathbf{B} \in \mathbb{C}^{(l-l') \times l'}$

AF protocols can be divided into two classes: 1) orthogonal AF in which the BS remains silent while the RS transmits, and 2) non-orthogonal AF in which the BS and RS transmit simultaneously. Orthogonal AF (OAF) was studied by Laneman et al [30] and Sendonaris [28]. In this case, $\mathbf{A}_2 = \mathbf{0}$ and there is no interference at the destination. Azarian et al [67] proposed and studied the performance of non-orthogonal AF (NAF), and showed that

despite the interference, an improvement in performance can result. We discuss this further in Section 2.3.4 after introducing some performance measures.

The two-hop MH case can be described similarly by setting $\mathcal{E}\{h_{01}\} \approx 0$. Very often $\mathcal{E}\{h_{01}\} \ll \mathcal{E}\{h_{11}\}, \mathcal{E}\{h_{12}\}$ in practical cellular geometries, so the approximation is valid. Since there are no parallel paths we can drop one of the subscripts, and use h_1 and h_2 for the hop channel gains. Describing more than two hops is an extension of this model. This will be covered in more detail in Chapter 4.

2.2.4 Decode and Forward Protocols

DF for a two-hop system can be described generally as follows [67]. As with AF, the BS transmits a sequence of l symbols, $s_{j1}, j = 1, \ldots l$. The RS listens for the first l' symbols, decodes the data and re-encodes and transmits new symbols $s_{j2}, j = l'+1, \ldots l$. The symbols received at the BS are

$$r_{j2} = \begin{cases} h_{01}s_{j1} + v_j & 0 \leq j \leq l' \\ h_{01}s_{j1} + h_{12}s_{j2} + v_j & l'+1 \leq j \leq l \end{cases} \tag{2.9}$$

The DF technique described by Laneman et al considers $l' = l/2$, while the Azarian dynamic DF (DDF) protocol finds an optimum l' based on the rates that can be supported by the channel conditions on each hop. The performance found by [30, 67] will be given in Section 2.3.4. As before, $\mathcal{E}\{h_{01}\} \approx 0$ in typical cellular systems, so the MS only receives a symbol from the RS, and part of the transmission frame is wasted. However, we will see later in this thesis that in a larger two-dimensional cellular system, wasted portions of time can be used for transmissions in other areas of a cell, which improves the aggregate network rate. This results from the use of spatial reuse scheduling, which will be discussed in Section 3.2.3.

One advantage of the DF technique is that the MS only needs to know h_{01} and h_{12}, but not h_{11} whereas with AF the MS needs to know all three. Another advantage is that the modulation order and coding scheme (using adaptive modulation and coding - AMC) can be optimized for channel conditions on each hop, while with AF the end-to-end coding scheme must be chosen based on the end-to-end SNR, which is limited by the worst hop. The main disadvantage of DF is that it suffers from error propagation. Decoding errors on one hop are passed onto subsequent hops. However, this is only a significant issue at higher error rates. Also, a DF relay requires more complex circuitry than an AF relay.

2.2.5 Ad Hoc Mesh Networks

In an ad hoc network, there is no central mechanism for routing, coordinating and scheduling transmissions in the network. This type of network is common in computer local area networks and sensor networks. The nodes of an ad hoc mesh network are scattered throughout

an area, and one node may communicate by relaying data through some number of intermediate nodes to a distant final destination. A cellular system may also be organized this way, and the 802.16-2004 standard included mesh as an option, so we looked at some aspects of an ad hoc mesh network.

A very important result was published by Gupta and Kumar [25, 26]. They were the first to provide information theoretic analysis of large ad hoc wireless relaying networks. They showed that for n nodes randomly and uniformly scattered throughout a disk of 1 square meter area the capacity per node scales as $\frac{W}{\sqrt{n \log n}}$, where W is the normalized transmission rate (b/s/Hz) available to the nodes. Thus the network capacity is $\frac{W\sqrt{n}}{\sqrt{\log n}}$. They showed that even when nodes are optimally placed and the nodes' transmission ranges are optimally selected, the network can do no better than to provide $O(\frac{W}{\sqrt{n}})$ capacity to each node. Although their derivations are somewhat complex, these results were derived using a very simple physical model for data transmission. There was no consideration of a Rayleigh or Ricean wireless channel. So although this work gives some theoretical bounds for capacity of large wireless networks, it is difficult to extend it to a practical cellular system with more realistic channel models.

A major difficulty with ad hoc networks is finding optimal routes from any given data source to possible destinations. Routing is a difficult problem in computer communications, and it has been shown that finding routes and updating routing tables throughout an ad hoc network can absorb a significant amount of network capacity. As the network grows larger, an increasing proportion of the available network capacity is used up by routing efforts, and the delay in transmitting user data increases. These problems are obviously exacerbated when the nodes in the network are mobile. In a cellular system, traffic flow occurs between an MS and a BS that is connected to the wired backbone network. This type of traffic flow is more tree-like, with MSs as the leaves of the tree, RSs forming branches, and the BS as the root of the tree. For these reasons, we model a cellular system as a fixed infrastructure, tree-topology system with a centralized coordinating mechanism. One main disadvantage of a tree structure is the bottlenecking of data towards the BS. Examples of the tree structure in multihop cellular systems will be shown when we discuss the system model in Chapter 3.

2.3 Performance Measures

There are a number of different performance measures that can be used to determine the desirability of protocols and systems used in relaying. Different systems and protocols performance vary greatly in these measures. There are generally tradeoffs between different performance measures, and so all must be considered carefully when designing multihop relay systems.

2.3.1 Capacity of MIMO Systems

Shannon capacity is the maximum rate that can be supported error-free in a channel. Often it is normalized to channel bandwidth, so that it becomes spectral efficiency measured in bit per second per Hz (b/s/Hz). For a single antenna data link the Shannon capacity [68] is given as a function of signal to interference and noise ratio (SINR), ρ

$$C = \log_2(1+\rho) \quad \text{b/s/Hz} \tag{2.10}$$

For a $N_T \times N_R$ MIMO system, with a deterministic channel, $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ the capacity is [58],

$$C \triangleq \max_{f(\vec{x})} [I(\vec{x};\vec{y})] \tag{2.11}$$

where $\vec{x} \in \mathbb{C}^{N_T}$ is the transmitted symbol vector, $\vec{y} \in \mathbb{C}^{N_R}$ is the received symbol vector, $f(\vec{x})$ is the probability distribution of $\vec{x}$, and $I(\vec{x};\vec{y})$ is the mutual information between $\vec{x}$ and $\vec{y}$. It has been shown [42, 43] that $I(\vec{x};\vec{y})$ is maximized when $\vec{x}$ is zero mean circularly symmetric complex Gaussian (ZMCSCG). The capacity is then

$$C = \max_{\mathbf{R}_{xx}:Tr(\mathbf{R}_{xx})=N_T} \log_2 \left[\det \left(\mathbf{I}_{N_R} + \frac{E_x}{N_T N_0} \mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H \right) \right] \quad \text{b/s/Hz} \tag{2.12}$$

where $\mathbf{I}_{N_R}$ is the $N_R \times N_R$ identity matrix, $\mathbf{R}_{xx}$ is the covariance matrix of $\vec{x}$, E_x is the symbol energy, and N_0 is the noise variance. Often in practical systems the channel is not fully known at the transmitter, so the best the transmitter can do is to transmit independent signals with equal power on all antennas. In this case $\mathbf{R}_{xx} = \mathbf{I}_{N_T}$ and the capacity for equal power allocation is[2]

$$C_{EP} = \log_2 \left[\det \left(\mathbf{I}_{N_R} + \frac{E_x}{N_T N_0} \mathbf{H}\mathbf{H}^H \right) \right] \quad \text{b/s/Hz} \tag{2.13}$$

We use this expression to find the capacity for a given channel realization, $\mathbf{H}$.

The channel is a random variable in wireless system, and so the "capacity" given by (2.12) is also a random variable. We may now talk about average information rate or ergodic capacity

$$\bar{C} \triangleq \mathcal{E}_{\mathbf{H}}[C] \tag{2.14}$$

where $\mathcal{E}_{\mathbf{H}}[*]$ is the average over random variable $\mathbf{H}$.

[2]Better performance can be obtained with a waterfilling algorithm if the channel information is known perfectly at the transmitter. For this research, we use the equal power allocation technique because it is simpler to implement. The basic conclusions about MIMO and MH relaying will not change with the waterfilling technique, but the capacities will be higher throughout. In fact, the difference between ergodic capacity when the channel is known and the ergodic capacity when the channel is unknown decreases with increasing SINR.

With the commonly used iid Rayleigh channel model, elements of $\mathbf{H}$ are iid and circularly symmetric complex Gaussian random variables. For a Rayleigh channel, without channel knowledge at the transmitter, the ergodic capacity has been shown to be [59]

$$\bar{C}_{EP} \approx K \log \frac{\rho}{N_T} + \sum_{i=1}^{K} \mathcal{E}[\log \lambda_i^2] \tag{2.15}$$

where $K = \min(N_T, N_R)$, ρ is the SINR, and λ_i are the singular values of $\mathbf{H}$. At low SINR, the capacity is

$$\bar{C}_{EP} \approx N_R \, \rho \, \log_2 e \tag{2.16}$$

where e is Euler's constant. The above expression shows that there exists only receive array gain at low SINR, and that increasing the number of transmit antennas provides no advantage. We can, however, decrease co-channel interference in the system with the use of transmit-side beamforming, which will increase the SINR.

In our work we are also interested in the performance of MIMO in Ricean channels. As discussed in Section 2.1, the channel matrix is composed of two parts: a deterministic LOS component and a random NLOS component, the proportions of which are controlled by a Rice factor, K_r. In [60], expressions are given for the upper bound on the capacity of the Ricean channel[3]. Assuming equal power allocation among the transmit antennas, the upper bound is given as

$$R(\mathbf{H}) = \mathcal{E}_{\mathbf{H}}[I_{\mathbf{H}}] \leq \log_2 \left[\sum_{p=0}^{K} \left(\frac{\rho b^2}{N_T} \right)^p \sum_{j=0}^{p} K_r^j \, (L-p+1)_{(p-j)} \times \binom{K-j}{p-j} \mathrm{tr}_j(\mathbf{T}) \right] \tag{2.17}$$

where ρ is the SINR at the receiver, $K = \min(N_R, N_T)$, $b = \sqrt{\frac{1}{K_r+1}}$, $L = \max(N_R, N_T)$, $\mathbf{T} = \mathbf{H}_{LOS}\mathbf{H}_{LOS}^H$, $\mathrm{tr}_j(\mathbf{T})$ is the j^{th} elementary symmetric function of $\mathbf{T}$ (see [60] and [69]), and $(m)_n$ is the Pochhammer symbol given by

$$(m)_n = m(m+1)\ldots(m+n+1) \tag{2.18}$$

The SINR on a particular data link is determined by a number of system parameters, such as system geometry, transmit power, antenna gains, receiver thermal noise and path loss. The system model and specific parameters used will be discussed in Chapter 4. This capacity is largest if both $\mathbf{H}_{NLOS}$ and $\mathbf{H}_{LOS}$ are full rank. However, since $\mathbf{H}_{LOS}$ is usually low rank in practical systems, the Rice factor will determine the proportion of power in the

[3]The authors of [60] refer to this quantity as the "average mutual information." They are careful not to use the term "capacity" here since that means the maximum mutual information achievable on the channel over all possible distributions. If one considers maximizing over all possible channels, the maximum rate is achieved on a Rayleigh channel. However, we refer to this quantity as the "capacity of the Ricean channel."

LOS component, and thus the rank of $\mathbf{H}$. Special case number 2 (Corollary 2) in [60] is the case of a Ricean channel with rank 1 $\mathbf{H}_{LOS}$

$$R(\mathbf{H}) = \mathcal{E}_{\mathbf{H}}[I_{\mathbf{H}}] \leq \log_2\left[1 + \sum_{p=1}^{K}\sum_{j=0}^{1}\left(\frac{\rho b^2}{N_T}\right)^p (K_r KL)^j \times (L-p+1)_{(p-j)}\binom{K-j}{p-j}\right] \quad (2.19)$$

2.3.2 Capacity of Multihop Relaying Systems

In this section we look at the capacity performance of multihop relaying systems. Calculation of the network capacity of a MH relay system depends on the type of relaying used. With AF, the noise accumulates on each hop, resulting in an equivalent end-to-end SINR which determines the capacity [70, 71]. The signal received on each hop is

$$r_k = h_k s_k + n_k \quad (2.20)$$

where s_k is the transmitted signal having energy $\mathcal{E}_k$. A gain of G_k can be applied to the signal to invert the loss due to the channel [70]:

$$G_k = \sqrt{\frac{1}{|h_k|^2 + N_0}} \quad (2.21)$$

where N_0 is the power spectral noise density. The resulting SINR was shown to be [70]

$$\rho_{net} = \left[\prod_{k=1}^{n}\left(1 + \frac{1}{\rho_k}\right) - 1\right]^{-1} \quad (2.22)$$

where ρ_k is the received SINR on the k^{th} hop given by

$$\rho_k = \frac{\mathcal{E}_k^2 |h_k|^2}{N_0} \quad (2.23)$$

The upper bound of 2.22 was shown to be

$$\rho_{net} = \left[\sum_{k=1}^{n}\frac{1}{\rho_k}\right]^{-1} \quad (2.24)$$

The end-to-end capacity can be found using the Shannon formula, or the throughput using AMC. It is apparent that the end-to-end SINR and thus the resulting capacity/throughput are dominated by the hop with the lowest SINR. The capacity of AF is [24]

$$R_{net,AF} = \frac{1}{n}\log_2(1 + \rho_{net}) < \frac{1}{n}\log_2(1 + \rho_{min}) \quad (2.25)$$

where $\rho_{min} = \min(\rho_1, \rho_2, \ldots, \rho_n)$.

The following analysis applies to DF relaying. The capacity of the k^{th} hop is $R_k, k\epsilon[1,n]$ which can be determined by the Shannon formula (if AMC is used, then R_k is the throughput). The time required to transmit a frame containing n_{bits} bits on the k^{th} hop is

$$T_k = \frac{n_{bits}}{R_k}, k\epsilon[1,n] \tag{2.26}$$

With no spatial reuse, only one station in the network transmits at a given time. The total time required to transmit n_{bits} bits from end to end is

$$T_{total,NoSR} = \sum_{k=1}^{n} T_k = \sum_{k=1}^{n} \frac{n_{bits}}{R_k} = n_{bits} \sum_{k=1}^{n} \frac{1}{R_k} \tag{2.27}$$

The net or aggregate capacity with no spatial reuse is then

$$R_{net,DF} = \frac{n_{bits}}{T_{total,NoSR}} = \left(\sum_{k=1}^{n} \frac{1}{R_k} \right)^{-1} \tag{2.28}$$

If all the hop capacities are the same, $R_1 = R_2 = \ldots = R_n = R_{hop}$, then the aggregate capacity is

$$R_{net,DF} = R_{hop}/n \tag{2.29}$$

It is apparent that with more hops, more time is wasted in relaying and thus the aggregate capacity may suffer drastically unless the individual hop capacities can be increased greatly with reduced station spacing. Similar to AF, the network capacity in (2.28) is limited to the rate of the slowest hop. However there is an important difference between AF and DF. With AF, the end-to-end rate is limited by the accumulated SINR of all the hops, and that rate must be used throughout the network. It was shown in [24] that the ergodic capacity of a DF multihop system exceeds that of an AF multihop system regardless of the type of channel fading.

$$R_{net,DF} > R_{net,AF} \tag{2.30}$$

Although DF is limited by the slowest hop, individual hops in better fading conditions can still transmit at higher rates. This provides opportunity to feed data to other parts of the network until the slow hop recovers. On the other hand, DF suffers from error accumulation since errors made on one hop are passed onto the next hop. However, if we are considering Shannon capacity, which is defined as the maximum rate at which error-free transmission can occur, then no accumulation occurs. Of course zero BER is not possible in practice, but low BER can be achieved using AMC and hybrid automatic repeat request (H-ARQ).

Network performance can be improved with the use of spatial reuse in which more than one station in the cell may transmit at a time. Simultaneous transmissions can occur if the transmitting stations are separated sufficiently so significant interference is avoided. The minimum allowable separation is the spatial reuse distance. We will discuss network capacity and throughput of multihop systems with spatial reuse in Section 4.2.

2.3.3 Diversity Gain and Multiplexing Gain

In relaying and in MIMO systems, there may exist extra degrees of freedom that can be exploited to enhance performance. In a fading channel, degrees of freedom can be used to improve the reliability of communications (a reduction in the error rate or outage probability) by taking advantage of multiple paths. Diversity gain, d, is defined as [46, 59]

$$d \triangleq -\lim_{\rho \to \infty} \frac{\log(P_E(\rho))}{\log(\rho)} \tag{2.31}$$

where ρ is the SINR and $P_E(\rho)$ is the bit or frame error probability[4]. The diversity order is the negative slope of the error probability (or error rate) curve when plotted against SINR. Higher diversity order is desirable so that the error probability drops off more rapidly with increasing SINR. For large ρ, the error probability for a Rayleigh fading channel can be written as [59]

$$P_E(\rho) \approx c_1 \rho^{-d} \tag{2.32}$$

where c_1 is a constant that depends on the specific modulation scheme used. For binary phase shift keying (BPSK) on a non-fading channel with additive noise, the error probability is approximated by an exponential relationship [59]

$$P_E(\rho) \approx c_2 \exp(-c_3 \rho) \tag{2.33}$$

where c_2, c_3 are constants. Clearly the error probability drops off much faster with SINR in a non-fading channel than in a fading channel with low diversity order. For low SINR, there is often no real performance improvement due to diversity since the spectral efficiency is decreased in order to provide the diversity. This is an important consideration in cellular relaying systems in which the SINR is usually low. More detail on the performance of diversity will be given in Section 4.3.

The maximum possible diversity order is related to the number of diversity branches available. With CR this is related to n, the number of cooperating paths,

$$d_{max} = n \tag{2.34}$$

[4]Outage probability can also be used.

In general, d_{max} is not achieved. Maximum diversity order can only be obtained if the average SINR of all diversity paths are equal. We will see that for real-world two dimensional cellular geometries, the path lengths will generally not be equal, so this condition will not often be met.

With MIMO, diversity is bounded by the number of spatial channels available.

$$d_{max} = N_T N_R \tag{2.35}$$

Alternatively, additional data can be sent using degrees of freedom. Multiplexing gain is the suitable measure, which is defined by [59]

$$r \triangleq \lim_{\rho \to \infty} \frac{R(\rho)}{\log(\rho)} \tag{2.36}$$

where $R(\rho)$ is the rate achievable on the channel.

2.3.4 Diversity-Multiplexing Tradeoff

Diversity gain and multiplexing gain are considered together since there is a tradeoff between the two within the available degrees of freedom.

For a MIMO system with N_T transmit antennas and N_R receive antennas, the optimal diversity multiplexing tradeoff (DMT) curve, $d^*(r)$, was shown by Zheng et al [45] to be the line segments joining the points $(k, d^*(k)), k = 0, 1, \ldots, \min(N_T, N_R)$, where

$$d^*(k) = (N_T - k)(N_R - k) \tag{2.37}$$

The maximum diversity is $d^*_{max} = N_T N_R$ and the maximum multiplexing gain is $r^*_{max} = \min(N_T, N_R)$. The DMT curve for MIMO is shown in Fig. 2.5.

Similarly, DMT exists for cooperative relaying systems and it depends on the specific relaying protocol used. Laneman et al [30] found the DMT for single-relay OAF to be bounded by the line

$$d_{OAF} = \begin{cases} 2(1-2r) & 0 \le r \le 1/2 \\ 0 & \text{elsewhere} \end{cases} \tag{2.38}$$

No diversity gain is possible ($d = 0$) for $1/2 \le r \le 1$ due to the half-duplex assumption. Azarian's [67] single-relay NAF scheme was shown to achieve the optimal DMT for a single relay as

$$d_{NAF} = \begin{cases} (1-r) + (1-2r)^+ & 0 \le r \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.39}$$

where $(x)^+ = \max(x, 0)$. With NAF, $\mathbf{A}_2 \neq \mathbf{0}$ in (2.5), so there exists interference between the source and relay signals at the receiver during the second half of the transmission frame.

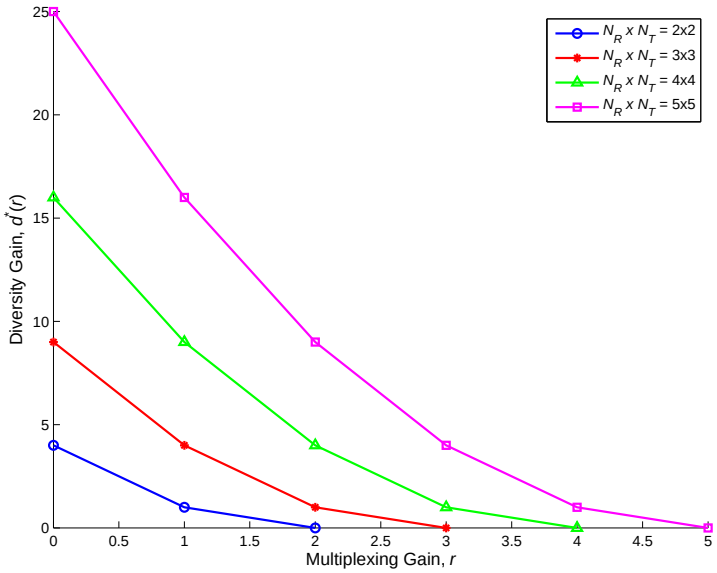


Figure 2.5: Diversity-multiplexing tradeoff for MIMO.

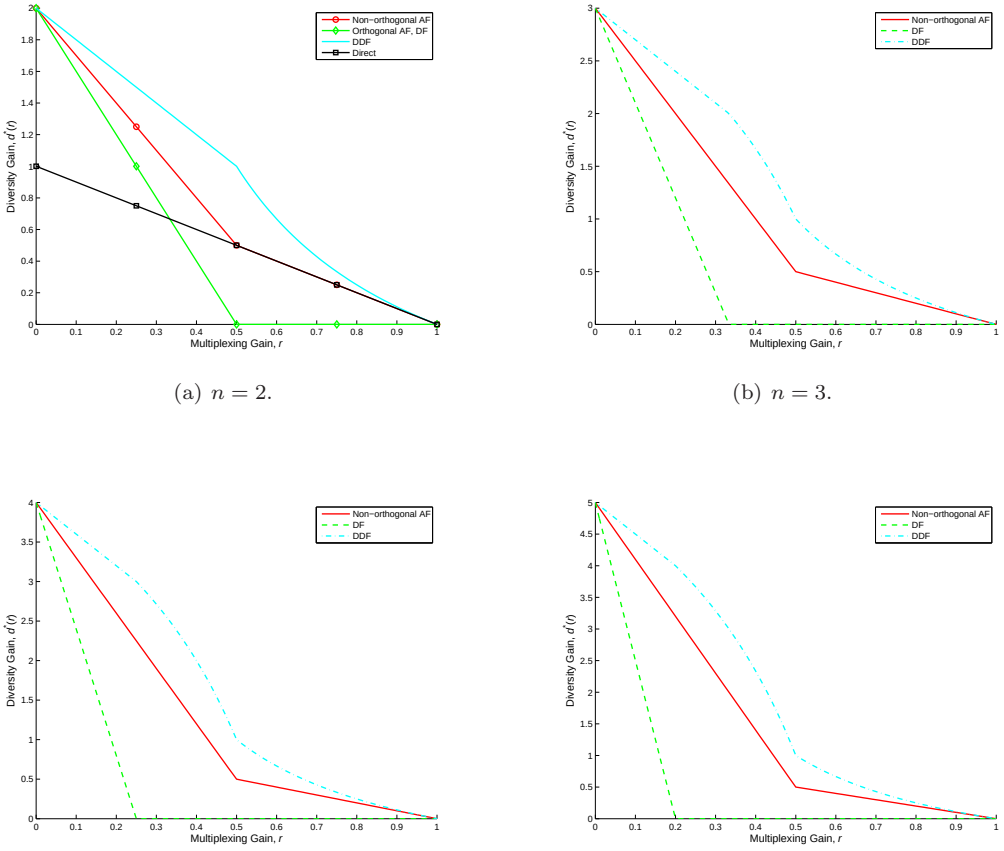
However, there is an increase in the DMT due to better use of the degrees of freedom in the channel. In [30] the DMT bound for a simple single-relay DF protocol was found to be

$$d_{DF} = \begin{cases} 2(1-2r) & 0 \leq r \leq 1/2 \\ 0 & \text{elsewhere} \end{cases} \tag{2.40}$$

which is the same as OAF. However, [67] proposes an improved protocol called dynamic DF (DDF). In this scheme, the relay receives for a time (l' symbols) long enough to maximize the probability of successful decoding before re-encoding and transmitting to the destination. This assures that the signal has a low error rate, which minimizes error propagation. The source (BS) continues to transmit in the second phase, and the destination (MS) decodes based on the combined received signal. It is called dynamic because l' will vary depending on the condition of the channel during that cooperating transmission block (l symbols). In practice this can be done by AMC in which the modulation scheme and error control code strength are selected based on the condition of the channel so that a particular BER is met. This method is more efficient because the length of time varies according to need, rather than being fixed ($l' = l/2$ as in Laneman's OAF and DF schemes), maximizing the use of the channel. The DMT for DDF is [67]

$$d_{DDF} = \begin{cases} 2(1-r) & 0 \leq r < 1/2 \\ (1-r)/r & 1/2 \leq r \leq 1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.41}$$

which clearly outperforms the other AF and DF protocols. The DMT bound was also found



(a) $n = 2$. (b) $n = 3$.

(c) $n = 4$. (d) $n = 5$.

Figure 2.6: Diversity-multiplexing tradeoff for various cooperative relaying schemes.

for $n-1$ relays (n cooperating paths). For NAF, the DMT is [67]

$$d_{NAF} = \begin{cases} (1-r) + (n-1)(1-2r)^+ & 0 \le r \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.42}$$

and

$$d_{DDF} = \begin{cases} n(1-r) & 0 \le r < 1/n \\ 1 + \frac{(n-1)(1-2r)}{1-r} & 1/n \le r < 1/2 \\ \frac{(1-r)}{r} & 1/2 \le r \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.43}$$

for DDF. Fig. 2.6 compares the DMT curves for the various relaying protocols for $n = 2, 3, 4, 5$.

It is interesting to compare the DMT for MIMO and cooperative relaying. Fig. 2.7 shows 2×2 MIMO compared to $n = 4$ cooperative relaying. Both can achieve a maximum diversity order of 4, but MIMO clearly outperforms due to its ability to provide multiplexing gain as well.

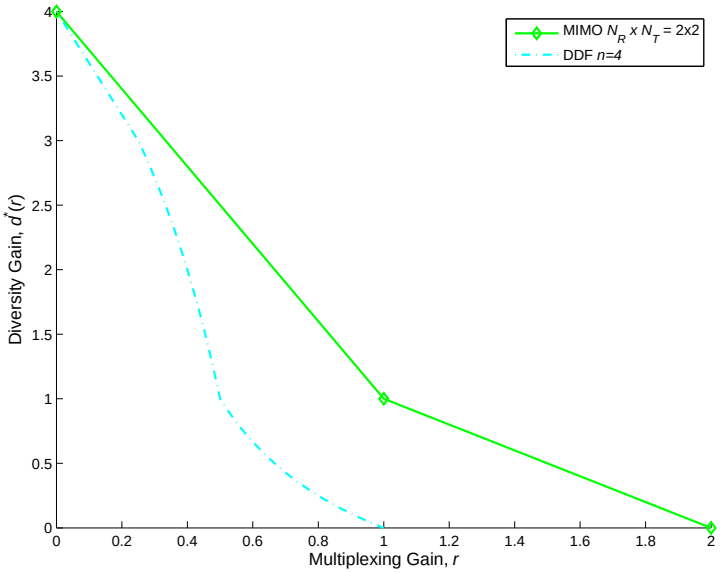


Figure 2.7: Diversity-multiplexing tradeoff comparison between MIMO and DDF.

2.3.5 Delay

One of the main disadvantages of MH relaying is the accumulation of delay between the transmitter and the intended end recipient. There is a tradeoff between network throughput and delay performance. If we wish to obtain higher aggregate network throughput then the delay performance will suffer. However, depending on the application, delay may or may not present a problem. For high speed one way services such as file downloading or one-way audio or video, excessive delay is not much of an issue. For two way audio or video conversations delay must be kept below a certain limit. 4G system requirements defined by the ITU [13–16] specify that delay should be less than 10 ms.

There are several sources of delay in a wireless network.

Physical layer delay Physical layer (PHY) delay is the time required to receive and retransmit packets from the source to the destination through intervening relays. The time required for the signal to propagate the physical distance is very low, since the speed of light is high and propagation distances are only a few kilometers.

ARQ delay Automatic repeat request (ARQ) is a mechanism that corrects errored packets. Retransmission of packets incurs additional delay.

Handshaking delay In 802.16 and other standards, a three-way handshaking scheme (Request-Grant-Confirm) is used between the requester and grantor of a connection.

MAC/Scheduling delay Medium access control (MAC) layer delay is the time between

a station's request to transmit data and the granting of the request by the scheduling mechanism.

Further discussion and some detailed results for delay in 802.16 networks in given in Section 4.5.

2.4 Multihop Relaying in Cellular Systems: Previous Work and Standards Development

Multihop relaying in cellular systems has been considered as early as about 2000 [72–74]. In [72] a multihop cellular network (MCN) is formed by allowing mobiles to communicate with each other, enabling them to relay packets destined for other mobiles. With some system-level analysis and simulations, it has been shown that the MCN can indeed improve throughput. In [73, 74], an ad hoc relaying system overlaid over an existing cellular system is proposed in order to relieve congestion in cells, improving the call blocking probability. If a particular cell has no channels available to serve a MS, an adjacent cell can serve that MS by relaying via ad hoc relay stations (ARSs) into the congested cell. Since then, interest in cellular relaying has grown significantly. Yanikomeroglu et al ([33, 64, 75–77] are just a few of many) have contributed a large body of work from about 2001 onward, in MH relaying and MH cellular systems.

Another group has published a large number of relevant papers concerning the use of MH relaying in HiPERMAN networks and cellular deployments [34, 78–82]. Much of their work is similar to the work presented in this thesis, but limited to two hops and using a simpler path loss model.

The IEEE's 802.16 Working Group recognized the potential benefits of relaying, and quickly added a Mesh[5] option in the 2004 revision [2] of the standard. This sparked interest among the research community in assessing the performance of 802.16-2004 Mesh mode. The IEEE group soon recognized that Mesh mode did not perform very well, and that a fixed infrastructure MH relay system was more suitable. Some of the main difficulties with Mesh are listed below.

- Only best effort (BE) quality of service (QoS) service flows are supported by Mesh.
- End to end delay and jitter are not controlled, and thus Mesh mode can be unsuitable for applications that are delay and jitter sensitive.
- MAC procedures and PHY frame structures are modified to support Mesh, and so Mesh is incompatible with other 802.16 modes.

[5]Following IEEE's nomenclature, *Mesh* is capitalized when referring specifically to 802.16-2004 Mesh mode, while the uncapitalized *mesh* refers to the more generic concept of mesh networks.

- There is no support for mobility.
- A Mesh-enabled system creates an ad-hoc mesh network which is not necessarily suitable for use in a cellular application.
- Only orthogonal frequency division multiplexing (OFDM) with time division multiple access (TDMA) is supported. There are compelling reasons to enable networks to use orthogonal frequency division multiple access (OFDMA).
- It was not completely specified in the standard.

In June 2006, the IEEE formed the 802.16j Working Group [83] in order to modify 802.16-2004 and 802.16e [3] to enable multihop relaying (MR[6]) based on fixed infrastructure. The structure of 802.16j is more suitable for wireless service providers such as cellular system operators and Mesh was completely dropped from the standard. In early discussion documents for 802.16j, several requirements were established for the IEEE Task Group to follow.

- OFDMA is the supported PHY mode.
- MR must be largely backward compatible with 802.16-2004 and backward compatible with all of 802.16e-2005.
- Communication paths must terminate on a BS which is connected to the wired infrastructure. This means that the intended topology is a tree topology and that MR is not a mesh nor an ad-hoc network.
- RSs, whether fixed, nomadic or mobile, are owned and operated by a service provider.
- Diversity can be obtained by using optional cooperative relaying.
- RSs must be low complexity in order to reduce overall system complexity and deployment cost.

Delay and relay complexity is minimized by using layer 2 (link and MAC layer) relaying rather than higher layer relaying. One of the notable differences in 802.16j is the use of OFDMA rather than TDMA as the multiple access method. Since multiple data streams can be allocated within an OFDMA symbol, rather than contending for a TDMA slot, delay can be minimized. MR will address several applications:

- Cell coverage enhancement.

[6]In order to avoid confusion, MR will refer specifically to multihop relaying as defined by IEEE 802.16j, while MH relaying will refer to multihop relaying in general.

- Filling coverage holes created by severe shadowing due to the presence of buildings, geographic features, mass transit tunnels, etc.
- Extending coverage beyond the cell boundary.
- Fast deployment of high bit rate networks for emergency situations.
- Extending coverage into buildings.

The latest version of IEEE 802.16-2009 [20] was published by IEEE on May 29, 2009 and the multihop relay portion exists as a separate amendment 802.16j (Multiple Relay Specification) [21] published on June 12, 2009. A couple of good review papers on 802.16j are [36, 84].

The 3GPP has also included multihop relaying [19] in their proposal to the ITU-R, but the details so far are fairly limited. Some research on relaying for LTE-Advanced has been presented in [37, 85].

Having studied this earlier work and observed the research and standards work that emerged concurrently with ours, we felt that MH relaying was a worthwhile study and there was much work to do to characterize more thoroughly its capabilities in cellular systems, both with single antenna stations and multiple antenna stations, with an eye on practical implementation and realistic modelling.

2.5 Multihop MIMO: Previous Work

When this work was first begun, there was a significant amount of research published on MIMO, a small number of publications on MH relaying, and virtually nothing on combining relaying with MIMO. During the course of our work, more publications appeared. Nearly all of the published work involves simulations since the development of closed form expressions for performance in complicated systems has remained largely unsolved. This section will summarize the most relevant work that we found.

One of the first researchers to consider relaying with multiple antennas is Dohler [86–88]. He recognized that, although MIMO promised great gains, realistically sized mobile terminals may only have one or two uncorrelated antennas. He studied virtual antenna arrays (VAAs, also called distributed-MIMO multistage relay networks) which achieve MIMO gains using single or multiple antenna mobiles which cooperate in transmitting data. In this system, shown in Fig. 2.8, a group of closely spaced mobiles can cooperate in order create a larger MIMO antenna array and data is relayed between tiers of mobiles using DF (also called regenerative) relaying. He derived algorithms that obtained power and resource (time or frequency resources) allocations that perform close to optimum performance.

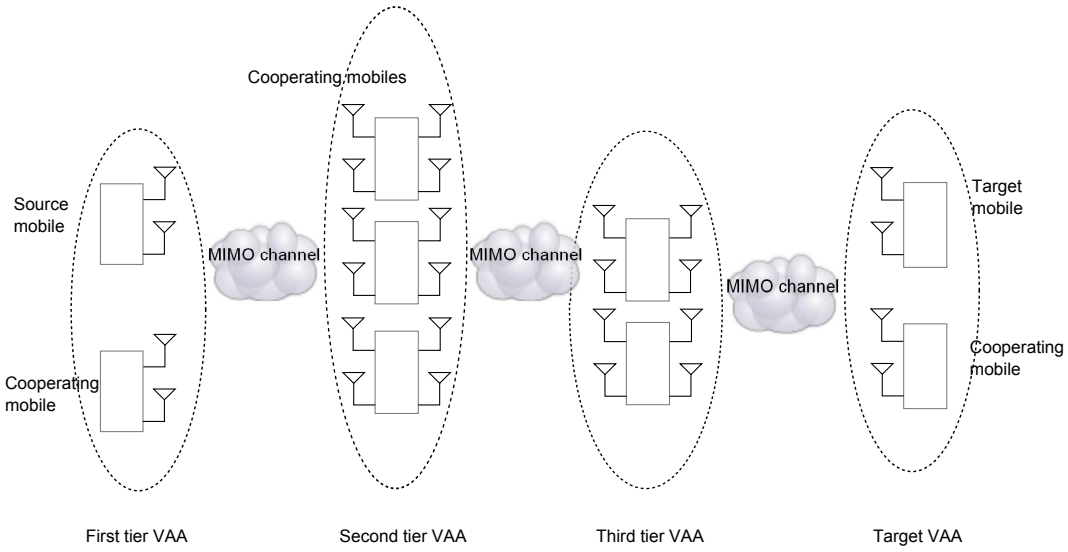


Figure 2.8: Virtual antenna array (VAA) model (three hop example).

Generally, his study of VAAs used a one-dimensional linear model, in which mobile terminals were structured so that a group of cooperating mobile terminals are much closer together than the distances between relaying tiers. His model was generic enough to include multiple hops, but in-depth study was done on two and three hops. This provided some useful insight into MH-MIMO, and the performance of algorithms for power and resource allocation between relaying tiers. Both Rayleigh and Nakagami channel models were used. This work was one of the most important because it was, to our knowledge, the only research looking at combining MH relaying and MIMO at the time, and gave some ideas for structuring further work. It did not give any significant information-theoretical analysis of the capacity of MH-MIMO, but it did illuminate numerous areas for further research. This work may also be applied to coordinated multipoint (CoMP) MIMO systems, which have more recently become an important research topic. One difficulty in implementing this system in practice is the need for coherent transmission of signals among physically separated cooperating mobiles.

The first theoretical analysis of MH-MIMO systems appeared in 2006 [8]. This paper gave upper and lower bounds on capacity for a two hop MH-MIMO system with one data source, one destination with M antennas each, and a close-clustered group of K relays dedicated to the delivery of data (see Fig. 2.9). The relays have N transmit and N receive antennas. Their protocol assumes that the K relays are partitioned so that a group of relays is assigned to a transmit-receive antenna pair on the source and destination. There is no description of how this is accomplished. The main result is that the capacity scales linearly with M as $K \to \infty$ in a Rayleigh channel.

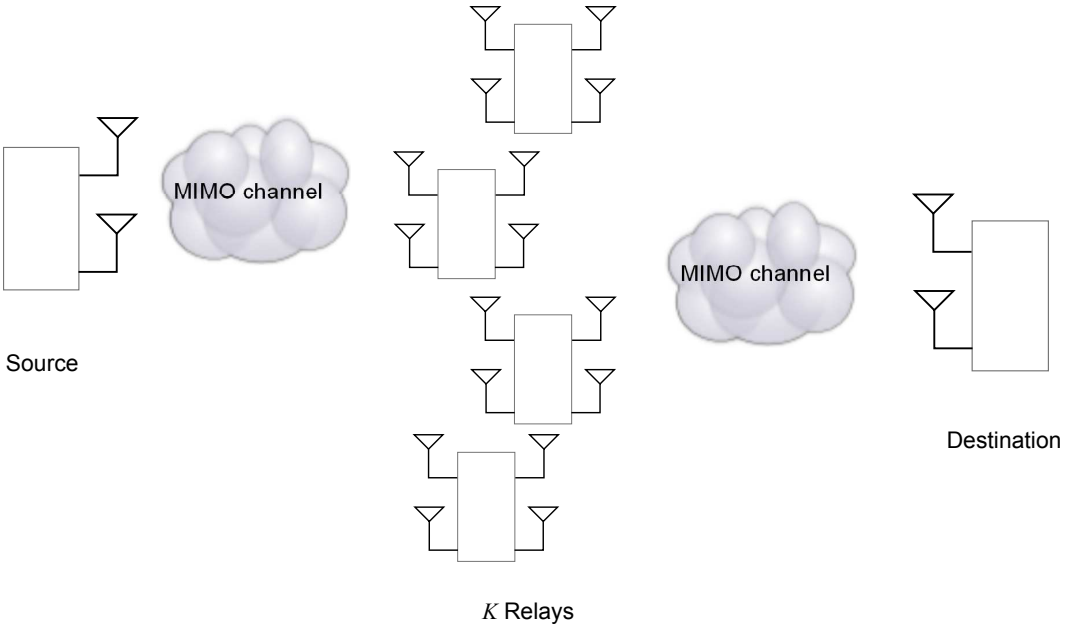


Figure 2.9: Two hop close-clustered relay MH-MIMO model (used in [8–11] and many others).

$$C = \frac{M}{2} \log K + O(1) \tag{2.44}$$

This is similar to the MIMO expression, but there is a factor of 1/2 due to half-duplex relaying. However, the results here have been derived under simplifying assumptions. It has been assumed that perfect channel state information is available at the relays and the destination, and that the distances between the relays is small compared to the distances from the relay cluster to each of source and destination. This geometry is not descriptive of a realistic cellular system. In addition, (2.44) is asymptotic: it holds only for very large K (well over 100) which is not likely to occur in a realistic cellular system. So although this paper shows that the capacity of MH-MIMO systems scales in a similar fashion to MIMO systems, their results cannot be used to calculate capacities of the systems we intend to study.

The outage performance of a two-hop MH-MIMO system similar to the one described above was studied in [9]. Versions of AF and DF were compared, and it was found that a digital selective relaying (DSR) method outperformed other relaying methods at low SINR, but was outperformed by direct transmission (no relays) at higher SINR. In DSR, DF is used to relay data, and the single cooperating path that is in the best channel condition is selected for a given transmission block. Of course all channel state information must be known at the relays in order for the path to be selected.

The DMT of MH-MIMO was studied in [10, 11]. The model used was the same as the one above (Fig. 2.9) with a close-clustered group of K relays deployed to assist in transmitting

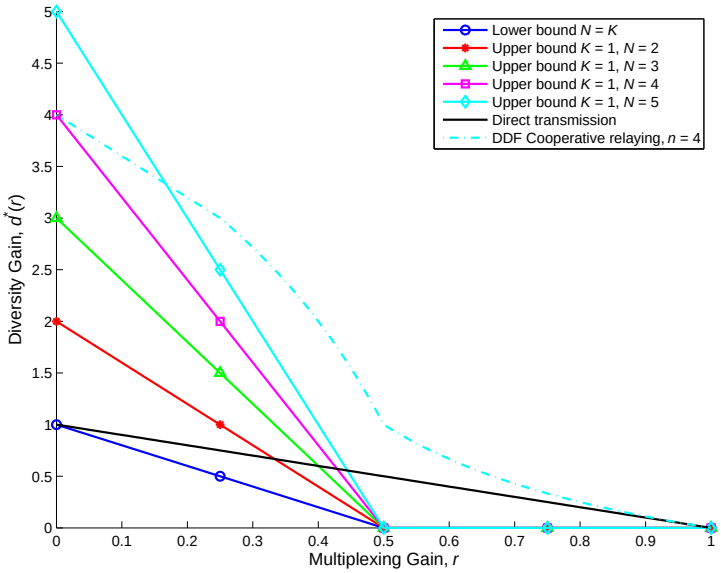


Figure 2.10: Diversity-multiplexing tradeoff bounds for MH-MIMO.

data from a single source to a single destination. In this case the source and destination each have a single antenna, and the total number of antennas at all the relays is N. The DMT curve is bounded by

$$1 - 2r \leq d_{MH-MIMO} \leq N(1-2r) \quad 0 \leq r \leq 0.5 \tag{2.45}$$

The lower bound occurs when $K = N$, i.e. when each relay has one antenna, and the upper bound when there is one relay with N antennas ($K = 1$). Fig. 2.10 shows the DMT bounds, which shows surprisingly poor performance when compared to the DDF protocol described in [67].

An important point to note about the DMT tradeoff curves shown so far is that the capacity for all protocols is normalized for fair comparison of diversity and multiplexing gain. For example, the direct transmission curve shown in Fig. 2.10 indicates that its performance is often better than that of certain relaying protocols at the same SNR. However, the existence of a relay in fact lowers the path loss significantly, increasing the SNR, which can increase the end-to-end capacity of relaying despite the half-duplex loss. Correct comparison of techniques in a system-level study requires consideration of the geometry and the path loss model.

Chapter 3

System Model

In this chapter, we put together a practical system model for studying multihop cellular systems. We start with a one dimensional model and extend it to two dimensional cellular layouts. Both single antenna systems and multiple antenna systems are modelled. The simulation platform chosen was MatLab since it provides a large number of built-in functions and is easily programmed.

3.1 System Model Overview

In order to obtain good results it is important to put together a model that accurately describes the system under study. The following are key aspects of a good model for relaying in a cellular system.

- It uses typical cellular topologies.
- A path loss model that incorporates both line of sight (LOS) and non-line of sight (NLOS).
- It captures both scattering effects and path loss as a function of distance.
- The model aligns with models adopted by standards bodies such as the Third Generation Partnership Project (3GPP) and IEEE 802.16.

There are three main components in a multihop cellular system: a base station (BS), multiple relay stations (RSs), and a large number of mobile stations (MSs). The MS, also known as user equipment or terminal device, is a wireless telephone handset or wireless computer terminal usually owned and operated by the system user. The BS (also known as a base transceiver station or a Node B in various standards) is a major installation comprised of an antenna, radio frequency (RF) transceivers and control equipment. Owned and operated by the cellular service provider, it provides wireless services to numerous MSs, and is connected to the switched telephone system via high capacity trunks. These trunks

can be provided via fibre optic systems, copper wire facilities, or point to point microwave links. The BS antenna is typically mounted fairly high (heights of 30 m or more are typical), for clear coverage to MSs. Data is transmitted from the BS to MSs in *downlink* channels, and in the reverse direction in *uplink* channels. Numerous regularly-spaced BSs exist throughput a cellular system in order to provide continuous coverage over a large area. The radius of coverage of a cell ranges from 300 m in an urban environment to several kms in rural areas. The specific frequencies used by cellular systems vary by country and technology, but they are often in the 700 MHz, 800-900 MHz, 1800-1900 MHz, 2100 MHz and 2500 MHz bands. In a multihop cellular system, RSs sharing those same frequencies are deployed to assist in capacity and coverage. RSs should be small in size, much less complex than a BS, and can be mounted at numerous locations such as on street lamp posts, sides of buildings, and inside mass transit stations. They are mounted at a lower height than BS antennas, typically 10 m or less.

In existing cellular systems, an MS is served directly by the BS that can provide the best quality channel to it. The MS is handed off to other BSs as the user moves. In a multihop cellular system, an MS can be served by either the BS directly, or the most suitable RS. RSs receive downlink transmissions from the BS, and pass the data along towards the MSs. Uplink transmissions from numerous MSs are collected by their serving RSs and retransmitted upstream towards the BS. Links between the BS and RSs, and between different RSs are called *transport* links, while links between BS and the MSs, and between RSs and MSs are called *access* links.

We are interested in the **system** level performance of multihop relaying and multiple antenna techniques in cellular layouts, and ultimately want to know the aggregate bit rate carried by the cell and the area averaged spectral efficiency. We use fairly simple **link** level calculations to find the capacities and throughputs of individual links: BS to RS transport links, RS to RS transport links, BS to MS access links, and RS to MS access links. With these link capacities and throughputs, we find the aggregate network capacities and throughputs considering non-spatial reuse and spatial reuse scheduling.

While second and third generation cellular systems have used time division multiple access (TDMA) or code division multiple access (CDMA), fourth generation systems will use orthogonal frequency division multiple access (OFDMA) and single carrier frequency division multiple access (SC-FDMA). We will consider OFDMA as the multiple access technique.

The following sections describe the system level and link level components of the model.

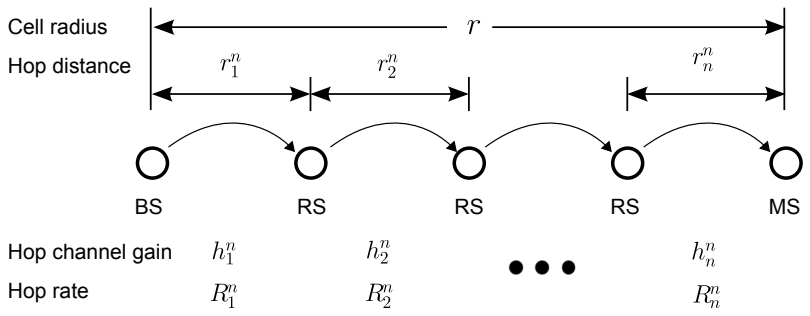


Figure 3.1: A one dimensional linear multihop relay system.

3.2 System Level

3.2.1 One Dimensional Multihop Topologies

A one dimensional linear MH relaying system with $n-1$ relays and n hops is shown in Fig. 3.1. The cell radius is r, and the hop distances are $r_k^n, k = 1, 2, ..., n$. To simplify calculations for the one dimensional case only, we often use equally spaced relays so that $r_k^n = r/n, k = 1, 2, ..., n$. The complex channel gains on the hops are $h_k^n, k = 1, 2, ..., n$, the rates on the hops are $R_k^n, k = 1, 2, ..., n$ and the noise at the receiver of each hop is $n_k^n, k = 1, 2, ..., n$. The superscript, $\bullet^n$ emphasizes that there are n hops in the system under study, and often we drop this superscript when the number of hops is understood. In this research R_k^n may refer to Shannon capacity, ergodic capacity, or throughput as necessary. Section 4.1 will give further details.

This model is suitable for analyzing a single cell. However, we must take into account that in a full cellular system there are numerous similar cells which create interference in the cell under study (the *subject* cell). The severity of interference will depend on the cluster size for radio resource reuse, N. With universal frequency reuse, $N = 1$, significant co-channel interference from one cell to another occurs. As an example, Fig. 3.2 shows a two-hop relay system with four tiers of interference. The frequency assignments for cluster sizes $N = 1, 2, 3, 4$ are shown. One dimensional topologies for two, three, four and five hop systems are shown in Appendix B. These drawings were used to find the distances to interferers for both forward link and reverse link transmission in Matlab simulations.

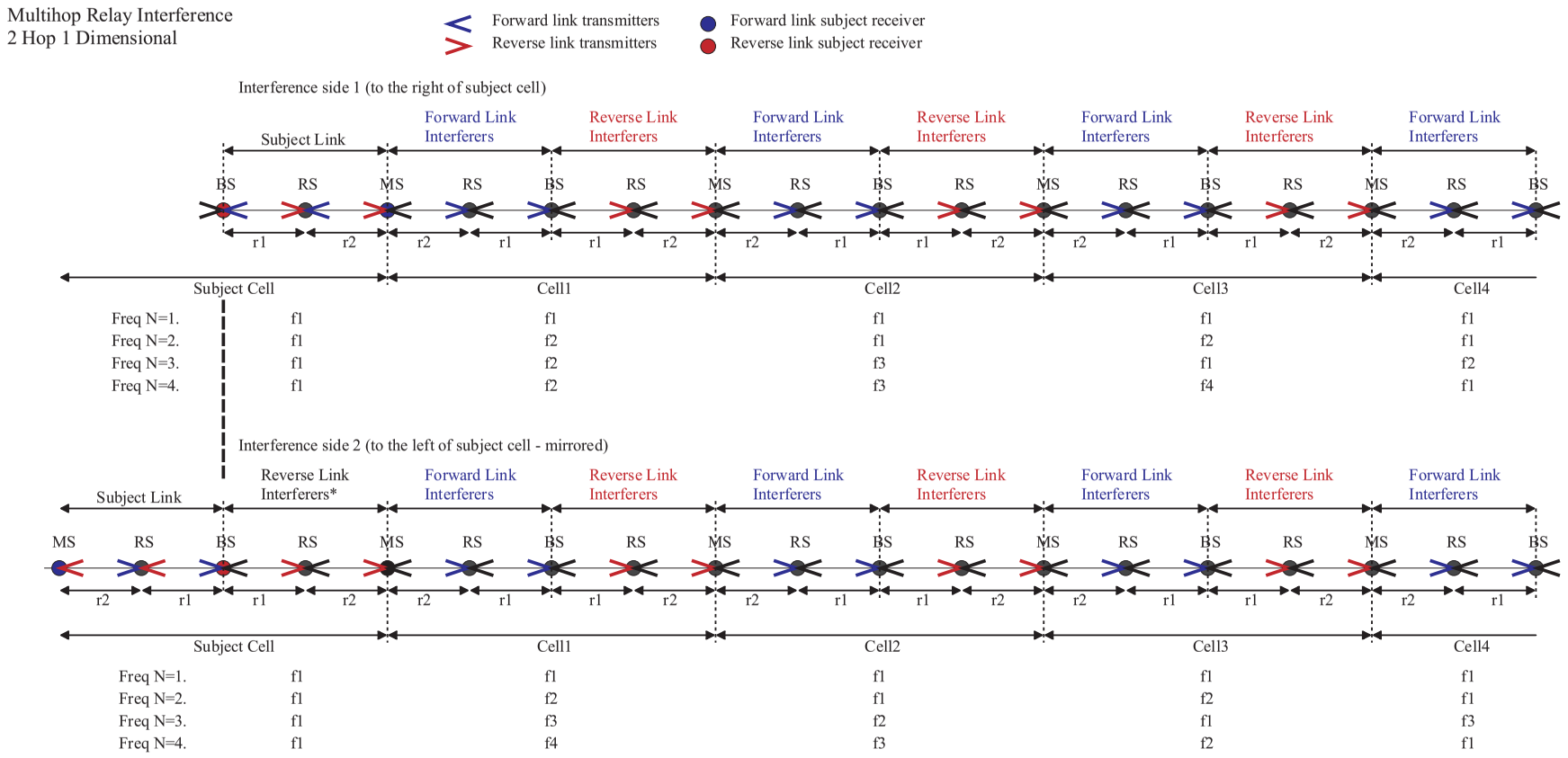
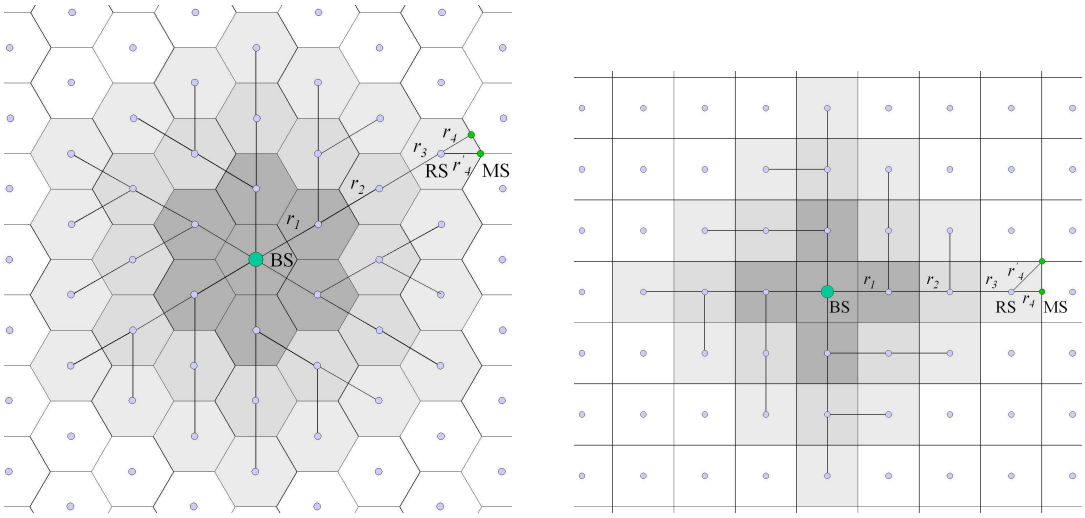


Figure 3.2: A two-hop one-dimensional linear multihop relay system, showing four tiers of interference.



(a) Hexagonal topology: 37 shaded subcells comprise one cell.

(b) Manhattan topology: 25 shaded subcells comprise one cell.

Figure 3.3: Four-hop relay cellular topologies.

3.2.2 Two Dimensional Multihop Topologies

A two dimensional cellular system is composed of numerous cells covering a large area. Typically these cells are approximated as tessellating equal-size hexagons in most greenfield scenarios, or as equal-size squares in a downtown urban street scenario (Manhattan). A BS is deployed in the centre of each cell, which serves numerous MSs in that cell. In a MH relaying cellular system, numerous RSs are deployed throughout the cell, which subdivides the cell into numerous subcells. A cellular system is best served using regularly-placed fixed relays (infrastructure-based relaying). Fig. 3.3 shows how a cell is subdivided into subcells for hexagonal and Manhattan topologies for four-hop relaying. Cellular topologies for two, three, four and five hop systems are shown in Appendix C.

MSs will be served by the closest RS or BS, handing off as necessary to a closer station as the MS moves. As a result, some MSs will obtain service directly from the BS (one hop), some MSs will be served via two hops, and so on. Transport links contain the data aggregated from numerous subcells, while access links serve user MSs directly. In a four-hop cell, a given MS may be served via one, two, three or four hops depending on its location in the cell. Along a straight line in a n-hop cell we have distances $r_1 = r_2 = \ldots = 2r_n$. From the figures we can see that last hop to the cell edge is one-half the distance of the other hops. The cell radius is $r = r_1 + r_2 + \ldots + r_n = A_n r_n, A_n = 2n - 1$. The inter-relay distances are fixed but the last hop distance, r'_n, is variable depending on the MS location. $0 \leq r'_n \leq \frac{2r}{\sqrt{3}(2n-1)}$ for the hexagonal topology and $0 \leq r'_n \leq \frac{\sqrt{2}r}{2n-1}$ for Manhattan. In the earlier simulations, we uniformly scattered MSs throughout the subcells, calculated the r'_n for each, and then calculated the resulting network aggregate rate. This required a lot of

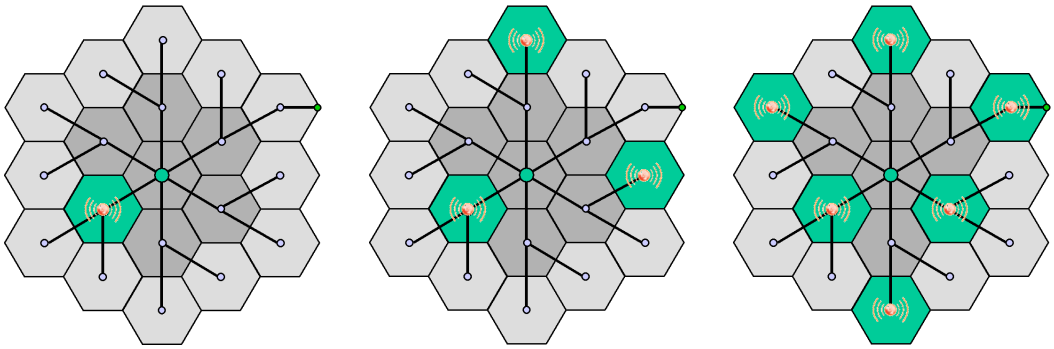
computation and a complicated MatLab program. Later we used a single "lumped up" MS for each cell at cell edges: either $r_n = r/(2n-1)$ or at the worst case location, $r'_n = \frac{2r}{\sqrt{3}(2n-1)}$ (hexagonal), $r'_n = \frac{\sqrt{2}r}{2n-1}$ (Manhattan) in order to calculate the total rate delivered to each cell. The difference in the results between these approaches is not significant. As shown in (2.28) the end-to-end rate for a MS is dominated by the worst hop rate. If the MS is fortunate enough to be close to the RS, it enjoys a good rate for that hop only, but this does not affect the end-to-end rate significantly.

Note that the geometry of the two dimensional cases can be approximated by repeating the one-dimensional topologies at suitable angles. For example, when studying the four hop Manhattan topology, we can take two one-dimensional topologies at right angles to each other. Similarly a hexagonal topology can be approximated by taking three one-dimensional topologies with 60° between them. This is an approximation since it omits stations that are off the one-dimensional lines. However, these contribute little interference since the gain of omni-directional antennas is fairly low, and the off-axis gain of directional antennas is very low.

Because these topologies are tree structures with end-to-end connections between the BS and many MSs, bottlenecks occur on links nearer the BS. We consider only decode and forward relaying, in which the data stream is decoded and re-encoded at RSs before transmitting on the next hop. This is necessary in order to enable the use of adaptive modulation and coding (AMC) on the individual hops. All relay stations are wireless, and may not transmit and receive simultaneously (half-duplex). We can calculate the SINR at each station's receiver, and then find the Shannon capacity or throughput attainable on the link using a suitable formula (described in Section 4.1). In this work, we have normalized the rates and throughputs by the channel bandwidth, so that they are expressed as spectral efficiency with units b/s/Hz. Ultimately we are interested in the cell's aggregate data rate, R_{net} (b/s/Hz), which can be calculated knowing the physical data rates achievable on each of the hops, and by considering the *spatial reuse* (described in Section 3.2.3) schedule imposed by the MAC layer.

3.2.3 Spatial Reuse

To find the aggregate network rate/throughput, we need to consider MAC layer scheduling of data transmissions throughout a cell. MH relaying allows spatial reuse of channels, in which multiple subcells may transmit simultaneously. A spatial reuse neighbourhood defines the neighbourhood within which stations must coordinate their transmissions, which in turn determines the level of inter-subcell interference due to simultaneously transmitting stations. We first consider a very conservative two-hop neighbourhood (as described in 802.16 Mesh mode), in which stations greater than two hops away from a transmitter may simultaneously



(a) No spatial reuse. (b) Spatial reuse SR2. (c) Spatial reuse SR1.

Figure 3.4: Illustration of spatial reuse in a three-hop hexagonal cell: a) no spatial reuse, b) spatial reuse case SR2, c) spatial reuse case SR1.

transmit (we call this SR2). We also consider the case in which stations greater than one hop away from a transmitter may simultaneously transmit (SR1).

Fig. 3.4 shows a three-hop hexagonal illustration of the benefit of spatial reuse. In Fig. 3.4(a), only one station in the cell transmits in a given time slot (no spatial reuse). Although inter-subcell interference is very low and thus coverage and reliability for MSs is improved, all subcells must share the spectrum - a wasteful situation. The result is poor spectral efficiency over the coverage area of the cell. MH relaying can actually decrease the AASE dramatically as compared to single hop. Spatial reuse scheduling, SR2, allows a few subcells to transmit simultaneously. Fig. 3.4(b) shows one of many possible schemes for SR2. Since up to three stations may transmit simultaneously, AASE is improved with SR2. Fig. 3.4(c) shows one of many possible SR1 schemes - in this example, six stations transmit simultaneously in a given time slot. Despite the additional inter-subcell interference, SR1 provides a significant improvement in AASE. The schedule of simultaneous transmissions depends on the geometry and the subcell spatial reuse distance. With multiple hops in two dimensions, spatial reuse time division multiple access becomes possible, and we must employ optimization techniques in order to determine a schedule that maximizes the network throughput.

Spatial reuse schedules, together with link level capacities and throughputs, are used to calculate the cell's aggregate throughput, R_{net} (b/s/Hz), and AASE in b/s/Hz/km^2 [89, 90]. While most researchers have used R_{net} calculations to assess multihop performance, we also look at AASE, which gives a meaningful measure of the system-wide use of spectrum. The definitions and more detailed discussions of R_{net} and AASE, and details of calculation methods are given in Chapter 4.

3.3 Link Level

3.3.1 Channel Model

Since we will be studying the performance of a wireless system in realistic cellular scenarios, we need a good model for RF propagation. We use components that are well established in the literature.

There are three main components of a wireless model:

Bulk path loss This is a deterministic value that gives the signal loss experienced due to distance between the transmitter and receiver.

Large scale fading This is random effect describing the shadowing impact of large objects (buildings, trees, hills, etc.) that impede and attenuate the signal.

Small scale fading This describes the random effect of the superposition of multiple received RF signals at the receiver.

The bulk path loss characteristic is very dependent on the specific propagation environment. So for real world applications, a system designer would need detailed local knowledge of the actual propagation environment. For the purposes of general research, empirical models are used that model generic environments such as urban, suburban, and rural areas. There are several models that have similar characteristics, but we have chosen to use a model adopted by the 3GPP [1], IEEE 802.16 [91] and other standards bodies. It is based on the Okumura-Hata and Walfish-Ikegami models for urban macrocell and microcell environments. Since a benefit of multihop relaying is the ability to relay around obstacles, it is necessary to use a dual slope model, which selects non-line-of-sight (NLOS) or line-of-sight (LOS) path loss as appropriate for the distance. Combining LOS and NLOS, the dual slope path loss model for frequencies up to 1.9 GHz is [1] is

$$PL_{dB}(x) = -20\log_{10}[\gamma(x)] = \begin{cases} 34.5 + 38.0\log_{10}(x) + \psi_{dB} & b < x < 5000\text{ m}, \\ 30.2 + 26.0\log_{10}(x) + \psi_{dB} & 20\text{ m} < x < b \end{cases} \tag{3.1}$$

where x is distance, $\gamma(x)$ is the path gain, and b is the distance *breakpoint* below which a NLOS path becomes LOS (typically about 300 m in an urban environment [1, 92]). A log-normal random variable, ψ_{dB}, in (3.1) models large scale fading. ψ_{dB} has zero mean, and its standard deviation, $\sigma_{\psi_{dB}}$ is typically 10 dB in an urban NLOS microcell scenario, and 4 dB in an urban LOS microcell [1]. We can see that the path loss exponent is $\beta_{NLOS} = 3.8$ for NLOS and $\beta_{LOS} = 2.6$ for LOS.

There has also been some interest in systems based on higher carrier frequencies such as 5 GHz. The path loss model extended to a frequency of 5 GHz [4] is

$$PL_{dB}(x) = -20\log_{10}[\gamma(x)] = \begin{cases} 42.5 + 38.0\log_{10}(x) + \psi_{dB} & b < x < 5000\text{ m}, \\ 38.2 + 26.0\log_{10}(x) + \psi_{dB} & 20\text{ m} < x < b \end{cases} \tag{3.2}$$

The research presented in this thesis is done primarily at 5 GHz since the loss is higher - a worst case scenario. With the inclusion of small scale fading, the complete channel, h, for a single antenna system with dependency on distance, x (in metres), is given by the Ricean model

$$h(x) = \gamma(x) \cdot \left(\sqrt{\frac{K_r(x)}{1 + K_r(x)}} e^{j\phi} + \sqrt{\frac{1}{1 + K_r(x)}} \mathcal{CN}(0, 1) \right) \tag{3.3}$$

where $\gamma(x)$ and $K_r(x)$ are area-averaged path gain and Rice factor, respectively, $\mathcal{CN}(0, 1)$ is a normalized complex Gaussian (i.e. normal) random variable and ϕ is an arbitrary phase. The Rice factor is the ratio of signal power in the LOS component to the power in the NLOS component. When $K_r(x) = 0$, we have a Rayleigh channel. The Rice factor, $K_r(x)$, is modeled as a function of distance [1], [93]

$$K_r(x) = \begin{cases} 0 & b < x < 5000\text{ m}, \\ 10^{1.3 - 0.003x} & 20\text{ m} < x < b \end{cases} \tag{3.4}$$

From (3.3) and (3.4) we can see that the path gain, $h(x)$, is a Rayleigh random variable when $b < x < 5000$ m and Ricean when $20\text{ m} < x < b$.

In order to find the SINR at a receiving station, the classical link budget with typical parameters is evaluated. Interference from all stations in all other cells, and in simultaneously transmitting subcells is included. With the inclusion of ψ_{dB}, SINR is a random variable, and averages can be found using Monte Carlo simulation.

The SINR for a receiver, q, receiving from transmitter, p, is

$$\begin{aligned} SINR(x_{pq}) =& P_{TX,dBm} + G_{TX,dB} - PL_{dB}(x_{pq}) \\ &+ G_{RX,dB} - N_{RX,dBm} - I_{RX,dBm} \end{aligned} \tag{3.5}$$

where x_{pq} is the distance between the transmitter and receiver, $P_{TX,dBm} = 10\log_{10}(P_{TX})$ is the transmit power, $G_{TX,dB} = 10\log_{10}(G_{TX})$ is the transmit antenna gain, $PL_{dB}(x)$ is the path loss from equation (3.1) or (3.2), $G_{RX,dB} = 10\log_{10}(G_{RX})$ is the receive antenna gain, $N_{RX,dBm}$ is the receiver noise floor, and $I_{RX,dBm}$ is the total interference at the receiver q. In order to improve spectral efficiency on frequency selective fading channels, standards such as LTE-A and 802.16 employ OFDM. The noise floor is calculated for OFDM256 as described in 802.16 documents [2] and Section 8.3.11.1 of [3].

$$\begin{aligned} N_{RX,dBm} =& 10\log_{10}(kT) + F_{dB} + M_{dB} \\ &+ 10\log_{10}\left(F_s \frac{N_{used}}{N_{FFT}} \right) \end{aligned} \tag{3.6}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, $T = 300$ K is the noise temperature, F_{dB} is the noise figure, M_{dB} is the margin, F_s is the sampling frequency, N_{used} is number of OFDM subcarriers used for data and N_{FFT} is the fast Fourier transform (FFT) size used in OFDM. With OFDM256, we have $N_{FFT} = 256$, $N_{used} = 200$ and $F_s = 11.4$ MHz. This results in $N_{RX,dBm} = -91.5$ dBm.

The interference level at receiver q is given by

$$\begin{aligned} I_{RX,dBm} =& P_{TX,dBm} + G_{TX,dB} + 10\log_{10}\left[\sum_{i\in\mathcal{I}}(PL(x_{iq}))\right] \\ &+ G_{RX,dB} - G_{FB,dB} \end{aligned} \tag{3.7}$$

where $\mathcal{I}$ is the set of interferers, $P_{TX,dBm}$ is the transmit power, $G_{TX,dB}$ is the antenna gain, x_{iq} is the distance between interferer, i, and the receiver, q, and $G_{FB,dB}$ is the front to back gain ratio of the antenna (included in the equation for interferers that are outside of the receiver's main beam). The set of interferers depends on the geometry of the system being studied. We have included all interferers that contribute significant interference at the receiver. It is worth noting here that the interferer distances, x_{iq}, are usually much larger than the link distance, x_{pq}. So with carefully designed system geometry, it can be arranged that most or all of the transmission links are LOS (lower path loss), while the interfering paths are NLOS (high path loss).

Link budgets for each hop of a MH communication link have been calculated using the parameters summarized in Table 3.1 (parameters based on [1–4]). Five different antenna configurations have been used, as outlined in Table 3.2. Noise and interference from other cells have been included to calculate SINR on each link (BS-RS, RS-RS, and RS-MS) for various one-dimensional, hexagonal and Manhattan layouts. Interference from all subcells in cells outside the studied cell has been included, since we assume no coordination of transmissions occurs beyond cell boundaries. Interference from other subcells within the studied cell only occurs when spatial reuse is employed.

3.3.2 Multihop MIMO Channel

The model we use for multihop relaying with multiple antennas is a modification of the single antenna model given above. Fig. 3.5 shows a one-dimensional multihop multiple-input multiple-output (MH-MIMO) system. $\vec{x}_0$ is the transmitted data stream and $\vec{x}_n$ is the received data stream after n hops. For the k^{th} hop, vector $\vec{s}_k \in \mathbb{C}^{N_{T,k}\times 1}$ contains the symbols sent, $\mathbf{H}_k^n \in \mathbb{C}^{N_{R,k}\times N_{T,k}}$ is the channel matrix, and $\vec{r}_k \in \mathbb{C}^{N_{R,k}\times 1}$ contains the symbols received, given by

$$\vec{r}_k = \mathbf{H}_k^n \vec{s}_k + \vec{n}_k \tag{3.8}$$

Table 3.1: Model parameters (based on [1–4]).

System Parameters	
Carrier frequency, f_c	5 GHz
Channel bandwidth, W	10 MHz
Receiver noise figure, F_{dB}	8 dB
Receiver noise floor, $N_{RX,dBm}$	-91.5 dBm
Maximum transmit power, $P_{TX,dBm}$	30 dBm
Omni antenna gain, $G_{TX,dB}$, $G_{RX,dB}$	9 dBi
Directional antenna gain, $G_{TX,dB}$, $G_{RX,dB}$	17.5 dBi
Directional antenna front-back ratio, $G_{FB,dB}$	25 dB
Link margin, M_{dB}	5 dB
Multiple access	TDMA
PHY mode	802.16 OFDM256
BS antenna height	32 m
MS antenna height	1.5 m
Dimensioning Parameters	
Cell radius (r)	0.5, 1.0, 2.0, and 3.0 km
Cell cluster size (N)	1, 2, 3, and 4
Number of hops (n)	1, 2, 3, and 4
Antenna pattern	Five configurations (Table 3.2)

Table 3.2: Antenna configurations: O=omni, D=directional.

Configuration (A)	BS Antenna	RS Antenna	MS Antenna
1	O tx, O rx	O tx, O rx	O tx, O rx
2	D tx, D rx	O tx, O rx	O tx, O rx
3	D tx, D rx	O tx, D rx	O tx, O rx
4	D tx, D rx	D tx, D rx	O tx, O rx
5	D tx, D rx	D tx, D rx	D tx, D rx

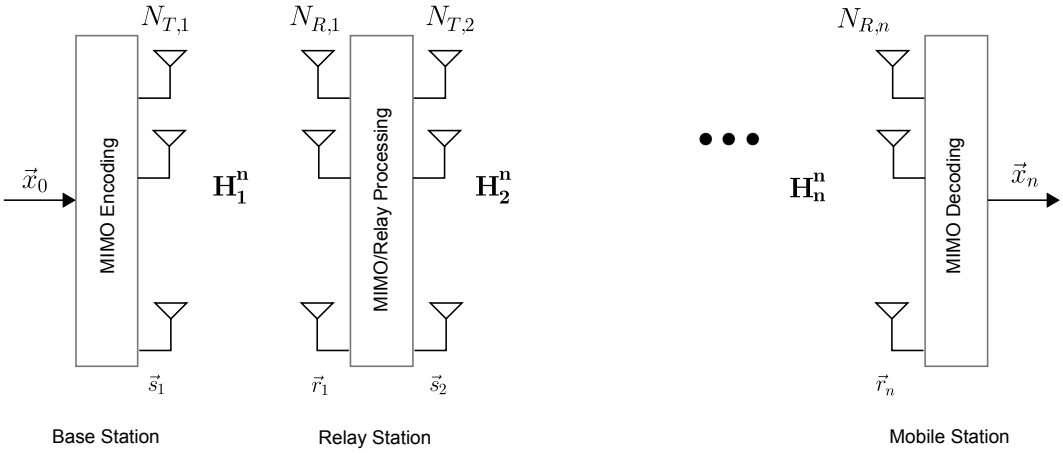


Figure 3.5: A general one-dimensional multihop MIMO system model.

where $\vec{n}_k \in \mathbb{C}^{N_{R,k}\times 1}$ is the noise received on the hop (which in general may include co-channel interference).

Following (2.1) and (3.3), the channel matrix, $\mathbf{H}_k^n \in \mathbb{C}^{N_{R,k}\times N_{T,k}}$, is

$$\mathbf{H}_k^n = \mathbf{H}(r_k^n) \tag{3.9}$$

where r_k^n is the distance of the k^{th} hop,

$$\mathbf{H}(x) = \gamma(x) \cdot \left(\sqrt{\frac{K_r(x)}{1+K_r(x)}}\mathbf{H}_{LOS} + \sqrt{\frac{1}{1+K_r(x)}}\mathbf{H}_{NLOS} \right), \tag{3.10}$$

$\gamma(x)$ is given by (3.1) or (3.2), and $K_r(x)$ is given by (3.4). The elements of $\mathbf{H}_{NLOS}$ are complex Gaussian distributed with zero mean and unity variance.

$\mathbf{H}_{LOS}$ is deterministic and must be normalized appropriately for fair comparison. According to [94] antenna arrays can be designed to make $\mathbf{H}_{LOS}$ full rank which creates orthogonal rows.

For a $N_T \times N_R$ MIMO system, the normalized $\mathbf{H}_{LOS}$ is

$$\mathbf{H}_{LOS} = C_{norm}\mathbf{H}'_{LOS} \tag{3.11}$$

where

$$C_{norm} = \frac{\sqrt{N_T N_R}}{\|\mathbf{H}'_{LOS}\|_F} \tag{3.12}$$

and $\| \star \|_F$ is the Frobenius norm.

For a 2×2 MIMO system, a full rank channel matrix is

$$\mathbf{H}_{LOS,full} = \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.13}$$

and a rank 1 channel matrix is

$$\mathbf{H}_{LOS,rank1} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{3.14}$$

For a 3×3 MIMO system, a full rank channel matrix is

$$\mathbf{H}_{LOS,full} = \sqrt{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.15}$$

and a rank 1 channel matrix is

$$\mathbf{H}_{LOS,rank1} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{3.16}$$

Full rank $\mathbf{H}_{LOS}$ can exist in short distance wireless links, such as those encountered in wireless local area networks (WLANs), but for greater distances, such as those typical in cellular systems, $\mathbf{H}_{LOS}$ is usually rank 1. We will consider both full rank and rank 1 situations.

For MIMO, antenna configuration 1 is used since MIMO arrays generally use omni-directional antennas in order to maximize the scattering in the channel.

3.3.3 Calculating Link Capacities and Throughputs

Single Antenna

Calculation of the **capacity** of each individual link between transmitter p and receiver q in the system is done using the Shannon capacity formula:

$$C_{pq} = \log_2[1 + SINR(x_{pq})] \quad \text{b/s/Hz} \tag{3.17}$$

Note that there is no 1/2 factor in front as is normally required to take into account the half-duplex constraint. Calculation of network rate capacities as described in Section 4.2 takes care of the half-duplex loss. Unless otherwise stated, throughput is normalized by the channel bandwidth to give units of b/s/Hz, which is a measure of spectral efficiency.

AMC is an important technique used to obtain high spectral efficiency when the channel conditions are changing, as is the case in wireless networks. The SINR varies considerably throughout a wireless network, and often between successive transmissions on a particular link. On each link, AMC adapts the forward error correction (FEC) coding and modulation size used on each data burst. An appropriate modulation and coding scheme (MCS) is used to meet the required BER while maximizing the spectral efficiency. AMC has been applied to determine the **throughput** achievable for the SINR calculated on each individual link. The use of AMC will be described further in Section 4.1.

MIMO

We have taken two approaches to the calculation of capacity for MIMO channel links. The first method is the use of Monte Carlo simulation, in which a large number of samples for $\mathbf{H}$ are generated, and the capacity for each sample is calculated using (2.13). The aggregate

network capacity is calculated for each sample using the methods described in Section 4.2. The ergodic aggregate network capacity is calculated by taking the average of all of the samples. This may take a lot of computation time since a large number of samples must be generated to obtain accurate results. The second method used is direct computation of the ergodic capacity using expressions (2.17) and (2.19) which were given in Section 2.3.1. We compared the accuracy of these two methods in Appendix A.

Chapter 4

Single Antenna Relaying

In this chapter, we examine some performance aspects of single antenna multihop relaying.

4.1 Link Throughput

One method of calculating potentially achievable link rates in networks is by using the Shannon capacity formula (2.10). This gives the maximum error-free rate theoretically achievable on a Gaussian channel with a given SINR. We use this formula for many of the calculations in Section 4.2. However, we are also interested in realistically obtainable rates, and ultimately the throughput considering PHY and MAC overhead and inefficiencies. Shannon capacity and realistic throughput using AMC are both calculated in this work.

Table 4.1 shows the modulation and coding used for different received SINR values [5] based on 802.16. Column 3 gives AMC thresholds for a coded BER less than 10^{-6} (from Table 266 in [2, 3]). For example, if the SINR is above 21.0 dB for a particular transmission burst, then the system can adopt the highest rate AMC scheme: 64QAM with an FEC code having a code rate of 3/4. If during the next burst the SINR falls to 8.4 dB, then the transmission link must fall back to QPSK with a stronger FEC code, with a 1/2 code rate, in order to meet the BER requirement. Column 4 shows the theoretical spectral efficiency for each AMC case. Obviously this is an upper bound on the spectral efficiency of AMC. Each of the PHY and MAC layers add overhead in the frame so that fewer of the transmitted bits carry usable payload. Additionally, in OFDM, numerous subcarriers are not available to carry data, being either null carriers or pilot tones. Theoretical and simulation work in [5, 6] on HiperLAN, HiperMAN[1] and 802.16 gives more realistic throughput spectral efficiencies for OFDM. Columns 5 and 6 (derived from the results in [6]) are the spectral efficiencies calculated for the PHY and MAC layers considering overhead, and column 7 gives the overall spectral efficiencies obtained via simulation of an 802.16/HiperMAN system with OFDM. The MatLab code was built to allow selection of any of the columns, 4 through 7, but the

[1]The HiperMAN standard uses the same PHY layer as 802.16, and thus we use research results for both standards.

Table 4.1: AMC levels in 802.16 (adapted from [2, 3, 5, 6]).

Modulation	Code Rate	SINR Thresh. (dB)	Theor. Rate (b/channel use)	Throughput (Normalized)		
				PHY (b/s/Hz)	MAC (b/s/Hz)	Simulated (b/s/Hz)
BPSK	1/2	3.0	0.5	0.35	0.31	0.29
QPSK	1/2	6.0	1.0	0.69	0.61	0.59
QPSK	3/4	8.5	1.5	1.04	0.93	0.88
16QAM	1/2	11.5	2.0	1.38	1.23	1.19
16QAM	3/4	15.0	3.0	2.07	1.86	1.76
64QAM	2/3	19.0	4.0	2.77	2.48	2.35
64QAM	3/4	21.0	4.5	3.11	2.79	2.63

values in column 7 were used for the results presented here since these values are most realistic.

4.2 Network Capacity and Throughput of Multihop Cellular Systems

4.2.1 One Dimensional Multihop Relaying Systems

In Section 2.3.2, we found the capacity of multihop relaying systems without spatial reuse. Here we find the aggregate network throughput including spatial reuse. Consider that stations two hops apart (a spatial reuse distance of 2) may transmit simultaneously. This is the best that can be done, since if all stations transmit simultaneously, significant interference results, and the half-duplex constraint is not met. In a linear one-dimensional multihop network, there are two equal size transmission frames containing n_{bits} each. In the first frame, hops $1, 3, 5, ...,$ and p are active, and in the second frame hops $2, 4, 6, ...,$ and q are active, where

$$p = \begin{cases} n & \text{for } n \text{ odd, and} \\ n-1 & \text{for } n \text{ even} \end{cases} \tag{4.1}$$

$$q = \begin{cases} n-1 & \text{for } n \text{ odd, and} \\ n & \text{for } n \text{ even} \end{cases} \tag{4.2}$$

The time required to transmit the first frame is

$$T_{frame1} = \max\left(T_1, T_3, \ldots, T_p\right) = \max\left(\frac{n_{bits}}{R_1}, \frac{n_{bits}}{R_3}, \ldots, \frac{n_{bits}}{R_p}\right) \tag{4.3}$$

since the frame time must be long enough for the lowest capacity hop to finish transmitting all n_{bits} bits. The slowest hop is the bottleneck and some time is wasted. Likewise, the time required to transmit the second frame is

$$T_{frame2} = \max\left(T_2, T_4, \ldots, T_q\right) = \max\left(\frac{n_{bits}}{R_2}, \frac{n_{bits}}{R_4}, \ldots, \frac{n_{bits}}{R_q}\right) \tag{4.4}$$

The total time required is then

$$T_{total,SR} = T_{frame1} + T_{frame2} \tag{4.5}$$

$$= n_{bits}\left[\max\left(\frac{1}{R_1}, \frac{1}{R_3}, \ldots, \frac{1}{R_p}\right) + \max\left(\frac{1}{R_2}, \frac{1}{R_4}, \ldots, \frac{1}{R_q}\right)\right] \tag{4.6}$$

and the spatial reuse network capacity is

$$R_{net,SR} = \frac{n_{bits}}{T_{total,SR}} \tag{4.7}$$

$$= \left[\max\left(\frac{1}{R_1}, \frac{1}{R_3}, \ldots, \frac{1}{R_p}\right) + \max\left(\frac{1}{R_2}, \frac{1}{R_4}, \ldots, \frac{1}{R_q}\right)\right]^{-1} \tag{4.8}$$

If, for example, $R_1 = R_2 = \ldots = R_n = R_{hop}$ then

$$R_{net,SR} = R_{hop}/2 \tag{4.9}$$

When $n > 2$, there is an improvement over (2.29).

There exists wasted transmission time in a one-dimensional DF relaying system, but we will see in the next section that SR makes better use of this wasted time when there are two dimensions of SR transmission opportunities.

The capacity of a two dimensional cellular system can be calculated similarly to that of a one dimensional system, with the added complication of considering the geometry of the network topology, and more complicated possibilities in spatial reuse scheduling. We have taken two approaches to calculating the throughput achievable in MH cellular networks. The first approach uses an analytical method to find an approximation to aggregate network, while the second uses a linear optimization method and can be employed for live spatial TDMA (STDMA) scheduling.

4.2.2 Calculating Aggregate Network Throughput

A key measure of a network's performance is the aggregate network throughput, R_{net}, which is the rate at which bits are delivered in the whole network. AASE [89, 90] is given by

$$AASE = \frac{R_{net}}{A \cdot N} \quad \text{b/s/Hz/km}^2 \tag{4.10}$$

where A is the area of the cell, and N is the cell cluster size. This measure must be considered in addition to the network throughput since, in cellular systems, we are concerned not only with spectral efficiency, but efficiency of spectrum use over the system coverage area. Both R_{net} and $AASE$ will be presented in the results section.

When calculating the aggregate network throughput, we must consider the PHY layer throughput achievable on each link, and the MAC layer scheduling of transmissions among the network nodes. Relays must operate in half-duplex mode, since in practical systems a

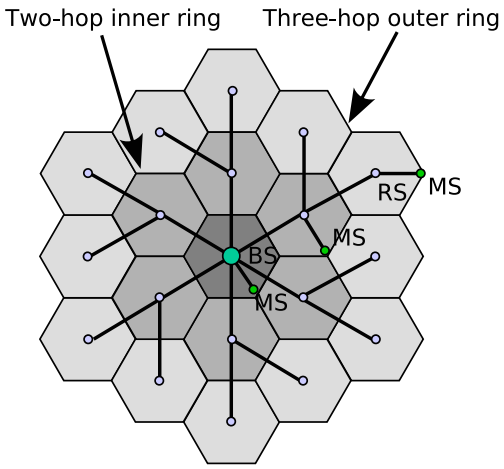


Figure 4.1: Three-hop hexagonal relay topology.

relay cannot simultaneously receive and transmit data in a channel. As a result, relaying incurs inefficiency since additional radio resources (orthogonal channels such as TDMA or FDMA slots, or code division multiple access - CDMA - codes) must be used to relay data. The aggregate network throughput for one cell is calculated by

$$R_{net} = \frac{n_{delivered}}{T_{delivery} \cdot W} \quad \text{b/s/Hz} \tag{4.11}$$

where $n_{delivered}$ is the number of bits delivered to all subscribers in the cell in one full schedule cycle, $T_{delivery}$ is the total time required to deliver the bits (includes the time required for relaying and scheduling), and W is the bandwidth used. The detailed expression used to calculate this differs for each topology (Manhattan and hexagonal), for different number of hops, and depending on the form of spatial reuse employed. Here we show the derivation for the forward link of the three hop hexagonal case. Fig. 4.1 shows the topology and Table 4.2 lists the symbols used. R_1 to R_5 are normalized throughputs on various links calculated for the specific propagation and interference environment using the technique described in Section 4.1, with units b/s/Hz. The centre BS delivers data to MSs within its range and relays data to the inner RSs (RS_1s), these six RSs adjacent to the BS deliver data to inner ring MSs within their range and relay data to the outer RSs (RS_2s), and the outer 12 RSs deliver to outer ring MSs within their range. We assume that MSs are evenly distributed throughout the cell, so that each subcell has the same number of mobiles.

A schedule is followed in order to deliver bits from the BS to all MSs in the network, and it must be assured that buffers at RSs neither overrun or underrun in the process. Buffer overrun occurs in the downlink (forward) direction when throughputs on upstream links are too fast for downstream links to keep up. If this is the case, upstream stations must wait for bits to be cleared out of downstream RS, which results in wasted time slots. Similarly,

Table 4.2: Symbols used for R_{net} calculation.

Symbol	Description
R_1	Throughput on link between BS and first RS
R_2	Throughput on link between first RS and second RS
R_3	Throughput on link between second RS and MSs within range of that RS
R_4	Throughput on link between BS and MSs within range of the BS
R_5	Throughput on link between first RS and MSs within range of that RS
n_1	Number of MSs within range of the BS (one-hop MSs)
n_2	Number of MSs within range of the first set of RSs (two-hop MSs)
n_3	Number of MSs within range of the second set of RSs (three-hop MSs)
η	Density of MSs, stations per subcell

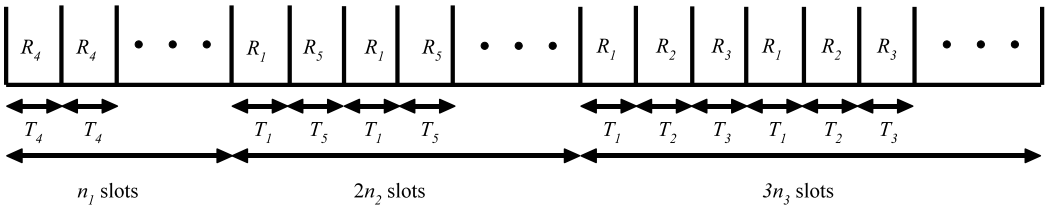


Figure 4.2: Three-hop relay bit delivery schedule - no spatial reuse.

buffer underrun occurs when downstream links are too fast for upstream links to keep up. Figure 4.2 shows a simple schedule without spatial reuse, which delivers an equal number of bits, n_b in each time slot. Note that

$$T_j = n_b/(R_j \cdot W), \quad j = 1, 2, \ldots, 5 \tag{4.12}$$

In the first n_1 time slots, n_b bits are delivered in turn to each of n_1 MSs near the BS. This occurs in one hop. Two-hop transmissions to inner ring MSs require $2n_2$ time slots: n_2 transmissions from BS to the RSs, and another n_2 transmissions from RS to the MSs. Similarly, three-hop transmissions to outer ring MSs require $3n_3$ time slots. Since the number of bits delivered in each slot are equal, from (4.12) we have

$$R_1T_1 = R_2T_2 = R_3T_3 = R_4T_4 = R_5T_5 \tag{4.13}$$

No Spatial Reuse

Many of the time slots in Fig. 4.2 are responsible for relaying only and thus reduce the aggregate throughput. During *last-hop* time slots bits are delivered to MSs. To calculate the aggregate network throughput, we only count the total number of bits transmitted during these last-hop slots. The total number of bits delivered to the MSs network-wide in one complete scheduling cycle is

$$n_{delivered} = (n_1R_4T_4 + n_2R_5T_5 + n_3R_3T_3) \cdot W \tag{4.14}$$

and the total time required to do this is

$$T_{delivery} = n_1 T_4 + n_2(T_1 + T_5) + n_3(T_1 + T_2 + T_3) \tag{4.15}$$

Using expression (4.11), it is easy to show that

$$R_{net,noSR} = \frac{(n_1 + n_2 + n_3)R_1}{n_1 \frac{R_1}{R_4} + n_2(1 + \frac{R_1}{R_5}) + n_3(1 + \frac{R_1}{R_2} + \frac{R_1}{R_3})} \tag{4.16}$$

For a hexagonal topology

$$n_1 = \eta \tag{4.17}$$

$$n_2 = 6\eta \tag{4.18}$$

$$n_3 = 12\eta \tag{4.19}$$

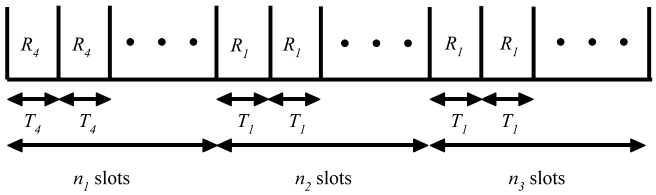
where η is MS density per subcell. We can further simplify equation (4.16) if we assume that in the worst case, all MSs are at subcell edges, and thus the average link throughputs on all last hops are equal: $R_3 = R_4 = R_5$. Expression (4.16) becomes

$$R_{net,noSR} = \frac{19}{\frac{18}{R_1} + \frac{12}{R_2} + \frac{19}{R_3}} \tag{4.20}$$

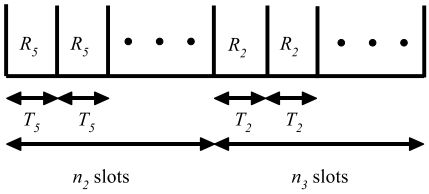
Spatial Reuse Cases

With a multihop cell, spatial reuse can be employed since simultaneous transmissions may occur for transmitters that are far enough away to cause minimal interference. Network throughput with spatial reuse, $R_{net,SR}$ is more complicated to calculate since it depends on the geometry, the effectiveness of the scheduler, the spatial reuse distance chosen (which results in a subcell "cluster"), and the rates, R_1 to R_5. In Section 4.2.3, we show how network throughput can be determined using an optimization approach. Here we use a simplified analysis method to approximate the throughput. The method shown here guarantees no buffer overrun or underrun, but the SR cases are approximations since there may be some inefficient use of time slots.

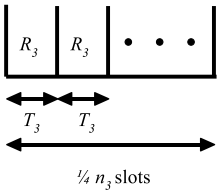
We show here an example derivation for the forward link of a three-hop hexagonal network and the SR2 case. Fig. 4.3 shows the time frames required. While the BS is transmitting, no other station may transmit in the same channel. The base station transmits three sets of bits. The first set is transmitted to n_1 MSs within range of the BS at a rate of R_4. The second set of bits is destined for MSs within range of the two-hop inner ring - there are n_2 of these MSs and the transmission rate is $R1$ between the BS and the inner ring of RSs. The third set of bits is destined for MSs within range of the three-hop outer ring - there are n_3 of these MSs and the transmission rate is again $R1$. The total time required



(a) BS transmission.



(b) RS_1 transmissions.



(c) RS_2 transmissions.

Figure 4.3: Illustration of SR2 scheduling in a three hop hexagonal cell.

for the BS to deliver n_{bits} to one-hop MSs and sufficient bits to be relayed via each RS_1, is therefore

$$T_{BS} = n_1 T_4 + n_2 T_1 + n_3 T_1 \tag{4.21}$$

Next, the RS_1s are scheduled to transmit. While any RS_1 is transmitting, two distant RS_2s may transmit at the same time, but no more than one RS_1 may transmit at a time. The length of time required for all RS_1s to deliver all bits to destination MSs within their range and relayed bits to the RS_2s is

$$T_{RS_1} = n_2 T_5 + n_3 T_2 \tag{4.22}$$

While any RS_2 is transmitting, three other RS_2s may transmit at the same time. The maximum time required to deliver all queued bits to all of the three-hop MSs is then

$$T_{RS_2} = \frac{n_3 T_3}{4} \tag{4.23}$$

Some of this time can be saved due to some RS_2s being able to transmit simultaneously with certain RS_1s. The amount of time saved depends on the relative link rates (i.e. whether T_{RS_1} is long enough to allow many bits be transmitted from the RS_2s), the actual distribution of MSs throughout the cell, and whether or not sufficient bits have been queued at the RS_2s in time for transmission (buffer underrun). To come up with simple expressions for this section, we pessimistically assume there is minimal overlap between T_{RS_1} and T_{RS_2}

Summing up the above three transmission times, using (4.14), and the simplification, $R_3 = R_4 = R_5$ as before, we find

$$R_{net,SR2} = \frac{19}{\frac{18}{R_1} + \frac{12}{R_2} + \frac{10}{R_3}} \tag{4.24}$$

A similar approach is used to find the aggregate throughput for SR1.

$$R_{net,SR1} = \frac{19}{\frac{18}{R_1} + \frac{6}{R_2} + \frac{4}{R_3}} \tag{4.25}$$

The denominator is smaller for SR1, since there is more opportunity for simultaneous transmissions. The expressions derived for all cases, one to four hops, are shown in Table 4.3. As in the above three hop example, it has been assumed that all last hop throughputs are equal. Note that in the one hop case, no spatial reuse is possible, and in the two hop case, SR2 is not possible.

These simple expressions, derived with the simplifying assumptions described above, are useful for off-line system design. In a live network the link conditions change continually, so a scheduling mechanism must be more sophisticated, taking into account any arbitrary

Table 4.3: Summary of R_{net} calculations.

System	Manhattan	Hexagonal
1 hop, no SR	R_1	R_1
2 hop, no SR	$5\left[\frac{4}{R_1}+\frac{5}{R_2}\right]^{-1}$	$7\left[\frac{6}{R_1}+\frac{7}{R_2}\right]^{-1}$
2 hop, SR1	$5\left[\frac{4}{R_1}+\frac{3}{R_2}\right]^{-1}$	$7\left[\frac{6}{R_1}+\frac{3}{R_2}\right]^{-1}$
3 hop, no SR	$13\left[\frac{12}{R_1}+\frac{8}{R_2}+\frac{13}{R_3}\right]^{-1}$	$19\left[\frac{18}{R_1}+\frac{12}{R_2}+\frac{19}{R_3}\right]^{-1}$
3 hop, SR2	$13\left[\frac{12}{R_1}+\frac{8}{R_2}+\frac{5}{R_3}\right]^{-1}$	$19\left[\frac{18}{R_1}+\frac{12}{R_2}+\frac{10}{R_3}\right]^{-1}$
3 hop, SR1	$13\left[\frac{12}{R_1}+\frac{4}{R_2}+\frac{4}{R_3}\right]^{-1}$	$19\left[\frac{18}{R_1}+\frac{6}{R_2}+\frac{4}{R_3}\right]^{-1}$
4 hop, no SR	$25\left[\frac{24}{R_1}+\frac{20}{R_2}+\frac{12}{R_3}+\frac{25}{R_4}\right]^{-1}$	$37\left[\frac{36}{R_1}+\frac{30}{R_2}+\frac{18}{R_3}+\frac{37}{R_4}\right]^{-1}$
4 hop, SR2	$25\left[\frac{24}{R_1}+\frac{20}{R_2}+\frac{5}{R_4}\right]^{-1}$	$37\left[\frac{36}{R_1}+\frac{30}{R_2}+\frac{7}{R_4}\right]^{-1}$
4 hop, SR1	$25\left[\max(\frac{24}{R_1},\frac{2}{R_3})+\frac{10}{R_2}+\frac{1}{R_3}+\frac{4}{R_4}\right]^{-1}$	$37\left[\max(\frac{36}{R_1},\frac{2}{R_3})+\max(\frac{5}{R_2},\frac{2}{R_3})+\frac{4}{R_4}\right]^{-1}$

set of rates on the links and finding the best schedules on an ongoing basis. Section 4.2.3 describes a scheduling system using an optimization method.

4.2.3 Scheduling and Throughput Regions for Multihop Relaying

Much work exists on the theoretical capacity of ad-hoc relay networks ([26, 95–99] are a few examples), generally under a number of simplifying assumptions. Bounds on the capacity of ad-hoc networks have been derived for classes of networks with random topologies, where the bounds hold with high probability as the network size gets large (asymptotically). In a network containing numerous nodes, each pair of which may communicate, the network capacity can be described by capacity regions. *Capacity* is defined as the maximum rate at which data transmission is achievable between two nodes. With N_n nodes in a network, there are $N_n(N_n-1)$ rates between node pairs. The set of all such rate combinations is called the *capacity region* and has $N_n(N_n-1)$ dimensions. The shape of capacity regions for wireless networks depends on numerous factors: data transmission schedule, propagation environment, etc., and is very difficult to derive analytically. Capacity bounds from the cited work are useful for classes of random and arbitrary ad-hoc networks, but do not give specific design rules or actual throughput for a realistic cellular system. The goal of our work is to develop design rules that could be incorporated into system design software. In order to do so, we find specific performance of networks by calculating SINRs for numerous system topologies and cluster sizes (using parameters from 802.16 and HiperMAN), finding each link's throughput using AMC, and calculating the resulting network throughput. Overhead in the physical and medium access control layers is included to determine usable *throughput*. We define a *throughput region*, similar to the capacity region, as the set of all usable throughputs between nodes or groups of nodes. Each point in the throughput region maps

to a particular *schedule*. It is sensible then to choose a schedule that maximizes the network throughput, and meets some criteria for service provision. Calculations for hexagonal and Manhattan tree topologies have been done for two to four relay hops. Throughput regions for a three-hop hexagonal cellular layout are presented as an example.

4.2.4 Calculation of Schedules and Throughput Regions

In order to calculate throughput regions [51], we have adapted the rate matrix approach from [100]. A brief outline of this approach as used in our research is presented here[2]. In order to describe the rates at which nodes in a general network may transmit to other nodes, a *link rate matrix*, $\mathbf{R}_L$, is used

$$\mathbf{R}_L = \{R_{L,mn}\} : m, n\epsilon[1, N_n], m \neq n \tag{4.26}$$

where N_n is the number of nodes (BS, RSs and MSs) in a network, and $R_{L,mn} \geq 0$ is the rate achievable between each transmitter m and receiver n in a network. The $R_{L,mn}$ are chosen according to any suitable channel model, such as Rayleigh, Ricean, etc. For each such "snapshot" of the channel conditions in the network, a schedule can be determined making best use of the inter-station links. For the results presented here, we use throughput or normalized throughput calculated as described in Section 4.1. When $R_{L,mn} = 0$ a usable link between nodes m and n does not exist. We also need to describe all possible *transmission schemes* available in a multihop relay network. A transmission scheme describes which node (or nodes) transmit simultaneously to which other node (or nodes) during a specific scheduling slot. The description of a transmission scheme for a multihop network must contain the following information: the transmitting node(s), the receiving node(s), the rate(s) of transmission, and the original source(s) of the data being transmitted during the slot. A *transmission scheme rate matrix*, $\mathbf{R}_k$, contains all of this information for the k^{th} transmission scheme. For a given network, there exist numerous (N_s) different transmission schemes $\{\mathcal{S}_k\}$ with corresponding rate matrices

$$\mathbf{R}_k = \{R_{ijk}\} : k\epsilon[1, N_s], i, j\epsilon[1, N_n], i \neq j \tag{4.27}$$

For each $\mathcal{S}_k$, with node A_i being the *original* data source,

$$R_{ijk} = \begin{cases} R & \text{if node } A_j \text{ receives at rate } R, \\ -R & \text{if node } A_j \text{ transmits at rate } R, \\ 0 & \text{otherwise.} \end{cases} \tag{4.28}$$

Each value of R is drawn from matrix $\mathbf{R}_L$ for the appropriate inter-node link. For example, if node 2 can transmit to node 4, then $R_{i2k} = -R_{L,24}$ and $R_{i4k} = R_{L,24}$. Each

[2]For more details on this method please refer to [100]

matrix (non-SR rate matrix), $\mathbf{R'}_k$, of a non-spatial reuse transmission scheme will contain only two non-zero elements, and they will be in the same row, since only one node may transmit at a time to one other node. With spatial reuse, multiple pairs of non-zero elements will exist in the matrix since more than one node may transmit simultaneously.

All non-SR transmission scheme rate matrices, $\mathbf{R'}_k : k\epsilon[1, N'_s]$, can be easily constructed for a particular network topology. Obviously, N'_s becomes very large for large networks. Next, spatial reuse rate matrices (SR rate matrices), $\mathbf{R}_k : k\epsilon[1, N_s]$, can be derived with the aid of a *compatibility matrix*, described by [101]

$$\mathbf{M_C} = \{M_{C,pq}\}, p, q\epsilon[1, N_n] \tag{4.29}$$

This matrix describes which nodes may transmit simultaneously without creating excessive interference for each other. Hence, it describes spatial reuse opportunities in the network. This matrix is geometry and topology-dependent, and its elements are

$$M_{C,pq} = \begin{cases} 1 & \text{if nodes } A_p,\ A_q \text{ may transmit simultaneously,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

A spatial reuse compatibility matrix $\mathbf{M_C}$ can be easily created by looking at the network topology. We wrote a small Matlab routine that generates all N_s matrices, $\mathbf{R}_k$, from $\mathbf{M_C}$ and $\mathbf{R}_L$.

The question now is to find a schedule that i) makes the best use of the link rates, and ii) makes the best use of spatial reuse opportunities in delivering all data from data sources to destinations. This can be formulated as a linear convex optimization problem. All possible transmission schemes, $\{\mathcal{S}_k\}$, are described by the set of rate matrices, $\{\mathbf{R}_k\}$. We now wish to determine what is the best fraction of time to allocate to each scheme. Let vector $\bar{a} = [a_1 a_2 ... a_{N_s}]$ describe the schedule, with a_k being the fraction of time that scheme $\mathcal{S}_k$ is allocated in one complete schedule cycle. We note that

$$0 \leq a_k \leq 1, \forall k\epsilon[1, N_s] \tag{4.31}$$

$$\sum_{k=1}^{N_s} a_k = 1 \tag{4.32}$$

Once the schedule, $\bar{a}$, is determined, a net rate matrix, $\mathbf{R}$, describing the net data flow from sources to destinations in the network as a result of that schedule, is calculated as

$$\mathbf{R} = \sum_k a_k \mathbf{R}_k \tag{4.33}$$

Whereas the individual matrix, $\mathbf{R}_k$, describes the transmission scheme for a scheduling slot, $k\epsilon[1, N_s]$, the net rate matrix, $\mathbf{R}$, contains information about end-to-end throughputs

that occur from each data source to each data sink after the completion of a complete scheduling cycle. This matrix contains the following elements

$$R_{ij}\begin{cases}<0 & \text{node } A_j \text{ is a net source of data,}\\ >0 & \text{node } A_j \text{ is a net sink of data,}\\ =0 & \text{node } A_j \text{ is acting as a relay only.}\end{cases} \tag{4.34}$$

These follow from the definitions in (4.28). A negative number indicates that node A_j is a transmitter (a source of data in the network wide context) while a positive number indicates a receiver (sink of data). Since a relay is simply passing on data, it is neither a source nor a sink of data and the matrix element must be zero.

As before, the row index, i, indicates the original data source and the column index, j, indicates the active node, so if $\mathbf{R}$ is to be a correct description of the network data flow, we must have

$$R_{ij}\begin{cases}\leq 0 & \text{when } i=j,\\ \geq 0 & \text{when } i\neq j.\end{cases} \tag{4.35}$$

This again follows from (4.28). When $i = j$, node A_j can only be a data source or a relay. If it is a data source then by (4.28) the matrix element must be negative, otherwise it is a relay and then the element must be zero. When $i \neq j$, node A_j is either a data sink in which case the matrix element must be positive, or a relay in which case the element must be zero.

Numerous solutions are possible depending on the specific formulation of the optimization problem. We have constraints (4.31) and (4.32) and further constraints $R_{ij} = 0$ for nodes acting as relays (from (4.34)). At first, it seems sensible to use the network sum throughput, R_{net} as the objective to maximize:

$$R_{net} = \sum_{i,j:R_{ij}>0} R_{ij} \tag{4.36}$$

Although the solution results in the highest network spectral efficiency, the resulting schedule will always favour one hop MSs (those MSs in close proximity to the BS) over multihop MSs (those MSs closer to the cell edge), which is unfair to MSs, and virtually useless. We can force equal throughput to all MSs by adding another set of constraints

$$R_{i,j} = R_{k,l} : i \neq j, k \neq l \tag{4.37}$$

when nodes i or j, and k or l are MSs.

Table 4.4: Hop distances (in m) for a 1 km radius Manhattan cell.

No. Hops	Hop 1	Hop 2	Hop 3	Hop 4
1	1000 NLOS	-	-	-
2	667 NLOS	333 NLOS	-	-
3	400 NLOS	400 NLOS	200 LOS	-
4	286 LOS	286 LOS	286 LOS	143 LOS

4.2.5 Aggregate Network Rate, Throughput and AASE of Multihop Cellular Networks

Using the system model and the methods described in the preceding sections, we performed a large number of calculations for Manhattan and hexagonal cellular topologies. We present here a subset of these results, using a few different examples for illustration. Numerous tables of $SINR$, network throughput (R_{net}) and $AASE$ values have been calculated for Manhattan and hexagonal cells using the parameters listed in Table 3.1. Since there are a large number of results (for all combinations of Manhattan and hexagonal cells, four different cell sizes, from one to four hops, five different antenna configurations and several different cluster sizes), we present here details of one example, and overall results for the remainder of the cases.

Example: Manhattan Cell

For a Manhattan cell with a radius of $r = 1$ km, antenna configuration 5, and a distance breakpoint of 300 m, several possible multihop designs are compared. We compare non-spatial reuse with the spatial reuse case SR1. Table 4.4 gives the hop distances for each multihop case. Recall that according to our dual slope path loss model, links are LOS if they are shorter than 300 m.

Table 4.5 gives the resulting rates using the capacity formula (2.10). The base case for comparison is Case A (no relaying, no spatial reuse is possible).

First consider the no-SR cases. The addition of relaying, with $n = 2$ (design B), link rates are increased somewhat due to the shorter distances (see Table 4.4), but the link rate increases are limited since the hops are all still NLOS. The aggregate network throughput is slightly worsened since some spectral resources are wasted due to relaying. The addition of a second relay (design C: $n = 3$) provides some improvement, but the greatest improvement is with four hops. In this case, all links become LOS, and thus link rates are significantly increased.

With SR, the improvement with multihop is more dramatic. The two-hop design provides

Table 4.5: R_{net} and $AASE$ for Manhattan design options, rates calculated using capacity formula, $r = 1$ km, antenna configuration 5 (Table 3.2).

Design		N	R_{net} (b/s/Hz)	$AASE$ (b/s/Hz/km^2)	Paths
	No SR				
A	$n = 1$	3	1.982	0.165	NLOS
B	$n = 2$	3	1.820	0.273	NLOS
C	$n = 3$	3	2.586	0.414	NLOS/LOS
D	$n = 4$	3	**5.603**	0.915	LOS
E	$n = 1$	1	1.498	0.375	NLOS
F	$n = 2$	1	0.957	0.431	NLOS
G	$n = 3$	1	1.387	0.667	NLOS/LOS
H	$n = 4$	1	4.503	**2.207**	LOS
	With SR				
B	$n = 2$	3	2.089	0.313	NLOS
C	$n = 3$	3	3.525	0.565	NLOS/LOS
D	$n = 4$	3	**11.375**	1.858	LOS
F	$n = 2$	1	1.072	0.482	NLOS
G	$n = 3$	1	1.807	0.869	NLOS/LOS
H	$n = 4$	1	9.276	**4.545**	LOS

a slight improvement over the single hop design, since only limited SR is possible. Greater increases in aggregate network rate occur with the three and four hop designs. The best SR network rate results from design D ($n = 4$, $N = 3$) with 11.375 b/s/Hz, which is 5.7 times the throughput of the single hop design. However, when we consider the $AASE$, the best design is H ($n = 4$, $N = 1$). Although co-channel interference is higher with universal frequency reuse, the network rate pays only a slight penalty, and the system-wide efficiency of spectrum use increases dramatically. $AASE$ in case H is 4.545 b/s/Hz/km^2, which is more than a 27-fold improvement over the single hop design. A design with $N = 1$ does not in all cases result in the highest $AASE$. We have found that in some geometries the penalty paid by higher co-channel interference is greater than the gain due to relaying.

If we look at incremental improvement, the greatest gains occur when all of the hops are short enough to be LOS. LOS lowers the path loss significantly and thus the link rates improve dramatically. Further addition of relays beyond that which achieves LOS is not as beneficial, since there will only be slight further improvement in link rates, and slight increase in network rate. In this example, four hops is the best design, but the conclusions will be different for different cell sizes.

Now we look at this design using the more realistic throughputs obtainable by AMC, including the effects of overhead. Table 4.6 gives the throughputs for the same designs as above. Obviously the throughputs are much lower than the rates obtained from the link capacity formula, but as before, design H is the best, giving a 34-fold improvement in $AASE$ over the single hop design. Note that design F resulted in no throughput. In that case, at least one of the hops did not have the minimum SINR (3 dB) required to transmit at the lowest AMC level, and so the MSs are in an outage condition.

Table 4.6: R_{net} and $AASE$ for Manhattan design options, using AMC, $r = 1$ km, antenna configuration 5 (Table 3.2).

Design		N	R_{net} (b/s/Hz)	$AASE$ (b/s/Hz/km^2)	Paths
	No SR				
A	$n = 1$	3	0.290	0.024	NLOS
B	$n = 2$	3	0.520	0.078	NLOS
C	$n = 3$	3	0.598	0.096	NLOS/LOS
D	$n = 4$	3	**0.812**	0.133	LOS
E	$n = 1$	1	0.290	0.072	NLOS
F	$n = 2$	1	0.000	0.000	NLOS
G	$n = 3$	1	0.335	0.161	NLOS/LOS
H	$n = 4$	1	**0.812**	**0.398**	LOS
	With SR				
B	$n = 2$	3	0.589	0.088	NLOS
C	$n = 3$	3	0.869	0.134	NLOS/LOS
D	$n = 4$	3	**1.686**	0.275	LOS
F	$n = 2$	1	0.000	0.000	NLOS
G	$n = 3$	1	0.454	0.218	NLOS/LOS
H	$n = 4$	1	**1.686**	**0.826**	LOS

Note that in all of the above cases, the use of directional transmit and receive antennas for all nodes (BS, RS, and MS) achieves good performance. Usually it is more practical for the MS (which may be mobile or nomadic) to have an omnidirectional antenna (antenna configuration 4). Table 4.7 shows a number of designs with an omnidirectional MS antenna. The results are similar to the previous ones, but the one-hop and two-hop cases have suffered due to the lower antenna gain at the MSs. Three and four hop cases can still maintain nearly the same network throughputs and $AASE$.

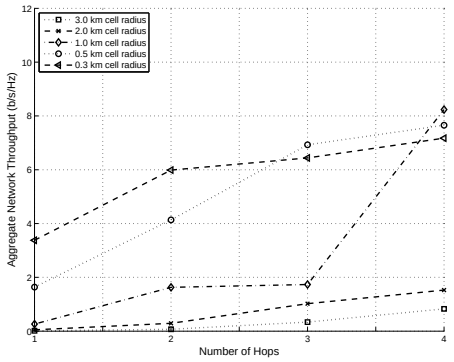
Fig. 4.4 compares the performance, R_{net} and $AASE$, for various cell sizes and numbers of relay hops for both Manhattan and hexagonal multihop cellular topologies. Antenna configuration 4 was used. Link capacities, calculated using (2.10), are used for this figure.

There are numerous conclusions that can be drawn from Fig. 4.4. For larger cells (3.0 and 2.0 km radius), the improvement in R_{net} is minimal since LOS paths are not created even with four hops. There is however a significant benefit due to spatial reuse, which is more evident in the $AASE$ shown in Figs. 4.4(c) and 4.4(d). A 1 km cell enjoys the benefits of LOS paths with four hops and the dramatic improvement is obvious. Cells with radii 0.5 km and 0.3 km see a large benefit with three and two hops respectively, again as the result of LOS relaying around obstacles. We can see from both R_{net} and $AASE$ plots that there is a levelling off after LOS is gained.

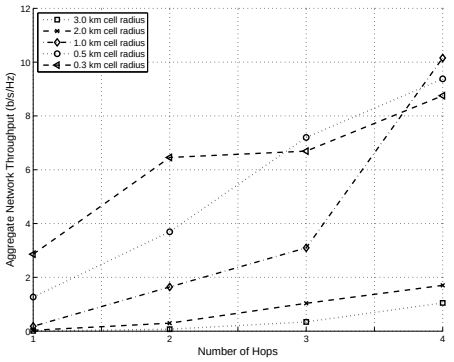
Fig. 4.5 shows the results using AMC and accounting for PHY and MAC overhead (using Table 4.1). The results are similar to the above capacity-based results, but with some slight differences due to the threshold nature of AMC. With the MSs at the cell edge (the worst case scenario) the 3 km radius cells cannot achieve the threshold SINR required to meet the BER target at the lowest AMC level, and so the MSs are in outage and the throughput for

Table 4.7: R_{net} and $AASE$ for Manhattan design options, using AMC, $r = 1$ km, antenna configuration 4 (Table 3.2).

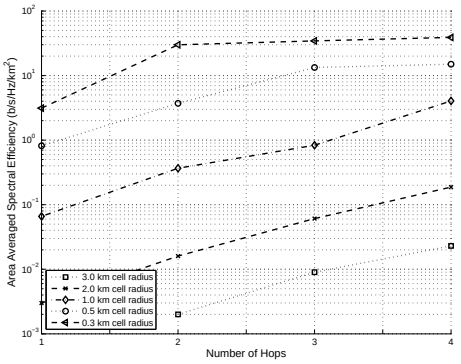
Design		N	R_{net} (b/s/Hz)	$AASE$ (b/s/Hz/km^2)	Paths
	No SR				
A	$n = 1$	3	0.000	0.000	NLOS
B	$n = 2$	3	0.401	0.060	NLOS
C	$n = 3$	3	0.598	0.096	NLOS/LOS
D	$n = 4$	3	**0.812**	0.133	LOS
E	$n = 1$	1	0.000	0.000	NLOS
F	$n = 2$	1	0.000	0.000	NLOS
G	$n = 3$	1	0.335	0.161	NLOS/LOS
H	$n = 4$	1	0.783	**0.384**	LOS
	With SR				
B	$n = 2$	3	0.491	0.074	NLOS
C	$n = 3$	3	0.869	0.139	NLOS/LOS
D	$n = 4$	3	**1.686**	0.275	LOS
F	$n = 2$	1	0.000	0.000	NLOS
G	$n = 3$	1	0.454	0.218	NLOS/LOS
H	$n = 4$	1	1.666	**0.816**	LOS



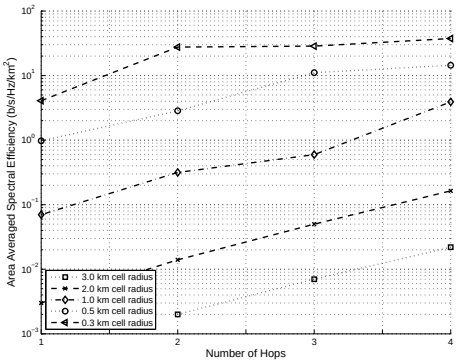
(a) Aggregate network rate, R_{net}, for Manhattan topologies.



(b) Aggregate network rate, R_{net}, for hexagonal topologies.

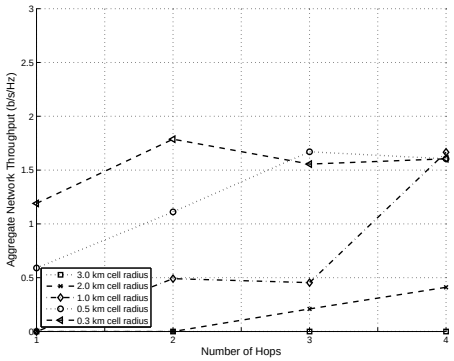


(c) Area averaged spectral efficiency, $AASE$, for Manhattan topologies.

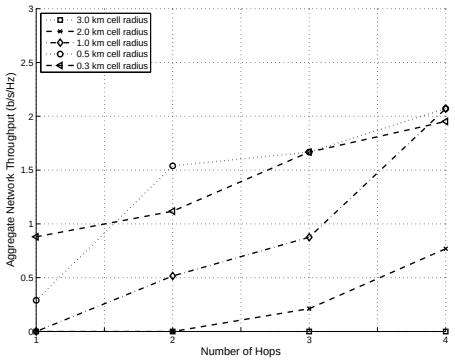


(d) $AASE$ for hexagonal topologies.

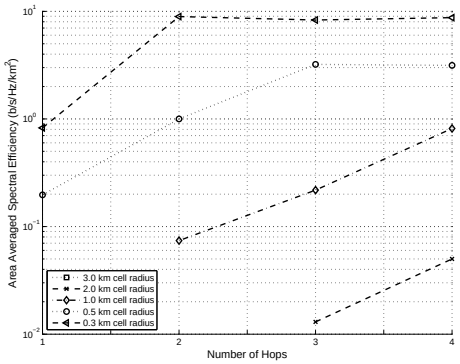
Figure 4.4: Multihop cellular results using the capacity formula. Antenna configuration 4.



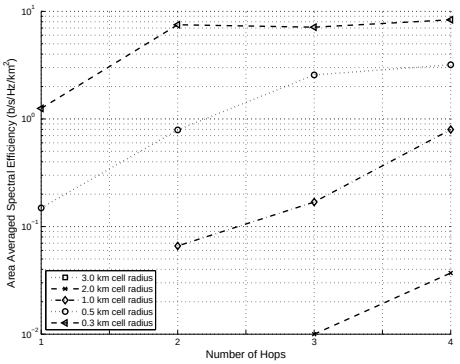
(a) Aggregate network throughput, R_{net}, for Manhattan topologies.



(b) Aggregate network throughput, R_{net}, for hexagonal topologies.



(c) Area averaged spectral efficiency, $AASE$, for Manhattan topologies.



(d) $AASE$ for hexagonal topologies.

Figure 4.5: Multihop cellular results using throughputs resulting from adaptive modulation and coding (AMC). Antenna configuration 4.

that size cell is 0 even with four hops. Likewise, a minimum of three hops is required for the 2 km radius cell, and two hops for the 1 km radius cell, to meet the lowest SINR threshold. Note also that for the 0.5 km radius Manhattan cell, the highest R_{net} is achieved with three hops. The addition of a fourth hop does not quite increase the link SINRs enough to obtain the next highest AMC level, and the added relay wastes some additional spectrum. A similar peak occurs at two hops for the 0.3 km Manhattan cell.

It is apparent from these examples that relaying design depends greatly on the specific propagation environment and the geometry.

Throughput Regions: Three Hop Hexagonal Example

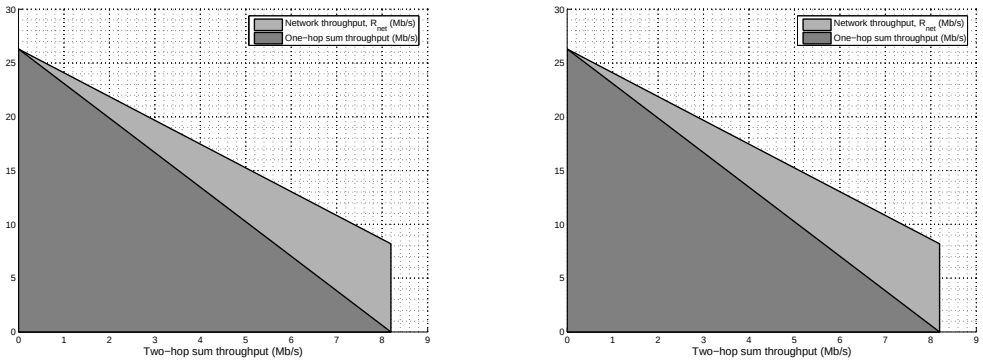
We have performed calculations for two, three and four hop topologies, using both hexagonal and Manhattan layouts. The example presented here uses the three hop network shown in Fig. 4.1 and considers the forward links from the BS to MSs. We lump MSs into one MS

per subcell and calculate sum throughput per subcell. Using the parameters in Table 3.1 with a cell of 1 km circumscribed radius and a channel bandwidth of 10 MHz (see [52]), an example set of link throughputs are $R_1 = R_2 = 11.9$ Mb/s, $R_3 = R_4 = R_5 = 26.3$ Mb/s. With symmetry and the BS as the single data source, the general formulation described in Section 4.2.3 simplifies greatly. This results in a network with one data source and 38 data sinks, so the throughput region has 38 dimensions. In order to display results that are easier to visualize, we sum up the throughput to MSs in each of the one-hop, two-hop and three-hop regions. For example, the two-hop sum throughput gives the total throughput to all MSs in all subcells that are in the inner ring in Fig. 4.1. This results in throughput regions with only three dimensions. We also show the aggregate network throughput, which is the effective throughput in the whole cell, taking into account additional resources required for multihop relaying.

Fig. 4.6(a) shows a two dimensional slice of the one-hop MS to two-hop (inner-ring) sum throughput region, and the aggregate network throughput with no spatial reuse. The line represents the upper boundary of the throughput region. All combinations of throughputs below that line are possible. At the point where maximum network throughput is achieved, no transmission occurs to the outer MSs, and although high in spectral efficiency, this network has poor coverage since it is serving only 1/19 of the cell area. It is necessary to trade network throughput for better (and fair) coverage, which means choosing a point on the region boundary line lower to the right. Fig. 4.6(b) shows the throughput region when spatial reuse is activated in the simulation algorithm. Spatial reuse provides no improvement in this case since, as discussed in Section 4.2.2, we have assumed that two-hop neighbours must coordinate, and thus no spatial reuse is possible within a two-hop distance from the BS.

Fig. 4.7(a) shows a two dimensional slice of the one-hop MS to three-hop (outer-ring) throughput region, and the network sum throughput with no spatial reuse. It shows a similar tradeoff. In this slice of the throughput region, enabling spatial reuse (Fig. 4.7(b)) does show improvement (the boundary has moved to the right) since transmissions in the outer ring can occur simultaneously with transmissions elsewhere in the network. And finally Fig. 4.8(a) and Fig. 4.8(b) show the results for two-hop MS (inner ring) to three-hop (outer-ring), without and with spatial reuse respectively. Again, spatial reuse allows improvement in the network throughput.

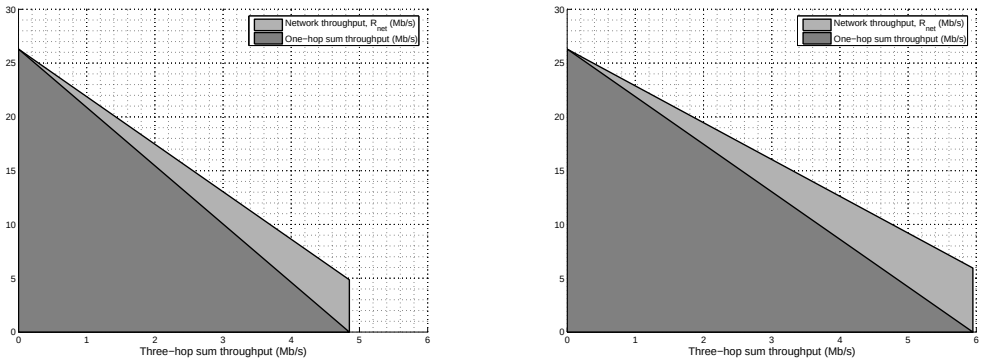
Fig. 4.9 shows the effects of spatial reuse. As three-hop MSs are served, there are opportunities for more than one node to transmit simultaneously and so we see the throughput region expand to the right along the horizontal axis.



(a) Without spatial reuse.

(b) With spatial reuse.

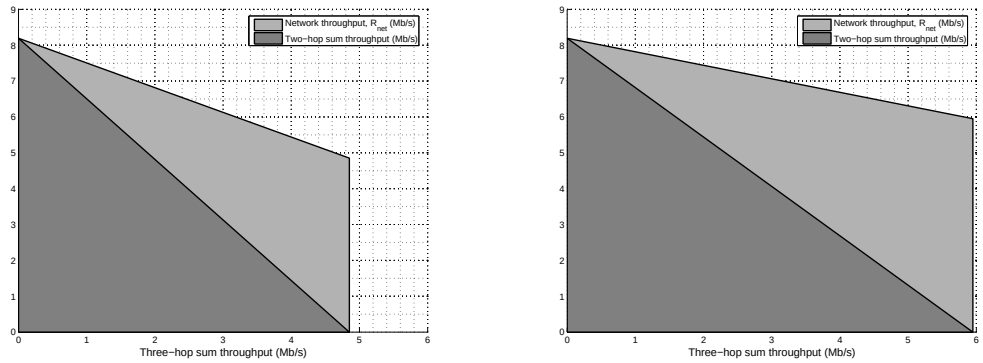
Figure 4.6: Throughput regions - one-hop vs inner ring (two-hop) nodes (y-axis labels are given in the legend).



(a) Without spatial reuse.

(b) With spatial reuse.

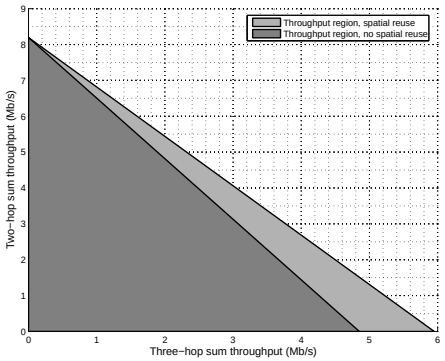
Figure 4.7: Throughput regions - one-hop vs outer ring (three-hop) nodes (y-axis labels are given in the legend).



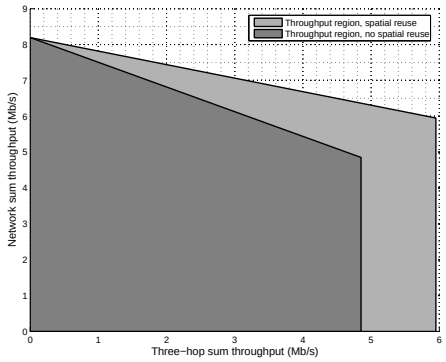
(a) Without spatial reuse.

(b) With spatial reuse.

Figure 4.8: Throughput regions - inner ring (two-hop) vs outer ring (three-hop) nodes (y-axis labels are given in the legend).



(a) Inner ring (two-hop) nodes vs outer ring (three-hop) nodes.



(b) Network throughput, R_{net}.

Figure 4.9: Throughput regions - with and without spatial reuse.

4.3 Diversity and Capacity in Ricean/Nakagami Channels

This section provides some analytical background on wireless cooperative diversity. As discussed in Chapter 2, relays in a wireless network can be used in two different ways. First, multihop relaying simply reduces the path loss and scattering components on a wireless link between a data source and a destination. Second, the existence of scattered relays creates multiple branches via which data can be sent. The existence of multiple branches enables the use of diversity to provide more reliable links. This section looks at some error rate and capacity aspects of these two options, enlarging upon the introductory discussion in Chapter 2.

First we look at bit error rate and capacity performance of diversity. With MRC, diversity branches are combined coherently to improve the reliability of fading channels. A key measure of channel reliability is the probability of a bit error. The bit error probability for MRC with BPSK in a Rayleigh channel is [27]

$$P_b^{MRC}(\bar{\gamma}_c) = \left(\frac{1-\mu}{2}\right)^d \sum_{k=0}^{d-1} \binom{d-1+k}{k} \left(\frac{1+\mu}{2}\right)^k \tag{4.38}$$

where

$$\mu = \sqrt{\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}} \tag{4.39}$$

and $\bar{\gamma}_c$ is the average symbol energy to noise ratio on each diversity branch (assuming equal $\bar{\gamma}_c$ on each branch), d is the number of diversity branches. The bit error probability for SC with BPSK in a Rayleigh channel is [27]

$$P_b^{SC}(\bar{\gamma}_c) = \frac{d}{2} \sum_{k=0}^{d-1} \frac{\binom{d-1}{k}(-1)^k}{1+k+\bar{\gamma}_c} \tag{4.40}$$

Compare these to the error probability for BPSK in a non-fading channel with additive white Gaussian noise:

$$P_b^{NF}(\gamma_b) = Q(\sqrt{2\gamma_b}) \tag{4.41}$$

where γ_b is the bit energy to noise ratio.

Figure 4.10 shows the bit error rates for non-fading and Rayleigh fading with MRC and SC. With a sufficient number of diversity branches, the BER performance of MRC and SC can approach the performance of a non-fading channel. SC performs a little worse than MRC, requiring about 2 dB more SNR at a BER of 10^{-3}.

MRC can also increase the average received SNR. If the average symbol energy to noise ratio on each diversity branch is $\bar{\gamma}_c$ (assuming equal on each branch), then d diversity branches gives a symbol energy to noise ratio of

$$\bar{\gamma}_s^{MRC} = d\bar{\gamma}_c \tag{4.42}$$

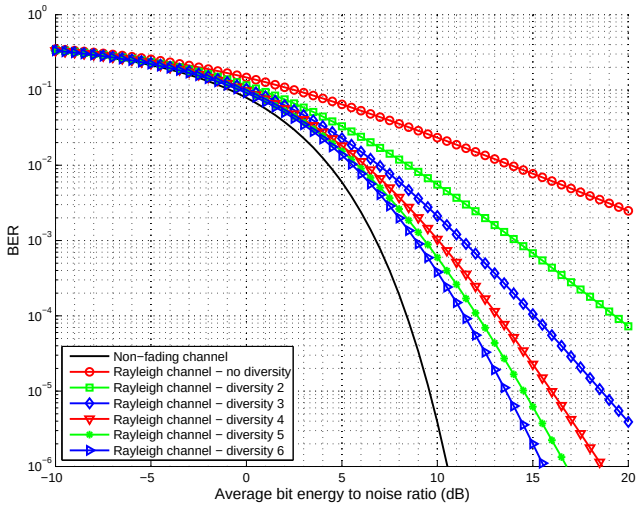
The capacity of the link is therefore

$$C^{MRC} = \log_2(1 + d\bar{\gamma}_c) \tag{4.43}$$

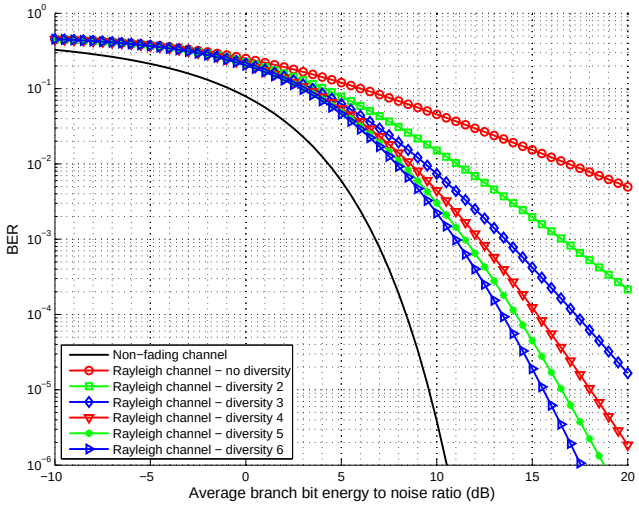
This is an improvement, but the capacity scales with the logarithm of d, so there are diminishing returns as d increases. However, this does not take into consideration the penalty in using multiple branches for diversity instead of for enhancing capacity. With repetition coding, d branches could have been used to transmit additional data, so the penalty in using diversity is $1/d$. The capacity is reduced to

$$C^{MRC} = \frac{1}{d}\log_2(1 + d\bar{\gamma}_c) \tag{4.44}$$

This can be improved with the use of space-time codes. However, most space-time codes are designed to maximize the diversity gain and coding gain, but not capacity. MRC can be used with AF relaying if the channel amplitudes and phases of the diverse branches are known. But since DF involves decoding and re-encoding at each hop, channel information for all hops except the last hops are lost. MRC involves coherently combining the diverse signals, so DF cannot use coherent combining until the last hop. Variations in path conditions on previous hops do not participate in MRC, which degrades the performance. SC makes sense for DF cooperative relaying, since the best branch can be selected for transmission based on end-to-end throughputs of each branch. With a block fading assumption (the channels



(a) BER of MRC in a Rayleigh channel.



(b) BER of SC in a Rayleigh channel.

Figure 4.10: Bit error rate performance of MRC and SC in a Rayleigh fading channel compared to a non-fading channel (BPSK).

on the hops do not change during a short block of time) the best branch can be selected, and only that branch is used for data transmission during that time block. This provides a significant saving in spectrum consumption. The average symbol energy to noise ratio for SC is

$$\bar{\gamma}_s^{SC} = \bar{\gamma}_c \sum_{k=1}^{d} \frac{1}{k} \tag{4.45}$$

With out-of-band transmission of channel state information at the transmitter (CSIT), the best branch can be chosen ahead of time, and data can be transmitted only on that branch. The capacity of the link with SC is then

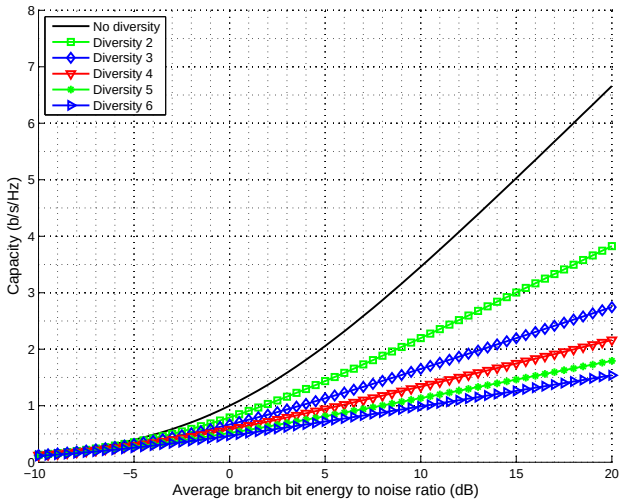
$$C^{SC} = \log_2(1 + \bar{\gamma}_c \sum_{k=1}^{d} \frac{1}{k}) \tag{4.46}$$

The capacity performance of MRC and SC in Rayleigh fading is shown in Fig. 4.11. So although SC performs slightly worse in BER, it can be more easily used in cooperative relaying, and with channel knowledge and advanced branch selection, SC has better spectral efficiency.

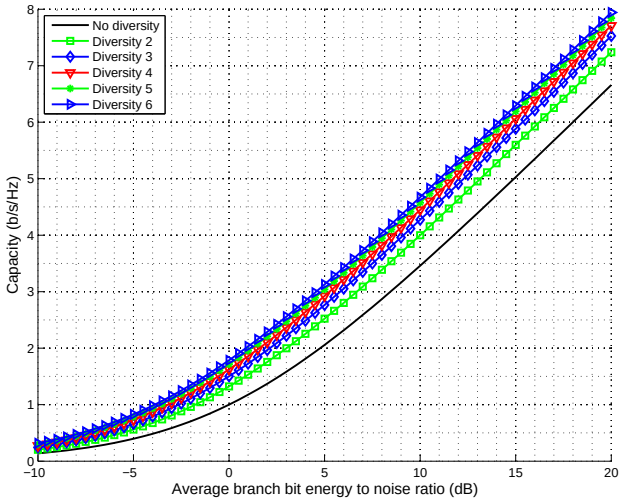
The preceding analysis has considered a NLOS Rayleigh fading channel, but as discussed earlier, cellular channels are more accurately modelled as LOS [27]. With the presence of a specular component diversity gains are diminished, but the path loss is reduced at the same time. Here we compare the BER performance of diversity using in a Nakagami channel. A Nakagami channel model [27, 102] is an alternate method of modelling a LOS channel, the form of which often lends itself more easily to analysis than the Ricean form. Instead of a Rice factor, K_r, it has a shape factor, m. The Ricean model can be approximated by a Nakagami model with the substitution [27]

$$m = \frac{(K_r + 1)^2}{2K_r + 1} \tag{4.47}$$

and a Rayleigh channel can be obtained with $m = 1$. In [103] an equivalence was shown between the BER performance of diversity in a Rayleigh channel and performance in a Nakagami channel. With a Nakagami shape factor m, the BER performance of d diversity branches in a Nakagami channel is equivalent to the performance of md diversity branches in a Rayleigh channel. For example, a typical Rice factor in a cellular system is $K_r = 6$ dB, which according to (4.47) is equivalent to $m = 2.8$ with a Nakagami model. So with no diversity, this Nakagami channel has almost the same BER performance characteristic as a Rayleigh channel having 3 diversity branches. If we were to use 3 diversity branches in this Nakagami channel, it would have nearly the same performance as 8 branches in a Rayleigh channel. Thus, as K_r and m increase, diversity combining techniques become much less useful.



(a) Capacity of MRC in a Rayleigh channel.



(b) Capacity of SC in a Rayleigh channel with out-of-band branch selection.

Figure 4.11: Capacity performance of MRC and SC diversity schemes.

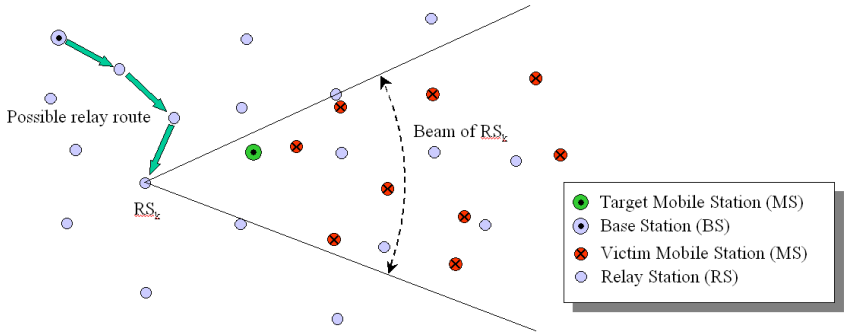


Figure 4.12: Generic system model.

4.4 Interference Avoidance

One of the many problems associated with multihop relaying is selecting the best relay from which to serve a particular MS. A simple method is to choose the relay for which the channel to the MS is the best. However, considering that there are a number of MSs in the cell which may suffer from co-channel interference, it seems sensible to modify this criterion. In this section we discuss an algorithm that attempts to find the best RS or RSs based on channel conditions to the target MS and the interfering effect the RS or RSs would have on other MSs in the cell.

4.4.1 Relay Selection Model

Fig. 4.12 shows a generic system model, with numerous RSs under control of a single BS as potential candidates for transmission to the target MS. There exists a transport route (thick green arrows) carrying data to the selected RSs, and there are possible access paths from the RSs to the target mobile (for clarity, only one such transport route is shown in the figure). The algorithms presented here are concerned with the access paths - selecting the most appropriate RS or set of RSs from which to transmit.

The normalized path gain from each RS_k to the target MS is

$$\gamma_{t,k} = \zeta_{t,k} d_{t,k}^{-\beta}, k = 1, 2, ...b \tag{4.48}$$

where $\zeta_{t,k}$ is a random variable modeling multipath fading, β is the path loss exponent, $d_{t,k}$ is the distance from the target MS to RS_k, and b is the number of RSs considered as potential candidates.

With P_k as the transmit power from RS_k, the total normalized power received at the MS is

$$P_r = \sum_{k=1}^{b} \gamma_{t,k} P_k \tag{4.49}$$

The receiver will require a minimum received power, P_{min}, to meet the bit error rate

requirement. This results in the constraint

$$P_{min} = \sum_{k=1}^{b} \gamma_{t,k} P_k \tag{4.50}$$

Each RS_k will cause interference I_k for a set of victim MSs, S_k. The victim MSs (in TDMA, those using the same time slot as the target mobile, or in CDMA, those using the same spreading code as the target mobile) in the set will be determined by the distance of the victims from RS_k and the angle and direction of the antenna beam.

$$I_k = \sum_{j \in S_k} \gamma_{v,jk} P_k = P_k \sum_{j \in S_k} \gamma_{v,jk} \tag{4.51}$$

where

$$\gamma_{v,jk} = \zeta_{v,jk} d_{v,jk}^{-\beta}, k = 1, 2, ...b \tag{4.52}$$

is the path gain between victim mobile j and its interference source (RS_k), $d_{v,jk}$ is the distance from RS_k to victim mobile j, and $\zeta_{v,jk}$ is a random variable modeling multipath fading.

The total normalized interference caused by all RSs is given by

$$I_{total} = \sum_{k=1}^{b} I_k = \sum_{k=1}^{b} \sum_{j \in S_k} \gamma_{v,jk} P_k \tag{4.53}$$

Solving for P_k in (4.51) and substituting into (4.50), we have

$$P_{min} = \sum_{k=1}^{b} I_k \frac{\gamma_{t,k}}{\sum\limits_{j \in S_k} \gamma_{v,jk}} \tag{4.54}$$

$$= \sum_{k=1}^{b} I_k \frac{\gamma_{t,k}}{\gamma_{v,k}} \tag{4.55}$$

where

$$\gamma_{v,k} = \sum_{j \in S_k} \gamma_{v,jk} \tag{4.56}$$

Before describing the proposed algorithms, it is necessary to describe two algorithms which will serve as the benchmarks for comparison. The following two benchmark algorithms have been adapted from [104], since they were intended for a similar system, and are fair comparisons to the proposed algorithms.

Benchmark Algorithm: Single Antenna Selection (SAS)

A simple algorithm for determining transmit powers P_k, $k = 1, 2, ...b$ minimizes the total transmit power

$$P_{total} = \sum_{k=1}^{b} P_k \tag{4.57}$$

subject to constraint (4.50). This results in a linear programming problem, the solution to which simply selects the RS that corresponds to the largest $\gamma_{t,k}$. For simplicity, and with no loss of generality, assume the $\gamma_{t,k}$ are ordered $\gamma_{t,1} > \gamma_{t,2} > ... > \gamma_{t,b}$. The result then is $P_1 = P_{min}/\gamma_{t,1}$, and $P_k = 0$ for $2 \leq k \leq b$.

Another way of formulating this result is

$$P_k = \frac{\gamma_{t,k}^{\alpha}}{\sum_{l=1}^{b} \gamma_{t,l}^{\alpha+1}} P_{min}, \alpha \to \infty \tag{4.58}$$

By minimizing (4.57), this algorithm expects to minimize interference averaged over all possible co-channel MS distributions. For a particular MS distribution, however, the interference is not necessarily minimized.

Benchmark Algorithm: Generalized Power Control (GPC)

The above algorithm results in only one RS transmitting at one time. However, it would be beneficial to allow more than one to transmit in order to take advantage of multiple antenna techniques. The author of [104] presented a simulcast system similar to that described herein, and proposed to improve upon the above algorithm by introducing an adaptation parameter $0 \leq \kappa \leq 1$ in the following manner

$$P_k = \left(\frac{\gamma_{t,k}^{\alpha}}{\sum_{l=1}^{b} \gamma_{t,l}^{\alpha+1}} \right)^{\kappa} P_{c1}, \alpha \to \infty \tag{4.59}$$

where P_{c1} is a scaling factor. In general, this results in $P_k \neq 0 \ \forall k$. It was found in [104] that $\kappa = 0.5$ gave the lowest probability of outage, and gave a 2 dB improvement in carrier to interference ratio (CIR) compared to the Single Antenna Selection algorithm.

Proposed Algorithm: Selective Transmit (ST)

The goal here is to choose the P_k values that minimize the total interference in (4.53) subject to the constraint (4.50). The solution selects the single RS corresponding to the highest $\gamma_{t,k}/\gamma_{v,k}$ which can be expressed as

$$P_k = \frac{(\gamma_{t,k}/\gamma_{v,k})^{\alpha}}{\sum_{l=1}^{b} (\gamma_{t,l}/\gamma_{v,l})^{\alpha+1}} \frac{P_{min}}{\gamma_{v,k}}, \alpha \to \infty \tag{4.60}$$

By minimizing (4.53), this algorithm will perform better since it takes into account additional known information: the paths to co-channel MSs that are victims of interference from each RS.

Proposed Algorithm: Simulcast (SC)

Once again, it is desirable to allow more than one RS to transmit simultaneously, so the ST algorithm is modified to include an adaptation parameter $0 \leq \kappa$. The resulting solution is

$$P_k = \left(\frac{\left(\gamma_{t,k}/\gamma_{v,k}\right)^{\alpha}}{\sum_{l=1}^{b} \left(\gamma_{t,l}/\gamma_{v,l}\right)^{\alpha+1}} \right)^{\kappa} P_{c2} \tag{4.61}$$

where P_{c2} is a scaling constant. Again, the general solution has $P_k \neq 0 \ \forall k$.

The adaptation parameters κ and α have been investigated. Numerous combinations of κ and α were tried. Via experimentation, it was found that the optimum results (lowest outage probability) are obtained when $\kappa = 1/\alpha$ with any $0.1 \leq \kappa \leq 1$, which corresponds to $1 \leq \alpha \leq 10$. The outage probability becomes worse when $1 \leq \kappa$.

4.4.2 Simulation Description

To show the efficacy of the proposed algorithms via simulation, a system considering three potential RSs ($b = 3$) was considered. Structuring the network with a regular RS distribution results in the system shown in Fig. 4.13. $b = 3$ is a convenient number to use since it gives good diversity gain - increasing this number beyond three has decreasing return. It also results in an equilateral triangle subcell tessellation which works in a hexagonal layout. (For Manhattan type layouts, $b = 4$ is useful, resulting in a square pattern.) For simplicity, only the closest sets of interference victims are shown in the figure. The three RSs closest to the target mobile are considered as candidates for transmission to that mobile. RSs use beam steering and channels can be reused in each triangular cell. For this simulation, 60° sectoring has been assumed.

Numerous sets of random mobile locations have been generated and the four algorithms run on each set. The CIRs for each algorithm have been calculated and the results analyzed. To investigate different system loading, sets have been generated with different "mobile probabilities" (probability that a channel is in use by a victim MS in a subcell): 10%, 70%, 90%, 100%. Path loss exponents (βs) used are 3.5, 4.0, 4.5, and 5.0.

Not included in the results are factors such as handoff gain, coding gain, and spread spectrum processing gain. It is expected that all algorithms tested will benefit similarly so the relative performance results are valid. All four algorithms simulated for the results presented here have been subjected to the same system model and parameters, and are thus compared on a fair basis.

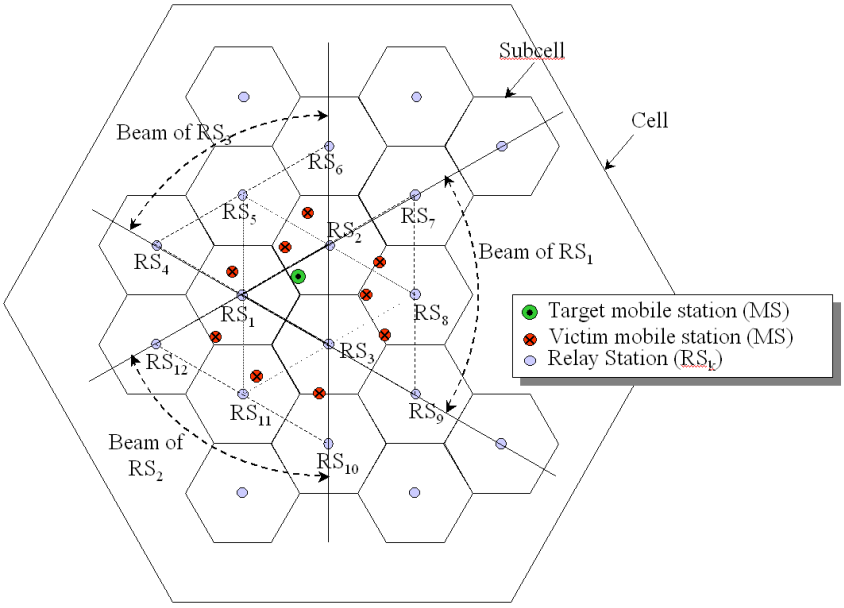


Figure 4.13: Simulation system model - example with $b = 3$.

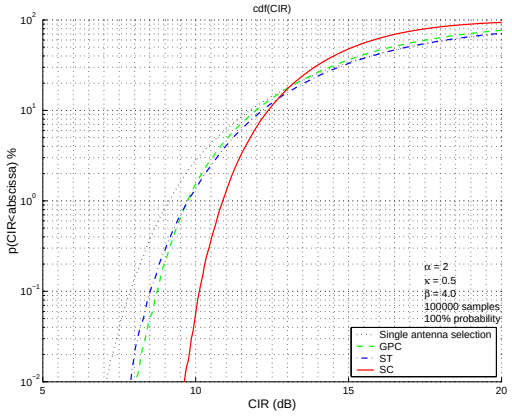
4.4.3 Simulation Results

A useful quantity to analyze when modeling random phenomena in radio networks is the CDF of the CIR. The CDF value on the vertical axis gives the probability that the CIR falls below the value on the horizontal axis, so one can directly read the probability of outage for a given CIR requirement.

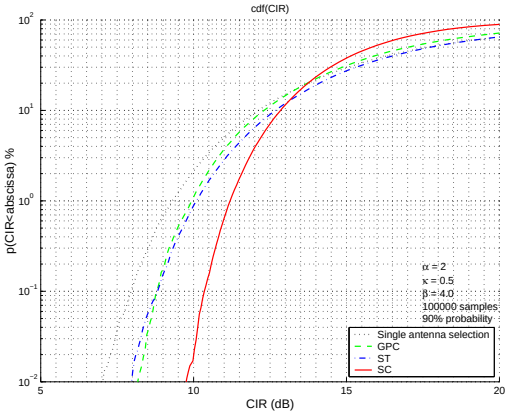
Fig. 4.14 shows the CDFs for the two benchmark algorithms and the proposed algorithms with several mobile probabilities (i.e the probability that a mobile is using a channel in subcells outside of the target subcell), and a fixed path loss exponent $\beta = 4.0$. As expected the SC algorithm very often out-performs the other ones, and both ST and SC algorithms show the best improvement when the mobile probability is low (i.e. a lightly loaded system). This is because both of these algorithms take advantage of unused channels by favouring selection of, or increasing the transmit power of, an RS pointing in the direction of subcells with channels not in use. SC performs worse than the others for higher CIR values. We believe that this is because the use of three simultaneous transmitters is more likely to cause worse interference when the CIR requirement is high. It is possible however that further tweaking of parameters κ and α would improve the performance of SC at high CIRs.

Fig. 4.15 shows the CDFs for the algorithms, with dependency on β for a fixed mobile probability of 70%. Fig. 4.16 shows the probability of outage for the algorithms, for a CIR requirement of 12 dB.

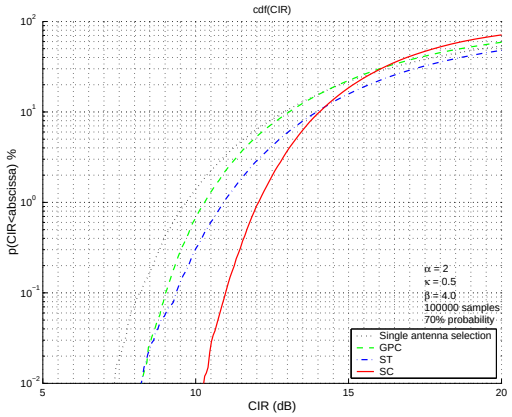
Table 4.8 shows the average CIRs (dB) for the simulation runs. Note that the average CIR was not necessarily increased with the proposed algorithms - in some cases it is slightly



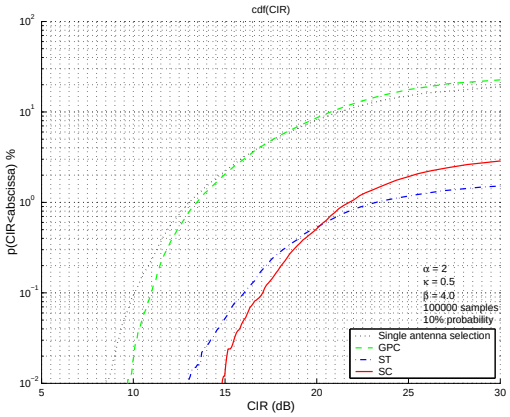
(a) Mobile channels occupied with 100% probability



(b) Mobile channels occupied with 90% probability

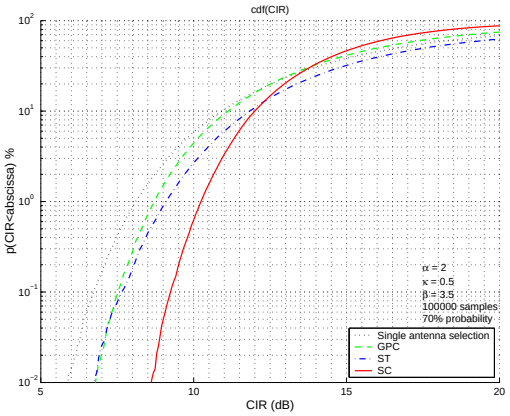


(c) Mobile channels occupied with 70% probability

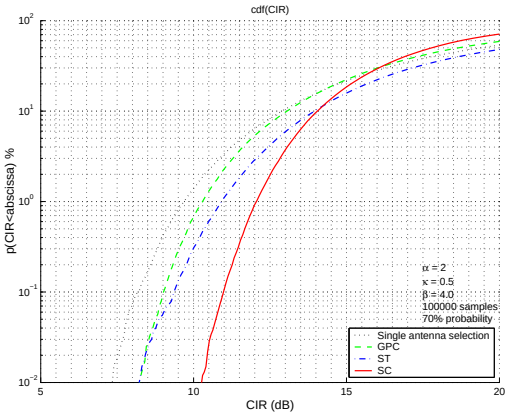


(d) Mobile channels occupied with 10% probability

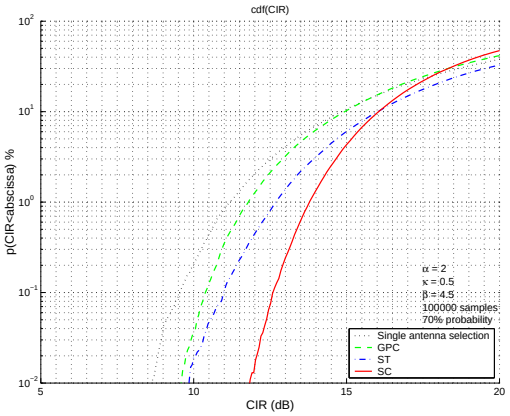
Figure 4.14: CIR CDFs for the four algorithms: path loss exponent $\beta = 4.0$, various mobile probabilities.



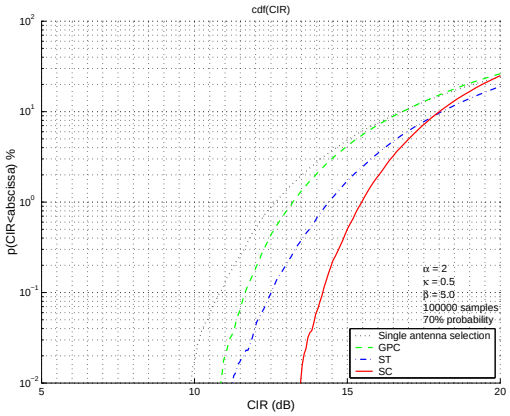
(a) $\beta = 3.5$



(b) $\beta = 4.0$

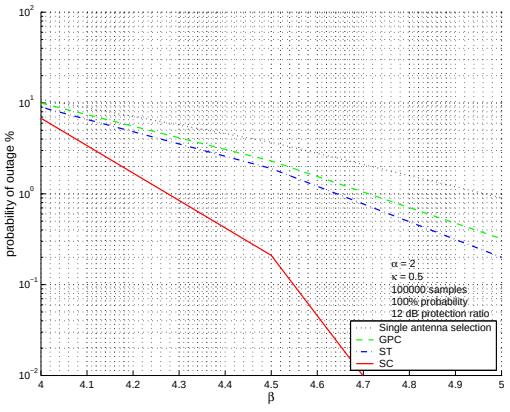


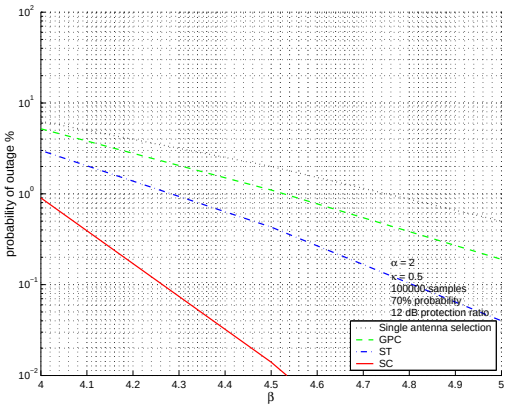
(c) $\beta = 4.5$



(d) $\beta = 5.0$

Figure 4.15: CIR CDFs for the four algorithms: 70% mobile probability, various path loss exponents (β).





(a) Mobile channels occupied with 100% probability (b) Mobile channels occupied with 70% probability

Figure 4.16: Comparison of the four algorithms: probability of outage vs β for a CIR of 12 dB.

$\beta = 3.5$				
prob	SAS	GPC	ST	SC
10%	23.1	23.0	35.1	34.6
70%	14.7	14.6	15.6	14.8
90%	13.6	13.5	14.0	13.1
100%	13.2	13.1	13.3	12.5

$\beta = 4.0$				
prob	SAS	GPC	ST	SC
10%	25.3	25.2	38.0	37.0
70%	16.8	16.8	17.9	17.0
90%	15.8	15.7	16.2	15.3
100%	15.3	15.2	15.6	14.7

$\beta = 4.5$				
prob	SAS	GPC	ST	SC
10%	27.3	27.3	40.3	39.3
70%	18.9	18.8	20.1	19.2
90%	17.9	17.8	18.4	17.5
100%	17.4	17.3	17.7	16.9

$\beta = 5.0$				
prob	SAS	GPC	ST	SC
10%	29.4	29.3	42.5	41.6
70%	20.9	20.9	22.2	21.3
90%	19.9	19.8	20.5	19.6
100%	19.4	19.3	19.8	19.0

Table 4.8: Average CIR (dB) for the algorithms.

reduced. However, the improvement shows in the shape of the CDF of the CIR, which was drastically steepened below the average CIR. So for a given CIR requirement, the probability of outage is greatly reduced using the proposed algorithm.

4.5 Delay in 802.16-2004 Mesh Mode

There are a number of mechanisms that affect the end-to-end delay in wireless multihop data transmission. The dominant mechanism emerges depending on frame length, packet error rates, number of nodes in the network, scheduling algorithm and the specific settings of protocol parameters. In this section we gather some results from literature and use them to find some delay characteristics of relaying.

When this work was done, 802.16-2004 was the latest published version. Whereas we have been using mobile station (MS) to denote the end user station, 802.16-2004 only defined a subscriber station (SS) so we will stick to that notation in this section. This section provides an overview of sources of delay in a multihop relay network based on 802.16-2004 Mesh, and some of the key parameters that affect delay.

4.5.1 Physical Layer Delay

Physical layer (PHY) delay is the time required to receive and retransmit packets from the BS through intermediate RSs to the destination SS. This can be calculated quite simply by multiplying the frame length by the number of hops. 802.16 frame sizes can be 2.5 ms, 4 ms, 5 ms, 8 ms, 10 ms, 12.5 ms or 20 ms.

Ordinarily in time division duplexing (TDD), the frames are divided into downlink (DL) and uplink (UL) subframes, with guard time gaps in between to allow for reversal of transmission direction. These gaps are called Receive/Transmit Transition Gap (RTG) and Transmit/Receive Transition Gap (TTG), and are specified as no less than 5 μs. Since

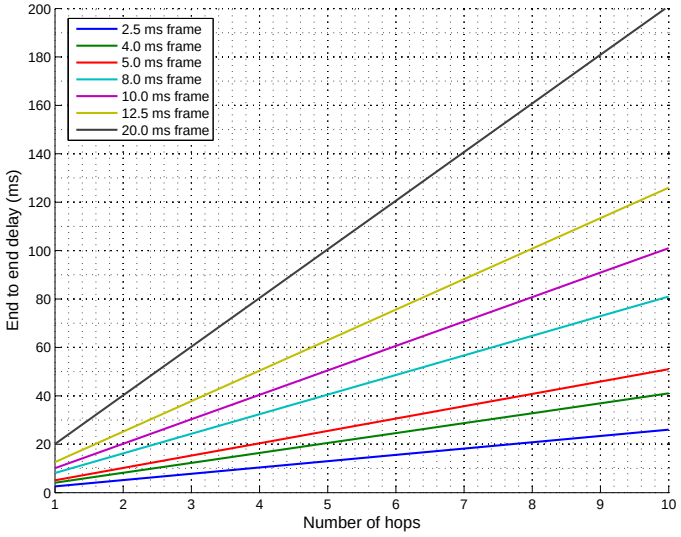


Figure 4.17: End-to-end, unidirectional physical delay for multihop 802.16-2004.

Mesh mode makes no distinction between UL and DL (there is no distinction between BS and SSs regarding scheduled access), there are turnaround times, subscriber station Receive/Transmit Transition Gap (SSRTG) and subscriber station Transmit/Receive Transition Gap (SSTTG), which according to the 802.16-2004 standard should be less than 100 μs.

Using this information we have plotted the end-to-end, unidirectional physical layer delay for one to ten hops, shown in Fig. 4.17. Naturally, this delay increases with frame size, and with the number of hops. A small frame size with a larger number of hops can be used. However, the use of a smaller frame size results in higher overhead for control data. Thus there is a tradeoff between delay and useable data throughput.

4.5.2 ARQ Delay

Automatic repeat request (ARQ) is an optional feature in 802.16 that corrects errored packets. Although ARQ improves the error performance and throughput of a wireless network, extra delay is incurred due to retransmission of errored packets.

4.5.3 Handshaking Delay

In 802.16, a three-way handshaking scheme (Request-Grant-Confirm) is used between the requester and grantor of a connection. Handshaking delay is often considered part of the MAC scheduling delay. Again, this three-way handshake occurs on every hop between the source and destination of the data.

4.5.4 MAC Scheduling Delay

MAC layer delay is the time between a station's request to transmit data, and the granting of the request by the scheduling mechanism[3]. It is often the most dominant source of delay, but unfortunately the most complicated to assess. This delay depends greatly on the type of scheduling used and a large number of system parameters, such as traffic load, network size, number of contending nodes, frame size, frame subdivision scheme (e.g. uplink/downlink ratio, percentage used for polling), etc. In Mesh mode, scheduled access delay occurs on every hop between the source and destination of the data.

There are four different QoS categories defined in 802.16-2004: unsolicited grant service (UGS), real-time polling service (rtPS), non-real-time polling service (nrtPS), and best effort (BE). Each of these uses a different scheduling technique, and therefore each has a different delay characteristic. Delay is generally longer on the uplink than the downlink, since multiple SSs must request and receive grants for bandwidth from the network. UGS is intended for delay sensitive applications such as voice over internet protocol (VoIP), while at the other end of the scale, BE is intended for services that can tolerate higher delay. UGS and rtPS services have maximum delay and jitter tolerance settings, while nrtPS and BE have no mechanisms for delay and jitter control. In the following sections, these QoS categories are analyzed separately. Unfortunately, only BE is defined for 802.16-2004 Mesh mode, which has been shown to have poor delay characteristics. Although not currently available for Mesh mode, we do provide some analysis of UGS and polling type techniques, as this may be relevant in future 802.16 work.

Scheduling delay depends on the specific scheduling algorithm used. In Mesh mode, three modes of scheduling are defined: centralized scheduling, coordinated distributed scheduling, and uncoordinated distributed scheduling. Some aspects of scheduling are defined in the standard, but key parts are left open for equipment manufacturers to choose.

4.5.5 Analysis of Published Work

As mentioned earlier, there are several sources of delay, and delay characteristics differ depending on the QoS type. Until recently, there had been little published work on delay specific to 802.16 networks. In mid- to late-2006, a larger number of publications specifically on numerous 802.16 topics (e.g. MAC layer design, scheduling, performance) appeared at conferences and in technical journals. Most do not cover Mesh mode, and although the scheduling mechanisms for Mesh are necessarily different, many of the results can be carefully extended to provide some useful information.

This section summarizes the most relevant published work available, with more emphasis

[3]In 802.16-2004 Mesh mode, the scheduling mechanism can be centralized at the BS, or distributed among the SSs.

on the BE work (since it is relevant to 802.16-2004 Mesh mode) and mixed service flows (since that represents more realistic networks).

Theory of Delay in Relay Networks

In [105], some theoretical tradeoffs are derived for general wireless ad hoc networks. This work gives a general relationship between delay (ignoring queuing delay at the source) and throughput in a network using an approach similar to Gupta and Kumar's approach to calculating throughput [25, 26]. The basic mechanism of the tradeoff is as follows. In a network with n nodes, the delay can be lowered by increasing the transmission range of each node so that data may reach the destination in fewer hops. However, this creates inter-node interference, and so the throughput decreases. Conversely, the throughput can be increased by reducing the transmission range of the nodes. In that case, data must be relayed through more nodes to reach its destination, increasing the delay. With the maximum throughput given in [25, 26], [105] shows that the delay, $D(n)$, scales with n according to

$$D(n) \propto \sqrt{n/\log(n)} \tag{4.62}$$

Although this work provides some useful insight, actual delay in a practical system, which is very much governed by the specific MAC implementation, cannot be inferred. More detailed analysis of relaying systems is presented in the following sections.

ARQ Delay

In [106], delay for a selective repeat ARQ scheme is analyzed for one and two hops. Average delays and complementary cumulative distribution function (CCDF) plots for networks with a PER of 0.2 are presented in their work. Traffic loads of 192 kb/s, 768 kb/s, 1536 kb/s, and 3072 kb/s in a 54 Mb/s system employing link adaptation were analyzed. The average delay was found to range from 2 to 4 ms, depending on the network traffic. 99.9% of the time, delay is less than 12 ms. Naturally, the delay increases with increasing traffic load. For two hops, the average delay is found to range from 4 to 9 ms. Not surprisingly, the average delay is approximately doubled and 99.9% of the time the delay is less than 20 ms. These results show that ARQ is generally not a major source of delay in 802.16 networks.

MAC Scheduling Delay: Best Effort

BE service has no constraints on delay or bandwidth and is the only QoS category defined for Mesh mode. SSs with BE connections contend for access to the network on each transmission and the network grants access the best it can. In 802.16, contention occurs in a control subframe (rather than in the data frame as in 802.11). 802.16 defines an algorithm that determines when a SS can make a transmission request in the control subframe. However,

802.16 does not define specifically how data transmissions are scheduled. A number of scheduling schemes are proposed in the literature. Naturally, delay can become problematic with BE, especially when the number of MSs is large.

As discussed earlier, three types of scheduling approaches are defined for 802.16-2004 Mesh mode: centralized scheduling, coordinated distributed scheduling, and uncoordinated distributed scheduling. With centralized scheduling, the BS collects SS bandwidth requests, determines and broadcasts the schedule to SSs via centralized scheduling configuration messages (MSH-CSCF). Specific scheduling algorithms are not defined in the 802.16 standard and numerous researchers are proposing and analyzing various algorithms.

With coordinated distributed scheduling, scheduling is coordinated within a two-hop extended neighbourhood of SSs. There are a number of transmission slots (the number of slots is set by a network configuration parameter - MSH-DSCH-NUM) in which SSs may transmit, and each SS determines its next transmission opportunity using an election algorithm defined in the standard. This algorithm results in a collision-free schedule. Access delay is determined by several system parameters (described a little later in this section) and the number of SSs within the two-hop neighbourhood.

With uncoordinated distributed scheduling, transmission is scheduled between two SSs. Collisions are possible with this form of scheduling, but it is suitable for simple ad-hoc network setup.

Not surprisingly, there exist tradeoffs between delay, efficient use of system resources, and fairness to all SSs in the network. In general, centralized scheduling results in better use of network radio resources, and thus better spectral efficiency since BS has the ability to determine the best use of resources on a system-wide basis. However, since resource requests must all make their way upstream to the BS, and the BS must disseminate the schedule back downstream, the delay between a SS transmitting a bandwidth request and receiving a bandwidth grant can be long. Distributed scheduling can reduce this delay since scheduling is organized within a two-hop neighbourhood. Naturally, distributed scheduling does not necessarily make best use of system-wide resources.

[7] appears to be the first analysis of 802.16 MAC distributed election algorithm for scheduling. Their delay results include both MAC scheduling delay and three-way handshaking delay, and are expressed as the average number of time slots. Conversion to time (in ms) depends on a number of network parameters. The resulting delay depends greatly on an adjustable parameter, XmtHoldoffTime, which controls the length of time a station must wait before sending a message requesting resources. With a small network, XmtHoldoffTime can be small since there is less contention for access, while for a larger number of stations, XmtHoldoffTime may be made larger to reduce delays due to request collisions.

The main results for three-way handshake delay are shown in Table 4.9, estimating the

time delay using a frame length of 10 ms, OFDM with a FFT size $N_{FFT} = 1024$, and a bandwidth of 10 MHz.

Table 4.9: Delay results derived from [7]

Number of stations (n)	10	30	50	100
Average delay (slots)	20	48	80	150
Average delay (ms) (100% control overhead)	13	31	52	97
Average delay (ms) (33% control overhead)	40	92	160	300
Average delay (ms) (20% control overhead)	61	160	261	500

[107] builds on the work in [7], proposing extensions to the 802.16 scheduling framework and analyzing the performance of the coordinated distributed scheduling algorithm. They present average scheduling delay results for a regular 8x8 grid of stations, with a base station in the centre of the grid (65 stations), and a frame length of 10 ms. The maximum number of hops in the network is four. The central base station experiences the worst delay, since it is surrounded by a large number of stations competing for resources. The best case average delay for the base station is found to be about 85 ms, while the average delay over the network is about 57 ms. The authors propose to lower the minimum XmtHoldoffTime given in 802.16 in order to reduce the delay. With a reduced XmtHoldoffTime, they show that the average delay for the base station reduces to about 63 ms. This is still high as this delay will occur on every hop. Adding another 10 ms for the physical transmission of a frame, the total average delay per hop is about 73 ms. For two hops, the average delay would be about 150 ms in one direction, and 300 ms over a round trip. This is sufficient for non-delay sensitive applications such as file transfer, but VoIP could not tolerate a delay this large.

MAC Scheduling Delay: Polling

There are two types of polling services defined in 802.16-2004: real time polling service (rtPS) and non-real time polling service (nrtPS). Both of these services use a polling mechanism in which SSs are given regularly recurring opportunities to request access to the network. Hence, request collisions are avoided, and scheduling delay can be better controlled. rtPS has delay and bandwidth constraints, while nrtPS does not. As a result, rtPS is expected to have slightly poorer performance than UGS, and nrtPS should perform somewhat better than BE. Although there are differences in the performance of these two schemes, we have not found published work that analyzes them explicitly.

In [12], a theoretical analysis of the average uplink delay for a polling scheme in 802.16 has been carried out. The authors give an expression for the average delay between an

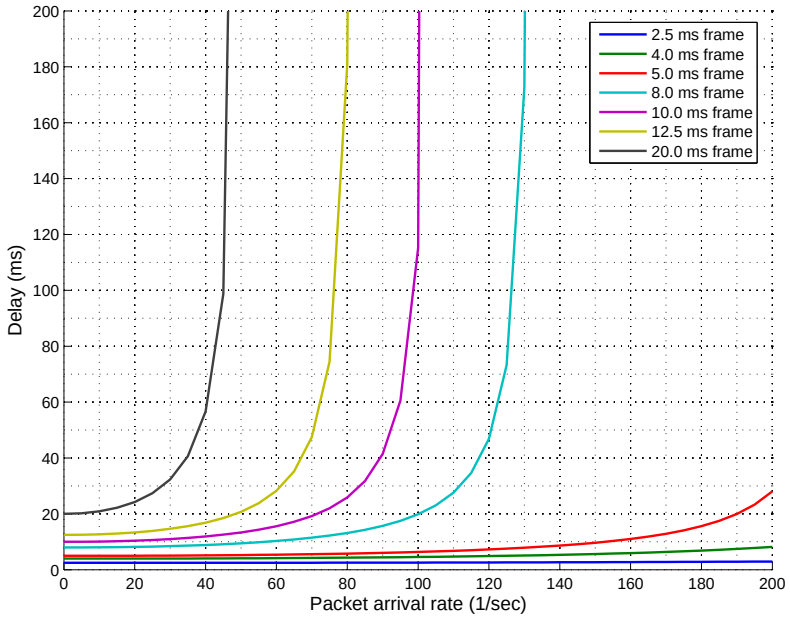


Figure 4.18: Average uplink delay in a single hop polling 802.16-2004 network with 10 SSs - adapted from [12].

SS's bandwidth request and the granted transmit slot. Their results have been verified by simulation. A simple round robin scheduling technique is assumed.

We have used their average delay expressions to analyze several different network configuration. Delay dependency on packet arrival rate, λ, is shown in Fig. 4.18 for a network with 10 SSs and the frame lengths specified in the standard. The lowest delay possible is about one frame length, since that is the minimum time it takes for the SS to receive a bandwidth grant from the BS after sending a bandwidth request. The delay increases slightly with packet arrival rate, with a sharp increase in delay as λ approaches $1/T_S$. At that point, packets are arriving faster than they can be cleared from the queues, so a backlog develops and the delay will approach infinity.

Fig. 4.19 shows the dependency of delay on network size for a frame length of 5 ms. As expected, additional SSs in the network can increase the delay. However, the delay remains tolerable even for a large number of SSs, since in its analysis it is implicitly assumed that enough polling and grant slots exist to handle all SSs. This is a key design issue for 802.16 networks.

MAC Scheduling Delay and Jitter: Unsolicited Grant Service

UGS provides the best delay characteristics and thus is most suitable for delay sensitive applications such as VoIP. Once a service flow is established via a connection request to the BS, packets can be transmitted in regularly occurring time slots without repeated requests. As a result, scheduling of UGS data flows is not an issue, but rather the call admission

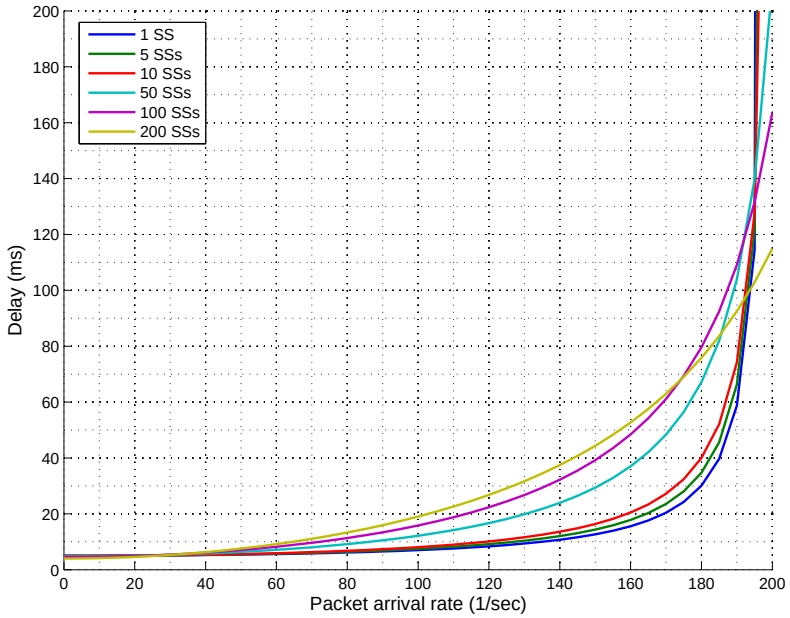


Figure 4.19: Average uplink delay in a single hop polling 802.16-2004 network with 5 ms frame - adapted from [12].

control upon initial connection.

[108] looks at delay and jitter for UGS. Although delay is low and dominated by ARQ and PHY layer delays, jitter can remain an issue, since in general the timing of allocated slots is not synchronous with the periodicity of the user data stream. In order to keep delay low for VoIP, 802.16 employs a Grant Synchronization mechanism. If the delay becomes too long, then the mechanism will shift the data grant ahead by one to three frames. In 802.16, several different frame sizes are allowed. Large frame sizes are desirable in order to keep the overhead lower, but the disadvantage is an increase in jitter. For example, if the frame size is 12.5 ms, then the timing of transmission may vary by about 12.5 ms. Jitter can be controlled by using smaller frame sizes. The authors propose a modification to the Grant Synchronization mechanism that reduces jitter. They show that for a frame size of 12.5 ms, the jitter can be limited to 5 ms.

MAC Scheduling Delay: Mixed QoS Flows

Some researchers have implemented and studied all four QoS categories in simulated wireless networks. These results are very useful since they indicate the relative delay characteristics of these categories. In general these results show that BE traffic performs poorly compared to the other types, and also how delay characteristics vary greatly with network size and design parameters.

Some simulated performance results of the four QoS categories were presented in [109]. In this study the BS grants uplink bandwidth to a SS, which in turn schedules amongst the

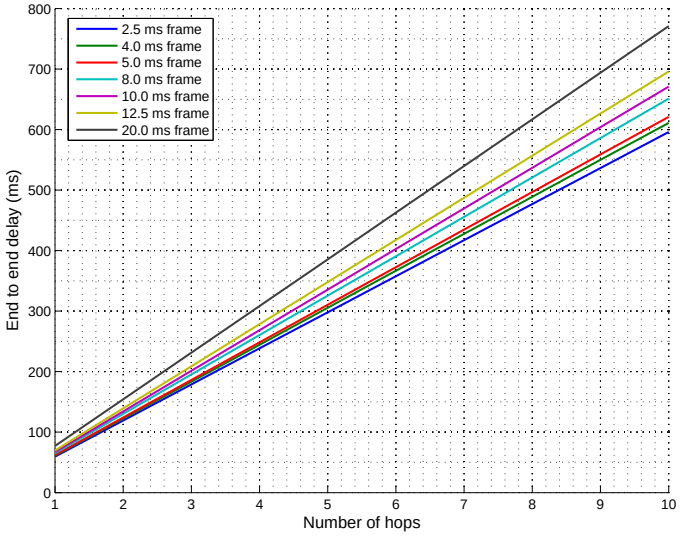


Figure 4.20: End-to-end, unidirectional physical and MAC delay for multihop 802.16-2004 using BE.

various service flows active at that SS. Each service flow (whether UGS, rtPS, nrtPS or BE) receives a weight which will determine the priority of bandwidth allocation.

In simulations, VOIP, motion picture experts group (MPEG) video, file transfer protocol (FTP), hypertext transfer protocol (HTTP) and telnet uplink flows are combined at several SSs connected to a BS. A 10 ms frame size is used. The delay is suitably low for up to 10 SSs, but it rises sharply for BE and nrtPS service flows in a large network. The delay can be as high as 4 seconds for 15 nodes. UGS and rtPS are less affected since they enjoy higher priority due to their delay limit.

Total Delay in a Multihop Network

Most of the research discussed above presents delay for a single hop network. These results can be extended to a multihop network since the MAC layer scheduling occurs at each hop. As discussed earlier, the MAC layer delay can vary widely depending on network parameters and network size. However, we will use a typical delay result given by [107]. For a four hop network, the average delay over the network was found to be 57 ms. The end-to-end unidirectional delay, including the results for physical layer delay, is plotted in Fig. 4.20.

It is apparent that the end-to-end unidirectional delay becomes significant for even a few hops when using BE in Mesh mode. Therefore Mesh cannot be used for delay sensitive services.

4.5.6 Extension to 802.16j

After completing this research, the IEEE 802.16 Working Group decided that Mesh mode performed poorly, and it was removed in 802.16j standard [21] issued in June, 2009. A major improvement in 802.16j is the implementation of all of the QoS categories so that delay sensitive services can be carried. So in 802.16j, relaying via multiple hops does not impact delay as severely as 802.16-2004 Mesh mode did. The above results can generally be extended to 802.16j but there are some differences. We have not studied delay in 802.16j since its publication was only recent, and our work was nearing completion at that time. There are two relay mechanisms defined: transparent relaying and non-transparent relaying. With transparent relaying, the control information is received by the MS directly from the BS at a low rate, but the payload data is delivered via intermediate RSs at a higher rate. This mode can increase the data rate to a MS, but the coverage range is not extended. So multiple relays multiply the PHY layer delay (including delay due to ARQ) but there is no impact on the MAC layer delay. With non-transparent relaying, a RS serving a MS acts as the BS for that MS, and so control information is received and regenerated on each hop. This method increases the data rate and extends the coverage range. However, the MAC layer delay is also affected by the number of intervening relays.

A few researchers have begun to study delay in 802.16j, and have proposed mechanisms for improving the delay. For example [110] showed that for a 10 hop 802.16j system using ARQ on each hop, the delay can be as high as 110 ms. They proposed a pipelined ARQ system that can limit the delay to about 50 ms.

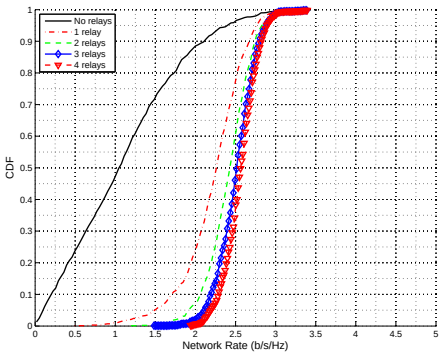
4.6 Cooperative vs Multihop Relaying

We discussed the relative merits of cooperative relaying vs multihop relaying for cellular applications in Section 2.2.2 and in Section 4.2 we have shown the benefits of multihop relaying. Here is a brief recap. Cooperative relaying provides diversity gain which improves the reliability of data transmission in a fading channel by reducing the BER. Multihop relaying can improve the aggregate network rate or throughput by reducing path loss, avoiding obstacles and allowing spatial reuse of spectrum throughout the cell. In this Section we compare CR and MH relaying on the basis of equal cost of deployment. When designing a cellular system with relaying, a system designer may wonder whether to deploy $n-1$ relays in parallel as in CR, or in series as in MH relaying. We assume the costs of the two types of relays are the same, and we compare the performance of the two options in delivering service from a BS to a MS a fixed distance away.

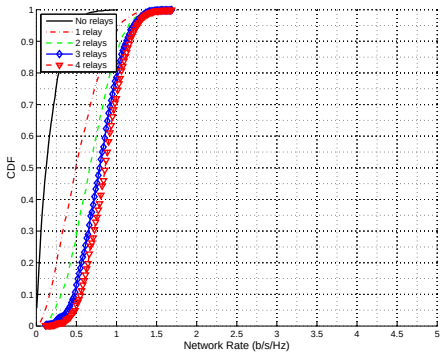
The system models used are based on the geometries shown in Figs. 2.2 and 2.3 and the dual slope path loss model described earlier in this Chapter. We use the Monte Carlo

technique to find the CDFs for the network rates, $R_{net,cr}$ and $R_{net,mh}$, for cell radii: 300, 500, 700, 1000, 2000 and 3000 m, and $n = 1, 2, \ldots 10$. Fig. 4.21 shows the Shannon capacity (2.10) performance comparisons between CR and MH relaying with spatial reuse (SR) for 300 m and 700 m cells. Comparing the 300 m cells in Fig. 4.21(a) and Fig. 4.21(c), it is apparent that CR with one relay (two diversity paths) provides a clear performance improvement over single hop and MH relaying with one relay. The addition of the relay provides a powerful diversity effect to overcome random channel fading. However using a second relay as a third diversity path provides only marginal improvement. With two relays (three hops), MH relaying outperforms CR, since although it does not provide diversity gain, it drastically increases the hop rates due to reduced path loss. Comparing 700 m cells, Fig. 4.21(b) and Fig. 4.21(d), we see a similar effect. With one relay, CR and MH relaying are close in performance. With further relay additions, MH relaying clearly outperforms, since the addition of relays creates short LOS paths instead of a NLOS path. Also, MH relaying makes more efficient use of spectrum with SR. CR also creates LOS paths, but the gain due to diversity is reduced because of reduced scattering, and the opportunity to use SR is limited.

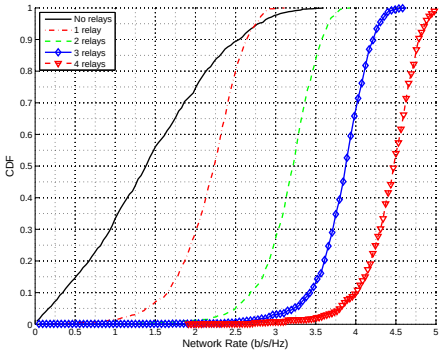
Fig. 4.22 again shows MH relaying outperforming CR. With large cells, the path loss can be very large, and so reducing the loss using MH relaying is highly beneficial, whereas CR, with only two hops, operates in low SINR and provides little diversity gain.



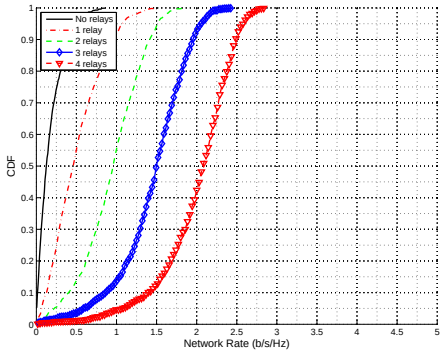
(a) Cooperative relaying for a cell radius of 300 m.



(b) Cooperative relaying for a cell radius of 700 m.

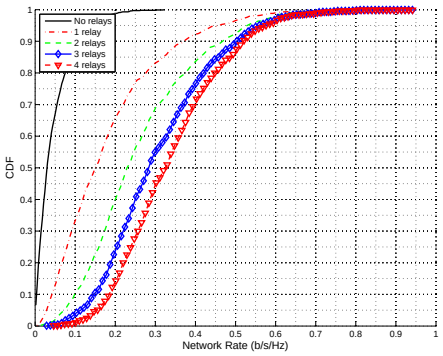


(c) Multihop relaying for a cell radius of 300 m.

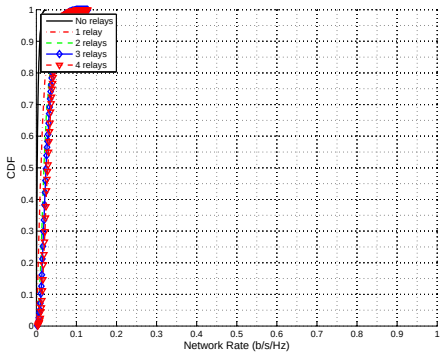


(d) Multihop relaying for a cell radius of 700 m.

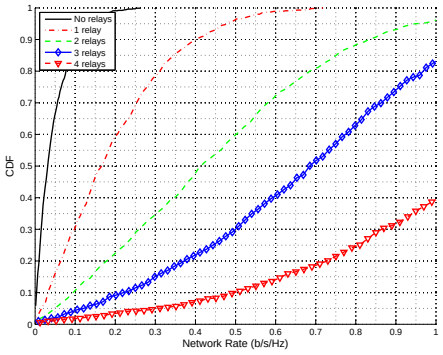
Figure 4.21: Cumulative distribution functions (CDFs) of network rates $R_{net,mh}$ and $R_{net,cr}$ (Shannon capacity formula).



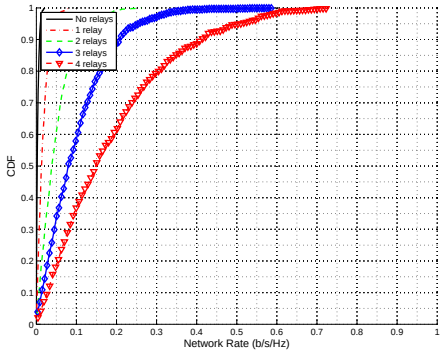
(a) Cooperative relaying for a cell radius of 1000 m.



(b) Cooperative relaying for a cell radius of 2000 m.



(c) Multihop relaying for a cell radius of 1000 m.



(d) Multihop relaying for a cell radius of 2000 m.

Figure 4.22: Cumulative distribution functions (CDFs) of network rates $R_{net,mh}$ and $R_{net,cr}$ (Shannon capacity formula).

Chapter 5

Multiple Antenna Relaying

With the model developed in Chapter 3 and some principles discovered for single antenna relaying in Chapter 4 , we extend the model and results to multiple antenna cellular systems.

5.1 MIMO Relaying Methods

Various forms of MIMO and relay processing are possible. As with single antenna relaying, we may consider amplify and forward (AF) in which the received signals are amplified and retransmitted, or decode and forward (DF) in which the data stream is recovered and re-encoded at each relay, or a number of hybrid approaches.

With AF, the transmitted vectors on the $(k+1)^{th}$ hop are

$$\vec{s}_{k+1} = \mathbf{P}_k^n \vec{r}_k \tag{5.1}$$

where $\mathbf{P}_k^n \in \mathbb{R}^{N_{T,k+1} \times N_{R,k}}$ is a matrix of gain values that create linear combinations of signals from the receive antennas to the transmit antennas. The received vector at the destination MS is

$$\begin{aligned} \vec{r}_n = &\mathbf{H}_n^n \mathbf{P}_{n-1}^n \mathbf{H}_{n-1}^n \mathbf{P}_{n-2}^n \mathbf{H}_{n-2}^n \quad \dots \quad \mathbf{P}_1^n \mathbf{H}_1^n \vec{s}_1 \\ &+ \mathbf{H}_n^n \mathbf{P}_{n-1}^n \mathbf{H}_{n-1}^n \quad \dots \quad \mathbf{H}_2^n \mathbf{P}_1^n \vec{n}_1 \\ &+ \mathbf{H}_n^n \mathbf{P}_{n-1}^n \mathbf{H}_{n-1}^n \quad \dots \quad \mathbf{H}_3^n \mathbf{P}_2^n \vec{n}_2 \\ &\dots \\ &+ \mathbf{H}_n^n \mathbf{P}_{n-1}^n \vec{n}_{n-1} + \vec{n}_n \end{aligned} \tag{5.2}$$

which can be rewritten as

$$\vec{r}_n = \mathbf{H}_{eq}^n \vec{s}_1 + \vec{n}_{eq} \tag{5.3}$$

This looks like a "single hop" $N_{T,1} \times N_{R,n}$ MIMO system with an equivalent channel matrix

$$\mathbf{H}_{eq}^{n} = \mathbf{H}_{n}^{n}\mathbf{P}_{n-1}^{n}\mathbf{H}_{n-1}^{n}\mathbf{P}_{n-2}^{n}\mathbf{H}_{n-2}^{n} \quad \dots \quad \mathbf{P}_{1}^{n}\mathbf{H}_{1}^{n} \tag{5.4}$$

and equivalent noise vector

$$\begin{aligned} \vec{n}_{eq} =& \mathbf{H}_{n}^{n}\mathbf{P}_{n-1}^{n}\mathbf{H}_{n-1}^{n} \quad \dots \quad \mathbf{H}_{2}^{n}\mathbf{P}_{1}^{n}\vec{n}_{1} \\ &+ \mathbf{H}_{n}^{n}\mathbf{P}_{n-1}^{n}\mathbf{H}_{n-1}^{n} \quad \dots \quad \mathbf{H}_{3}^{n}\mathbf{P}_{2}^{n}\vec{n}_{2} \\ &\dots \\ &+ \mathbf{H}_{n}^{n}\mathbf{P}_{n-1}^{n}\vec{n}_{n-1} + \vec{n}_{n} \end{aligned} \tag{5.5}$$

$\mathbf{H}_{eq}^{n} \in \mathbb{C}^{N_{R,n} \times N_{T,1}}$ and $\vec{n}_{eq} \in \mathbb{C}^{N_{R,n} \times 1}$. Standard MIMO techniques such as spatial multiplexing and diversity can be used end-to-end, and the capacity of this channel is given by

$$C = \log_2 \det \left(\mathbf{I}_{N_{R,n}} + \frac{\rho_{eq}}{N_{T,1}} \mathbf{H}_{eq}^{n} [\mathbf{H}_{eq}^{n}]^{H} \right) \quad \text{b/s/Hz} \tag{5.6}$$

where ρ_{eq} is the equivalent SNR. We can see from (5.5) that the noise accumulates just like the single antenna AF case.

Now we have the added requirement of determining the appropriate $\{\mathbf{P}_{k}^{n}\}$ in order to maximize the end-to-end capacity. To accomplish this all relays must have all of the channel state information and the search space has

$$S_{AF} = \sum_{k=1}^{n-1} (N_{T,k+1} N_{R,k}) \tag{5.7}$$

variables to be determined. Since $\mathbf{H}_{eq}^{n}$ is composed of a concatenation of MIMO channels, this channel acts like a series of keyhole channels which is degraded by the lowest rank MIMO channel in the system. It has been shown [58] that degraded MIMO channels suffer greatly from poor capacity when spatial multiplexing is used, and from reduced diversity order when diversity techniques are used.

We can see that there are a number of difficulties with the AF MIMO relay approach.

- There is a large search space for optimization or exhaustive search which grows with number of antennas and number of hops. This must be jointly optimized with the MIMO technique used.
- The noise is amplified and accumulates on each hop.
- All channel state information is required at all relays.
- The end-to-end link is a degraded MIMO channel if any of the hops are degraded, have fewer antennas or are in poor condition.

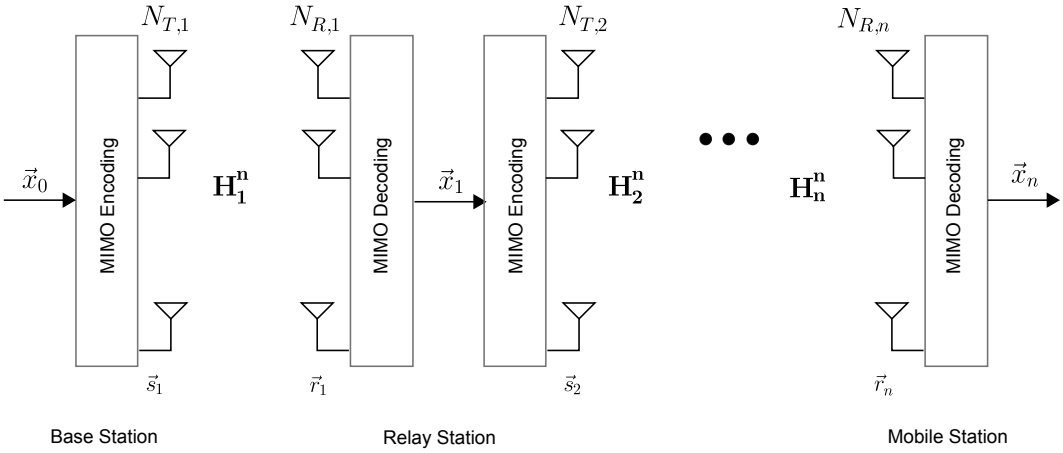


Figure 5.1: A one-dimensional DF multihop MIMO system model.

With DF, MIMO decoding and re-encoding occurs at each relay. The system model is shown in Fig. 5.1.

Data streams $\{\vec{x}_k\}$ are recovered at each hop, so instead of noise accumulation through the hop this system suffers from error propagation. However, channel state information is only required locally to each hop, so the system is much simpler. Another disadvantage is that the end-to-end rate is limited by the slowest hop. However, we will see that in a two-dimensional multihop system with spatial reuse (effectively a form of multiuser diversity), this has less impact on the aggregate network performance.

There appears to be no research done to date on the general case of AF MIMO relaying. However several authors have published work on the two-hop case. In [9, 10, 111] some results are given for a simplified protocol for two hops. In their work, $n = 2$, and they set $N_{T,2} = N_{R,1} = M$ and $N_{T,1} = N_{R,2} = N$. The gains at the relays are simply scalars so that $\vec{s}_2 = \sqrt{p}\,\vec{r}_1$. They investigate the capacity performance of DF, AF and various hybrid schemes.

A key result from [9, 10, 111] is that DF outperforms all of the other techniques for all SNRs.

We decided not to pursue AF MIMO techniques any further, and decided to consider only DF as the most practical alternative in cellular multihop MIMO system for the following reasons.

- The DF multihop MIMO system appears easier to implement than the general AF.
- The simplified version of AF and hybrids in [9, 10, 111] are less complex than DF, but their performance is consistently poorer.
- DF has a reduced requirement for dissemination of channel state information. Channel state information for a given hop is only needed at that hop. Channel quality

Table 5.1: Mixed MH MIMO case.

	N_T	N_R
BS	4	4
RS	3	3
MS	2	2

information may be needed elsewhere for scheduling purposes.

- Error accumulation in DF is only a serious problem for higher error rates. Error rates on each hop can be kept low by reducing the rate on a poor channel, or by exploiting transmission opportunities elsewhere in the network until a poor channel improves.

- DF allows the rate or diversity to be exploited fully on each link. Although a one dimensional system is limited by the lowest rate link, this is not a limiting factor in two dimensional systems due to spatial reuse.

As with the single antenna DF MH relaying systems, the rate or throughput can be found on each individual link using capacity formula (2.13) or the Salo bounds (2.17) and (2.19). The network throughput for a DF multihop MIMO network can be determined using the techniques discussed in Sections 4.2.2 and 4.2.3.

5.2 Results

We have simulated cases in which each hop uses $(N_{T,k} \times N_{R,k}) = (1 \times 1)$ (single antenna), (2×2), (3×3), (4×4), (5×5), and (6×6) MIMO. In practice, the BS can have a large antenna array, RSs must have a smaller array since they must be smaller and inexpensive, and MSs (laptop computers or mobile computing devices) are very limited in size. So we simulated a more realistic case (called the *Mixed* case in the figures), as described in Table 5.1. This creates hops with $(N_{T,k} \times N_{R,k}) = (4 \times 3)$, (3×3), and (3×2) on the downlink BS-RS, RS-RS, and RS-MS hops, respectively. The uplink will have $(N_{T,k} \times N_{R,k}) = (2\times3)$, (3×3), and (3×4) on the MS-RS, RS-RS, and RS-BS hops, respectively. This is a sensible design since transport hops nearer the BS can achieve higher rates, which is helpful in alleviating bottlenecks. Universal frequency reuse is used among the cells for all cases. We assume the use of omnidirectional (in the horizontal plane) antenna elements for the MIMO arrays since they provide the greatest spatial spread.

5.2.1 Ricean MIMO Hop

Here we look at the performance of a single Ricean MIMO hop. As discussed earlier, the addition of relays shortens the hop distances, which reduces path loss and scattering (i.e.

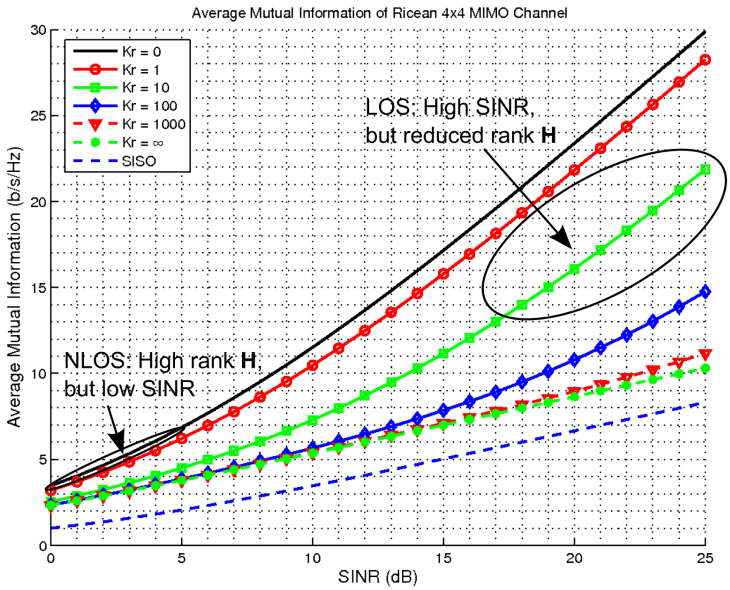


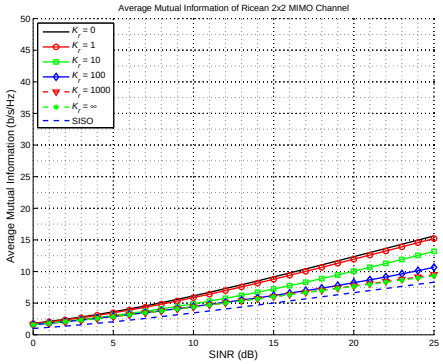
Figure 5.2: Upper bound on the average mutual information for (4×4) Ricean MIMO hop, with full rank $\mathbf{H}_{NLOS}$ and rank 1 $\mathbf{H}_{LOS}$, and a comparison to SISO.

increases the Ricean factor K_r). It is useful to look at this effect on a single hop link before studying the full network. Fig. 5.2 shows the average mutual information for a (4×4) MIMO link with full rank $\mathbf{H}_{NLOS}$ and rank 1 $\mathbf{H}_{LOS}$ calculated from (2.19) [60].

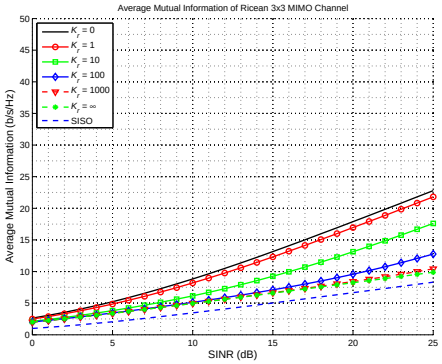
Wideband cellular systems generally operate at a low SINR. For example, a typical bit energy to noise power requirement in IS-95 (code division multiple access) is 6 dB [59]. OFDM systems may have slightly higher SINRs. It is easy to see from this figure that the rate advantage due to MIMO is relatively low at low SINR. We can increase the SINR on each hop by adding relays, but this may increase K_r, which reduces the rank of $\mathbf{H}$ and MIMO capacity gain is reduced, until at $K_r = \infty$, there remains only 6 dB array gain due to multiple receive antennas. From (3.4) we find that K_r is still about 10 at a fairly short distance of 100 m, and so MIMO gain, although reduced at this distance, is not completely lost.

Fig. 5.3 shows the dependence of capacity on the Rice factor and antenna configuration. More antennas do provide higher capacities, but the loss in capacity with increasing K_r is greater.

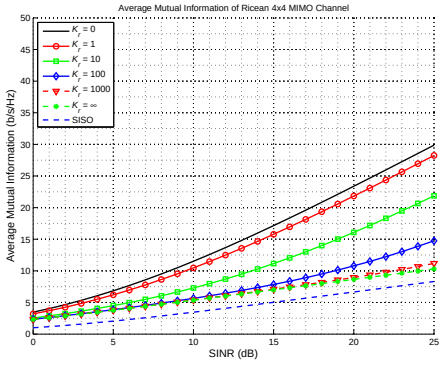
Fig. 5.4 shows the dependence of capacity on the Rice factor and SINR. The plots show that the capacity can drop off quite drastically with K_r at a fixed SINR, especially with a large number of antennas. Rice factor in cellular systems typically ranges from 3 to 20, which is in the range of steep reduction of capacity. The reason for this dropoff is that as K_r increases, more power is contained in the low rank LOS component of the matrix channel.



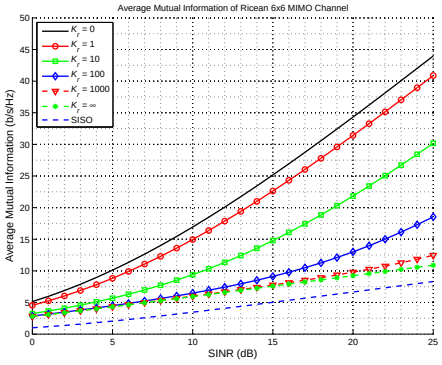
(a) $(N_R \times N_T) = (2 \times 2)$.



(b) $(N_R \times N_T) = (3 \times 3)$.

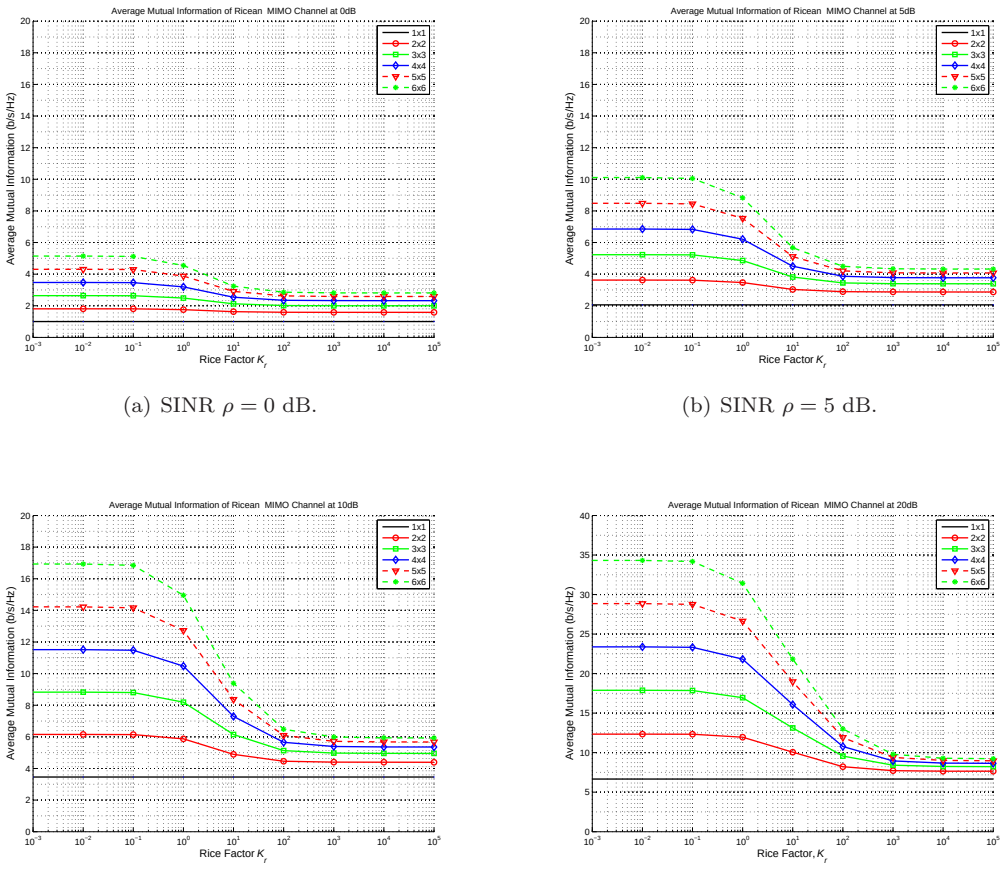


(c) $(N_R \times N_T) = (4 \times 4)$.



(d) $(N_R \times N_T) = (6 \times 6)$.

Figure 5.3: Upper bound on the average mutual information for a Ricean MIMO hop, with full rank $\mathbf{H}_{NLOS}$ and rank 1 $\mathbf{H}_{LOS}$.



(a) SINR $\rho = 0$ dB.

(b) SINR $\rho = 5$ dB.

(c) SINR $\rho = 10$ dB.

(d) SINR $\rho = 20$ dB.

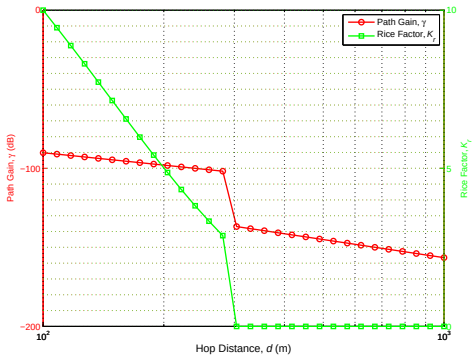
Figure 5.4: Upper bound on the average mutual information for a Ricean MIMO hop, with full rank $\mathbf{H}_{NLOS}$ and rank 1 $\mathbf{H}_{LOS}$.

This means that signal energy collapses into one dominant eigenmode of the channel matrix, $\mathbf{H}$.

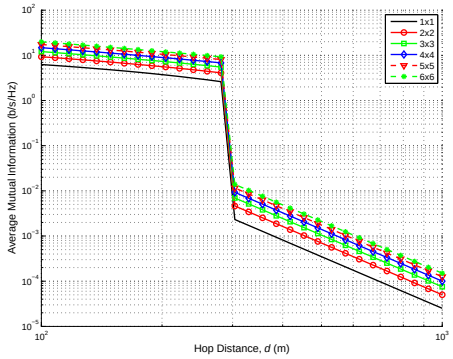
The previous results show the effects of K_r and SINR with one of them fixed while we vary the other. However, Rice factor and path loss change simultaneously with distance in a real propagation environment since a rich scattering environment (which is good for MIMO) becomes depleted with decreasing path loss. In the following figures we examine the effects of K_r and SINR jointly using the $K_r(x)$ and $\rho(x)$ models given by (3.2) and (3.4). Fig. 5.5(a) shows how K_r and path loss vary with distance, using a distance breakpoint of 300 m. Fig. 5.5(b) shows the resulting hop capacity. It is clear that the loss in MIMO gain is small compared to the gain due to increased SINR.

5.2.2 One Dimensional Multihop Relaying

In this section we look at how MIMO and MH relaying operate jointly in a one dimensional linear system with co-channel interference. Numerous cases have been simulated using the



(a) Dependence of path loss and Rice factor on hop distance.

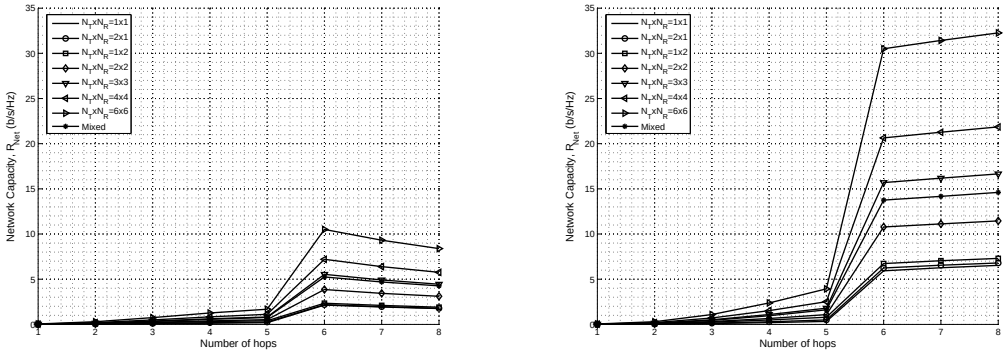


(b) Dependence of average mutual information for Ricean MIMO channels on hop distance, with full rank $\mathbf{H}_{NLOS}$ and rank 1 $\mathbf{H}_{LOS}$ and fixed transmit power, $P_{TX} = 0$ dBm.

Figure 5.5: Effect of hop distance, using dual slope model.

system model as described. We include here a sample of simulation results, for up to eight hops, and up to (6×6) MIMO. Figs. 5.6 and 5.7 show some sample results for a cell radius of 1500 metres, equally spaced relays and a distance breakpoint of 300 metres. For fewer than six hops, all hops are NLOS and so the path loss of each hop is high. All hop paths are uncorrelated Rayleigh channels, which should provide a good environment for capacity gain due to MIMO spatial multiplexing. However, the hops suffer from low SINR due to high path loss and co-channel interference. Since spatial multiplexing works best at high SINR, MIMO capacity gain is minimal. With the addition of another relay (a sixth hop), all hops become LOS and the path loss of each hop becomes drastically reduced. As a result, the hop SINRs increase and the network capacity increases greatly. Although SINR is much higher, spatial multiplexing and diversity gains suffer due to the largely correlated propagation environment. However, MIMO does assist in MH LOS situations because there remains some scattering component, and there exist receive array gain and interference control afforded by conventional transmit beamforming.

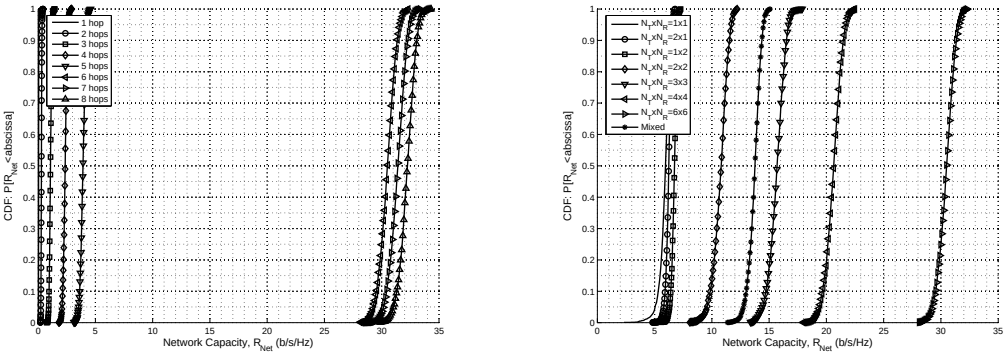
Fig. 5.6 clearly shows the importance of spatial reuse in MH relaying. When there are more than two hops, channels (time or frequency slots) can be reused at stations that are adequately separated in space, which provides great increases in network-wide spectral efficiency despite the introduction of interference between subcells. Without spatial reuse, interference is lower, but MH relaying is more wasteful of spectrum. As shown in Fig. 5.6(a), no spatial reuse case, R_{Net} decreases beyond 6 hops since relaying is increasingly wasteful of resources. With fewer than 6 hops, the addition of relays is slightly beneficial since the increase in SINR afforded by shortening the hop distances increases the MIMO gain. In Fig. 5.6(b), with spatial reuse, R_{Net} continuously increases with the number of hops. With more relays, there is more opportunity for channel reuse in distant parts of the cell.



(a) Multihop MIMO - no spatial reuse.

(b) Multihop MIMO with spatial reuse.

Figure 5.6: Multihop MIMO network capacities - with rank one LOS channel matrices.



(a) Multihop, (6×6) MIMO with spatial reuse.

(b) Six hop MIMO with spatial reuse.

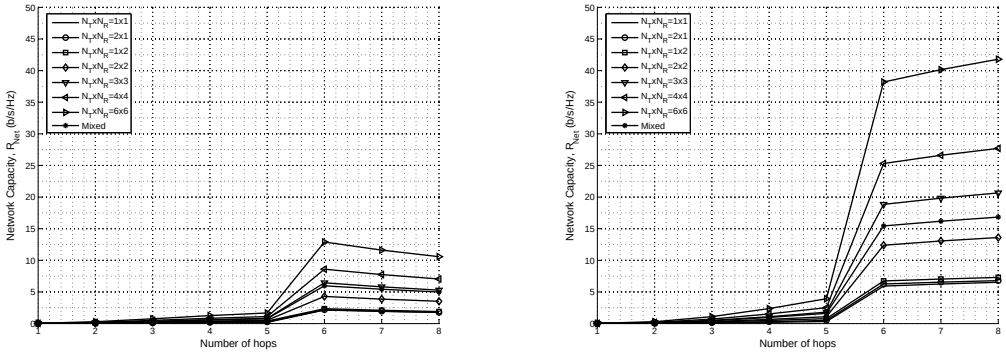
Figure 5.7: Cumulative distribution functions of MH MIMO network capacity - with rank one LOS channel matrices.

Cumulative distribution functions of MH MIMO network capacity for some cases are shown in Fig. 5.7. The figure demonstrates the drastic capacity increase that MH relaying can achieve by avoiding NLOS propagation and enabling spatial reuse, and the gradual increase in capacity afforded by MIMO.

Figs. 5.6 and 5.7 show the results using a rank one $\mathbf{H}^n_{LOS,k}$, while Fig. 5.8 and 5.9 show the results for full rank $\mathbf{H}^n_{LOS,k}$. The results are similar, but obviously R_{Net} is higher when the LOS matrix is high rank (although this is not likely to occur in a typical cellular system [60]).

5.2.3 Two Dimensional Multihop Cellular System

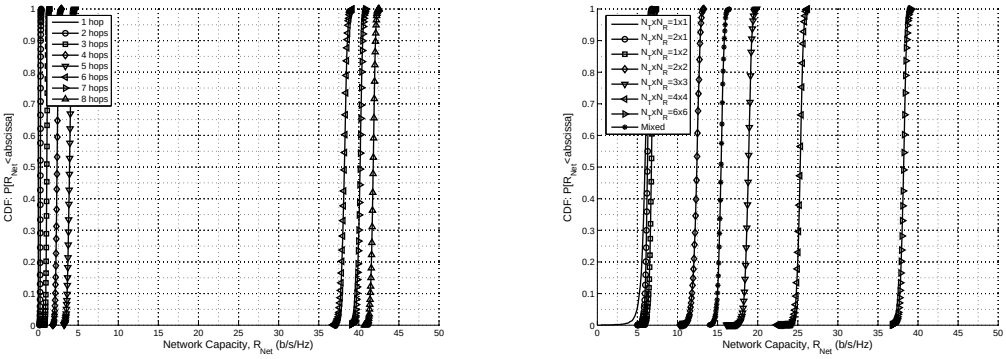
In this section we extend the calculations to a cellular system with tesselated Manhattan and hexagonal cells with one to four hops using the results of [53]. Universal frequency reuse is used among the cells for all cases. We assume the use of omnidirectional (in the



(a) Multihop MIMO - no spatial reuse.

(b) Multihop MIMO with spatial reuse.

Figure 5.8: Multihop MIMO network capacities - with full rank LOS channel matrices.



(a) Multihop, (6×6) MIMO with spatial reuse.

(b) Six hop MIMO with spatial reuse.

Figure 5.9: Multihop MIMO capacity CDFs - with full rank LOS channel matrices.

Table 5.2: Hop distances: 500 m radius hexagonal cell.

	Distance per hop (m) and path type (NLOS/LOS))			
n	r_1	r_2	r_3	r_4
1	500 - NLOS	-	-	-
2	333 - NLOS	167 - LOS	-	-
3	200 - LOS	200 - LOS	100 - LOS	-
4	143 - LOS	143 - LOS	143 - LOS	71 - LOS

Table 5.3: SINRs: 500 m radius hexagonal cell.

	SINR per hop (dB)			
n	Hop 1	Hop 2	Hop 3	Hop 4
1	-5.4	-	-	-
2	-12.9	19.4	-	-
3	28.9	19.8	17.4	-
4	32.5	27.9	18.4	16.1

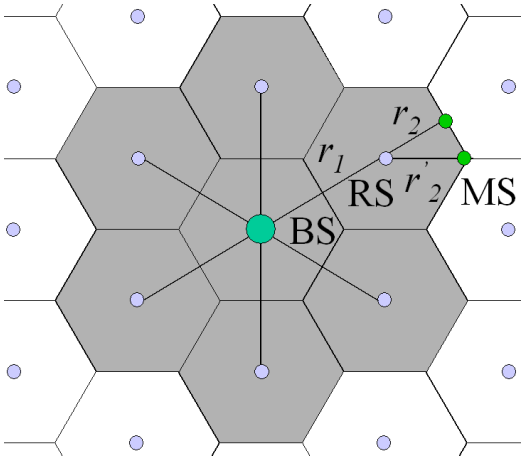
horizontal plane) antenna elements for the MIMO arrays since they provide the greatest spatial spread.

For a detailed example, we show calculations for a hexagonal topology with circumscribed cell radius of 500 m. The hop distances for this case are given in Table 5.2. SINRs calculated are given in Table 5.3.

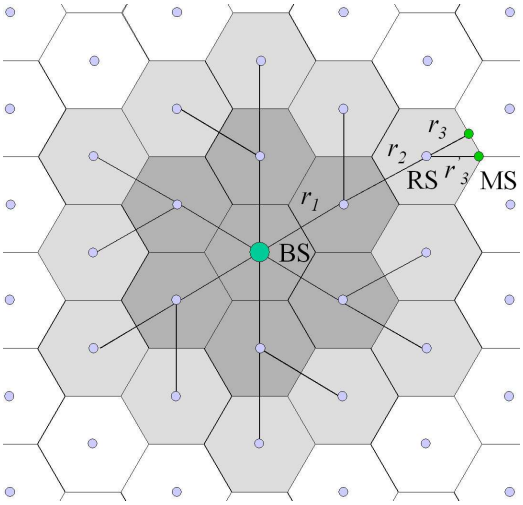
It is useful to observe how distances, path losses and SINRs change as relays are added to this system. The non-linear path loss model used, combined with the effect of scheduling transmissions among subcells within a cell gives some non-linear and somewhat surprising results.

With no relays ($n = 1$), an MS at the cell edge is 500 m from the BS, which gives a NLOS channel according to the path loss model (3.2). In this case, reception at the MS suffers from high co-channel interference from adjacent cells and a very poor SINR since we are considering universal frequency reuse among cells. The two hop ($n = 2$) hexagonal case shown in Fig. 5.10(a) has six RSs around the BS which gives two hops between the BS and any MS at the cell edge. The first hop, between the BS and any RS, is about 333 m and therefore is Rayleigh/NLOS according to the dual slope model. The second hop, between any RS and a cell-edge MS, is about 167 m and Ricean/LOS. The first hop link suffers from high path loss, and experiences high co-channel interference from numerous RSs in other cells. In fact there are three interfering RSs in other cells that are the same distance away as the BS. The interference is particularly bad from those RSs since the scheduling of RS transmissions in the other cells is not coordinated with the BS and RSs in the studied cell. Interference from within the studied cell is eliminated by scheduling. The second hop has a much better SINR since that link enjoys a much reduced path loss due to LOS, yet interfering signals are a greater distance away and experience higher loss due to NLOS.

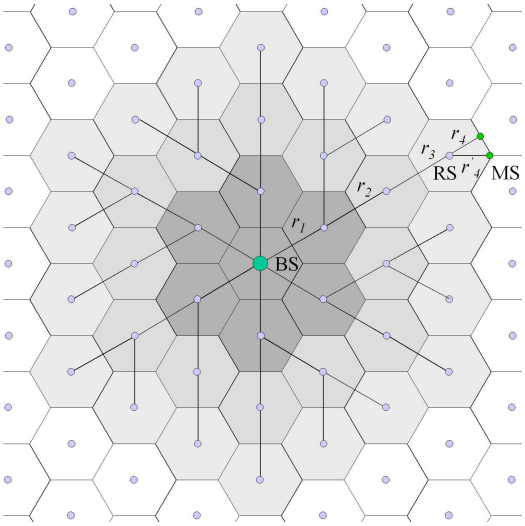
Adding 12 more RSs creates a three-hop hexagonal system as shown in Fig. 5.10(b). All three hops to an MS at the cell edge are LOS channels but the interfering channels



(a) Two-hop hexagonal topology.



(b) Three-hop hexagonal topology.



(c) Four-hop hexagonal topology.

Figure 5.10: Multihop hexagonal topologies.

Table 5.4: Rates: 500 m radius hexagonal cell, single antenna.

	R per hop (b/s/Hz)				R_{net} (b/s/Hz)
n	Hop 1	Hop 2	Hop 3	Hop 4	
1	0.30	-	-	-	0.30
2	0.058	6.2	-	-	0.067
3	9.3	6.3	5.5	-	5.2
4	10.5	8.9	5.8	5.1	7.7

Table 5.5: Rates: 500 m radius hexagonal cell, 3×3 MIMO on each hop.

	R per hop (b/s/Hz)				R_{net} (b/s/Hz)
n	Hop 1	Hop 2	Hop 3	Hop 4	
1	1.0	-	-	-	1.0
2	0.2	17.3	-	-	0.24
3	23.0	14.3	11.0	-	12.1
4	25.6	21.1	12.4	9.8	18.0

are still NLOS. Also, RSs within the studied cell can be scheduled to minimize co-channel interference. RSs in other cells, uncoordinated with transmissions in the study cell, are now a much greater distance away and so have much less impact than in the two-hop case. The resulting improvement in SINR on the links is dramatic.

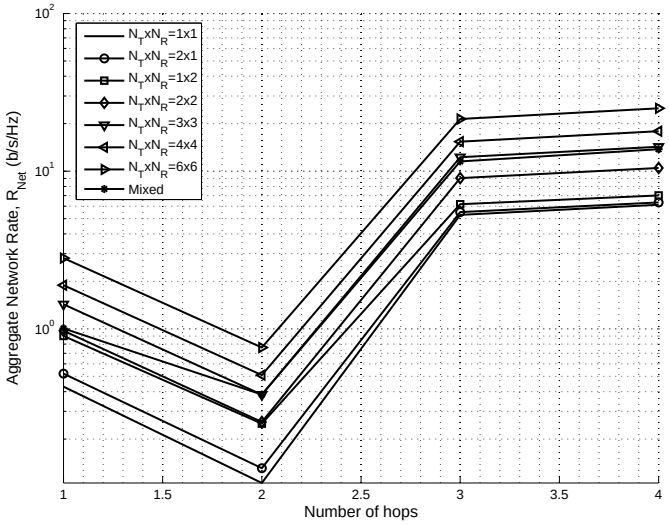
The next step, creating a four-hop hexagonal system as shown in Fig. 5.10(c), shortens the hops a little more. However, the incremental improvement over three-hop is less dramatic since LOS links were already obtained by the three-hop system. Notice that the SINR has improved on the first hop fairly significantly since the inner RSs become more insulated from the interfering transmissions from other cells. The last hop does not improve much in SINR because it is still quite near interfering subcells in the adjacent cells.

With the SINRs calculated above, we can now calculate the rates on each hop, and the aggregate network rate, R_{net}, with spatial reuse. Single antenna, 3×3 MIMO, and mixed MIMO cases are shown in Tables 5.4, 5.5, and 5.6 respectively.

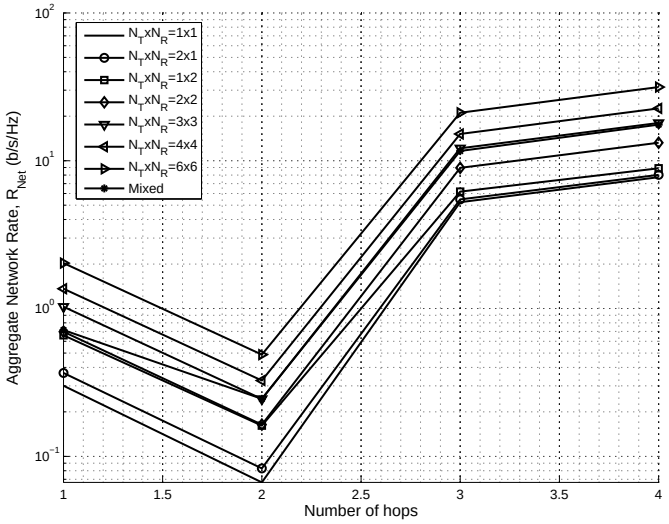
Fig. 5.11 compares the aggregate bit rates achievable by numerous MIMO configurations versus n, for Manhattan and hexagonal topologies with 500 m radius cells. Fig. 5.12 shows the results for 1000 m radius cells.

Table 5.6: Rates: 500 m radius hexagonal cell, mixed MIMO case.

	R per hop (b/s/Hz)				R_{net} (b/s/Hz)
n	Hop 1	Hop 2	Hop 3	Hop 4	
1	0.72	-	-	-	0.72
2	0.21	12.4	-	-	0.25
3	23.7	14.3	8.9	-	11.6
4	26.3	21.1	12.4	7.9	17.5

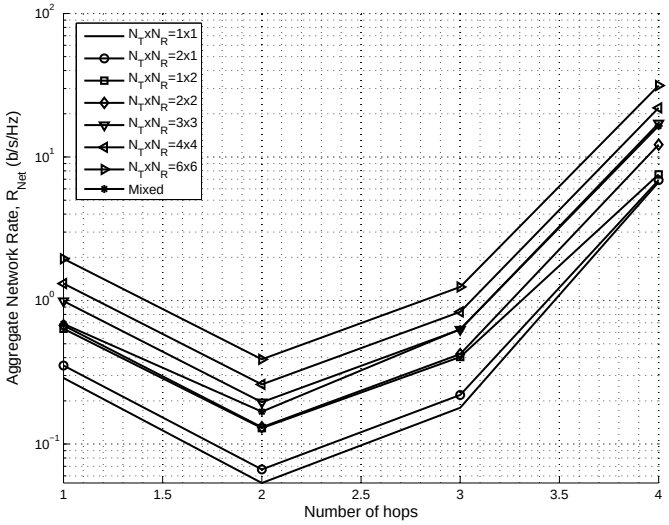


(a) Manhattan MH MIMO cell.

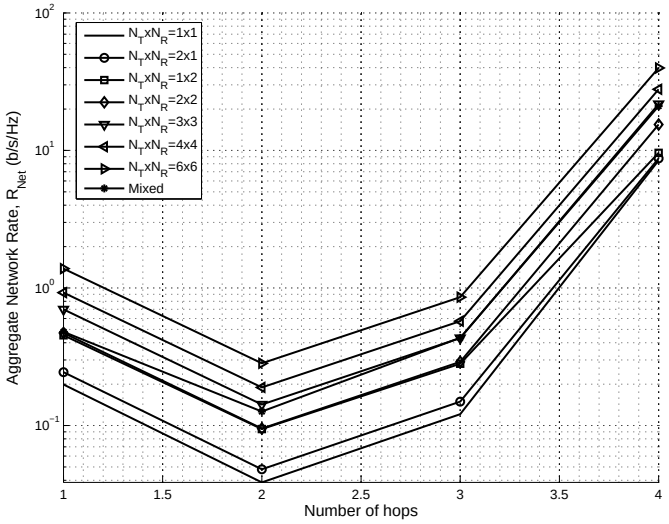


(b) Hexagonal MH MIMO cell.

Figure 5.11: Aggregate network rate for 500 m radius MH MIMO cells.



(a) Manhattan MH MIMO cell.



(b) Hexagonal MH MIMO cell.

Figure 5.12: Aggregate network rate for 1000 m radius MH MIMO cells.

5.3 Discussion

Results presented in this chapter show that there is a fundamental capacity tradeoff when using MIMO and MH relaying jointly. This may seem obvious, since the two techniques actually work using conflicting assumptions: MIMO works by exploiting the randomly scattering channel, while MH relaying attempts to mitigate that random behaviour and reduce path loss. A key effect is the loss of MIMO's diversity and spatial multiplexing gains as relaying is introduced. This is apparent from (2.13) since, with $r_{LOS} = 1$, one eigenmode of $\mathbf{H}$ dominates and MIMO capacity gain is lost as the Rice factor, K_r, increases. However, multiple antennas provide advantages due to receive array gain, and due to minimization of co-channel interference with conventional transmit beamforming methods. Also, the use of MH relaying shortens the hop distances, which increases the SINR. So although channel rank is reduced SINR increases. To put this another way, MIMO's spatial multiplexing and diversity gains are achieved at the expense of SINR: the uncorrelated signal that is key to MIMO gains occurs because the signal experiences rich scattering associated with high path loss.

At the outset one might expect that MH relaying should work best since it addresses the real root of the problem - a weak received signal - while MIMO tries to make the best of a bad situation by collecting and making best use of randomly scattered signals. Consider the ultimate MH system, in which there are an infinite number of relays spaced at zero distance. The signal received at the end destination at any distance from the sender would be perfect, but the cost of relay placement would be infinite, the delay long, and the algorithms and signalling overhead for routing prohibitively complicated.

As shown by example, the use of omnidirectional antenna elements in a MIMO array leaves the receivers in the system open to more severe co-channel interference. It would be preferable to use directional antenna elements since, as shown in Chapter 4, antenna gain and rejection of off-axis co-channel interference provides great advantage. However, using directional antenna elements reduces the scattering in the MIMO channel which reduces the MIMO gain. Classical beamforming with the multiple antennas can be used to make the whole array more directional, but this reduces potential MIMO gain. There is a tradeoff between MIMO capacity gain and power gain.

Hence a sensible application of MIMO with MH relaying in a cellular system may exploit the following approaches.

- Add just enough relays to achieve LOS and low path loss between stations. The resulting small subcells enable higher area averaged spectral efficiency (b/s/Hz/km^2).
- Use universal frequency reuse among the cells to increase area averaged spectral efficiency.

- Use spatial reuse scheduling among subcells throughout the cell in order to increase area averaged spectral efficiency.
- Classical beamforming with multiple antennas at the transmit side can be used to increase the gain of the array and reduce co-channel interference to off-axis receivers.
- Multiple antennas at the receiver will provide array gain and minimize co-channel interference from off-axis interferers.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

We have assembled a realistic model for single antenna and multiple antenna relaying in a cellular system. This model was used to determine the network capacity and investigate the tradeoffs associated with the combination of multihop relaying and MIMO. MIMO techniques can provide great gains in capacity, but only when rich scattering occurs, as is the case when the channel is NLOS. In addition, MIMO gains are best at high SNRs. On the other hand, multihop relaying provides great advantage by relaying around obstacles, reducing the path loss by creating LOS conditions, and enabling spatial reuse of spectrum. The path loss exponent is quite high in a NLOS channel ($\beta_{NLOS} = 3.8$ is typical), which means that reducing the signal transmission distance can reduce the path loss significantly. With the placement of sufficient number of relays, obstacles can be avoided, creating a series of LOS paths, reducing the path loss exponent to $\beta_{LOS} = 2.6$. With careful consideration of the local RF propagation environment, the system designer can add just enough relays to create individual LOS hop paths. Distances to interferers are larger and so the interfering signals experience higher loss. The result is a much improved SINR and a higher link data rate. However, multihop relaying also reduces the scattered component of the signal which degrades MIMO gains. We have shown that there is some tradeoff in using these methods simultaneously, but by understanding the nature of this tradeoff in a typical cellular system, we can leverage the benefits of both MH relaying and MIMO. A key aspect of this joint application is recognizing that high scattering (which is good for MIMO) necessarily means high signal loss (which is not good). So the great gains afforded by MIMO actually come at the expense of increased path loss.

6.2 Future Work

Much research has been carried out considering multiple antennas placed at one location. It has been observed that using multiple base stations in a cellular network can benefit

from multiple antenna techniques. This has been called coordinated multipoint (CoMP) transmission and reception[1] [47–49, 112]. By coordinating transmissions from multiple base stations (BSs), the SINR can be lowered and higher data rates can be achieved, especially to the user mobiles located at cell edges. The large body of work that exists on multiple antenna techniques can be adapted to CoMP transmission and reception. Three major methods of CoMP transmission and reception are possible. In the first method, joint transmission (JT), data destined for a particular mobile are jointly processed, precoded and transmitted from several BSs. This requires that the user data and channel state information be available at all coordinating base stations. This can be done fairly expediently if all BSs are connected by high speed fibre optic links, but the processing is very complex and the signalling load is enormous. Optimal joint processing is very difficult and may not be possible in a realistic network. A more practical method is coordinated beamforming (CB) in which multiple bases stations jointly optimize their beamforming vectors when transmitting to mobiles in order to reduce inter-cell interference. The cooperating base stations require channel knowledge of mobiles throughout the cooperating area. In a simpler method of interference avoidance, coordinated scheduling (CS), base stations jointly coordinate the scheduling of data transmission to mobiles.

CoMP transmission and reception techniques mitigate inter-cell interference and hence enable significant network capacity gains achievable through spatial multiplexing at relatively high signal-to-interference ratios. Both MH relaying and CoMP transmission and reception techniques have strengths, and it is expected that a suitable combination of these techniques can provide superior performance results. Since MH relaying subdivides cells into subcells with the addition of RSs, CoMP transmission and reception techniques can also be applied between RSs in separate subcells. The combination of MH and CoMP transmission and reception has great promise for future cellular systems, and it is necessary to study how they may be used jointly.

[1] CoMP transmission and reception is also known in the literature as network MIMO.

Bibliography

[1] "Universal mobile telecommunications system (UMTS); spatial channel model for multiple input multiple output (MIMO) simulations," ETSI, Tech. Rep. 3GPP TR 25.996 Version 6.1.0 Release 6, September 2003.

[2] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16-2004.

[3] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, IEEE Std. 802.16e-2005.

[4] D. S. Baum *et al.*, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *Proc. VTC2005-Spring*, Stockholm, Sweden, May 2005, pp. 3132–3136.

[5] J. Khun-Jush, P. Schramm, U. Wachsmann, and F. Wenger, "Structure and performance of the HiperMAN/2 physical layer," in *Proc. VTC1999-Fall*, Amsterdam, The Netherlands, September 1999, pp. 19–22.

[6] C. Hoymann and B. Walke, "WiMAX Wireless Metropolitan Area Networks," in *European Wireless 2005 (Tutorial)*, April 2005.

[7] M. Cao *et al.*, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *Proc. 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Urbana-Champaign, IL, USA, May 2005, pp. 78 – 89.

[8] H. Bölcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[9] Y. Fan and J. S. Thompson, "On the outage capacity of MIMO multihop networks," in *Proc. Globecom 2005*, vol. 4, December 2005, pp. 2204–2208.

[10] Y. Fan, J. Thompson, A. Adinoyi, and H. Yanikomeroglu, "On the diversity-multiplexing tradeoff for multi-antenna multi-relay channels," in *Proc. ICC '07*, June 2007, pp. 5252 –5257.

[11] Y. Fan, C. Wang, H. Poor, and J. Thompson, "Cooperative multiplexing: Toward higher spectral efficiency in multiple-antenna relay networks," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 3909 –3926, September 2009.

[12] R. Iyengar, P. Iyer, and B. Sikdar, "Analysis of 802.16 based last mile wireless networks," in *Proc. IEEE GLOBECOM '05*, vol. 5, December 2005, pp. 3132–3127.

[13] "Requirements, evaluation criteria and submission templates for the development of IMT-Advanced," ITU-R, Tech. Rep. ITU-R M.2133, November 2008.

[14] "Requirements related to technical performance for IMT-Advanced radio interface(s)," ITU-R, Tech. Rep. ITU-R M.2134, November 2008.

[15] "Guidelines for evaluation of radio interface technologies for IMT-Advanced," ITU-R, Tech. Rep. ITU-R M.2135, November 2008.

[16] "Correction of typographical errors and provision of missing texts of IMT-Advanced channel models in Report ITU-R M.2135," ITU-R, Tech. Rep. ITU-R IMT-ADV/3, July 2009.

[17] A. Urie, "The future of mobile networking will be small cells," Alcatel-Lucent, Tech. Rep., September 2009, presentation to IEEE IOFC Workshop, PIMRC'09, Tokyo, Japan.

[18] "Submission of a candidate IMT-Advanced RIT based on IEEE 802.16, parts 1 to 4," IEEE, Tech. Rep. IEEE L802.16-09/010Xr5, October 2009.

[19] "3rd generation partnership project; technical specification group radio access network; feasibility study for further advancements for E-UTRA (LTE-Advanced) (release 9)," 3GPP, Tech. Rep. 3GPP TR 36.912 V9.0.0, September 2009.

[20] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Broadband Wireless Access Systems*, IEEE Std. 802.16-2009, May 2009.

[21] *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Broadband Wireless Access Systems, Amendment 1- Multiple Relay Specification*, IEEE Std. 802.16j-2009, June 2009.

[22] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572–584, September 1979.

[23] G. Farhadi and N. Beaulieu, "On the ergodic capacity of wireless relaying systems over rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4462–4467, November 2008.

[24] G. Farhadi and N. C. Beaulieu, "On the ergodic capacity of multi-hop wireless relaying systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2286–2291, May 2009.

[25] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, March 2000.

[26] ——, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877–1849, August 2003.

[27] G. L. Stuber, *Principles of Mobile Communication.* Norwell, MA, USA: Kluwer Academic Publishers, 1996.

[28] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity–Part I: System Description," *IEEE Trans. Commun.*, vol. 51, no. 11, November 2003.

[29] ——, "User Cooperation Diversity–Part II: Implementation Aspects and Performance Analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, November 2003.

[30] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Effcient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, December 2004.

[31] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," in *Proc. ICC 2004*, vol. 6, June 2004, pp. 3369–3373.

[32] X. Zhang, W. Wang, and X. Ji, "Multiuser diversity in multiuser two-hop cooperative relay wireless networks: System model and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 1031–1036, June 2008.

[33] H. Yanikomeroglu, "Fixed and mobile relaying technologies for cellular networks," in *Proc. Second Workshop on Applications and Services in Wireless Networks (ASWN'02), Paris, France*, July 2002, pp. 75–81.

[34] R. Schoenen and B. Walke, "On PHY and MAC performance of 3G-LTE in a multihop cellular environment," in *Proc. WiCom 2007*, September 2007, pp. 926–929.

[35] O. Oyman, N. J. Laneman, and S. Sandhu, "Multihop relaying for broadband wireless mesh networks: From theory to practice," *IEEE Commun. Mag.*, vol. 45, no. 11, pp. 116–122, November 2007.

[36] S. W. Peters and R. Heath Jr, "The future of WiMAX: Multihop relaying with IEEE 802.16j," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 104–111, January 2009.

[37] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, "Relay architectures for 3GPP LTE-Advanced," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, pp. 1–14, March 2009.

[38] J. Sydir and R. Taori, "An evolved cellular system architecture incorporating relay stations - [WiMAX update]," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 115–121, June 2009.

[39] "IEEE 802.16m system description document," IEEE, Tech. Rep. 802.16m-09/0034r2, September 2009.

[40] S. Parkvall, E. Dahlman, A. Furuskar, Y. Jading, M. Olsson, S. Wanstedt, and K. Zangi, "LTE-Advanced - Evolving LTE towards IMT-Advanced," in *Proc. VTC 2008 Fall*, September 2008, pp. 1–5.

[41] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[42] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[43] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Tel.*, vol. 10, no. 6, pp. 585–596, November 1999.

[44] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[45] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[46] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1859 – 1874, September 2004.

[47] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56 –61, August 2006.

[48] J. Andrews, W. Choi, and R. Heath, "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Commun. Mag.*, vol. 14, no. 6, pp. 95 –104, December 2007.

[49] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748 –1759, May 2010.

[50] K. Jacobson and W. A. Krzymień, "Minimizing co-channel interference in wireless relay networks," in *Proc. Wireless 2005*, Calgary, Canada, July 2005, pp. 339–346.

[51] ——, "Realistic throughput of cellular multi-hop relay networks with spatial reuse," in *Proc. WPMC-2006*, San Diego, USA, September 2006, pp. 966–970.

[52] ——, "Cell dimensioning and network throughput in cellular multi-hop relay networks," in *Proc. VTC2006-Fall*, Montreal, Canada, September 2006.

[53] ——, "System design and throughput analysis for multi-hop relaying in cellular systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4514 –4528, October 2009.

[54] ——, "Multi-hop relaying and multiple antenna techniques - performance trade-offs in cellular systems," in *Proc. WPMC-2007*, Jaipur, India, December 2007.

[55] ——, "Multi-hop relaying and MIMO techniques in cellular systems - throughput achievable on Rayleigh/Ricean channels," in *Proc. Globecom 2008*, New Orleans, Louisiana, November 2008.

[56] ——, "Multihop relaying and multiple antenna techniques - performance trade-offs in cellular systems," *EURASIP Journal on Wireless Communications and Networking*, submitted September, 2010.

[57] ——, "Cooperative vs multihop relaying in cellular systems," in *Proc. PIMRC 2009*, Tokyo, Japan, September 2009.

[58] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.

[59] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.

[60] J. Salo, F. Mikas, and P. Vainikainen, "An upper bound on the ergodic mutual information of Rician fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1415–1421, June 2006.

[61] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, October 2003.

[62] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, October 2004.

[63] J. Zhang and T. M. Lok, "Performance comparison of conventional and cooperative multihop transmission," in *Proc. WCNC 2006*, vol. 2, April 2006, pp. 897–901.

[64] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1820–1830, October 2004.

[65] R. Nabar, F. Kneubuhler, and H. Bolcskei, "Performance limits of amplify-and-forward based fading relay channels," in *Proc. ICASSP '04*, vol. 4, May 2004, pp. iv–565 – iv–568 vol.4.

[66] R. Nabar, H. Bolcskei, and F. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1099 – 1109, August 2004.

[67] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152 –4172, December 2005.

[68] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, July 1949.

[69] R. A. Horn and C. R. Johnson, *Matrix analysis*. New York, NY, USA: Cambridge University Press, 1985.

[70] M. Hasna and M.-S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 216–218, May 2003.

[71] M. Hasna, "Average BER of multihop communication systems over fading channels," in *Proc. ICECS 2003*, vol. 2, 12 2003, pp. 723–726.

[72] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: a new architecture for wireless communications," in *Proc. INFOCOM 2000*, vol. 3, March 2000, pp. 1273–1282.

[73] C. Qiao and H. Wu, "iCAR: an integrated cellular and ad-hoc relay system," in *Proc. Computer Communications and Networks, 2000*, October 2000, pp. 154–161.

[74] H. Wu, C. Qiao, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 2105 –2115, October 2001.

[75] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "A characterization of multihop wireless communications channels," in *Proc. CWIT'01*, June 2001.

[76] V. Sreng, H. Yanikomeroglu, and D. Falconer, "Coverage enhancement through two-hop relaying in cellular radio systems," in *Proc. WCNC2002*, vol. 2, March 2002, pp. 881–885.

[77] H. Bolukbasi, H. Yanikomeroglu, D. Falconer, and S. Periyalwar, "On the capacity of cellular fixed relay networks," in *Proc. Canadian Conf. on Elect. and Comp. Eng. 2004*, vol. 4, May 2004, pp. 2217–2220.

[78] D. C. Schultz, B. Walke, R. Pabst, and T. Irnich, "Fixed and planned relay based radio network deployment concepts," in *Proc. 10th Wireless World Research Forum*, WWRF. New York, USA: The Institution of Electrical Engineers, ISBN 0-85296-753-5, October 2003.

[79] T. Irnich, D. C. Schultz, R. Pabst, and P. Wienert, "Capacity of a relaying infrastructure for broadband radio coverage of urban areas," in *Proc. VTC2003-Fall*, vol. 5, October 2003, pp. 2886–2890.

[80] N. Esseling, B. Walke, and R. Pabst, "Performance of MAC-frame-based protocols for mobile broadband systems using layer 2 relays," in *Proc. 11th Wireless World Research Forum, WWRF*, Oslo, Norway, June 2004.

[81] R. Pabst *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, pp. 80–89, September 2004.

[82] N. Esseling, B. Walke, and R. Pabst, "Performance evaluation of a fixed relay concept for next generation wireless systems," in *Proc. PIMRC 2004*, vol. 2, Barcelona, Spain, September 2004, pp. 744–751.

[83] "Call for contributions IEEE 802.16 relay task group," IEEE, Tech. Rep. IEEE-802.16j-06/001, May 2006.

[84] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16j relay-based wireless access networks: an overview," *IEEE Wireless Commun. Mag.*, vol. 15, no. 5, pp. 56–63, October 2008.

[85] O. Teyeb, V. Van Phan, B. Raaf, and S. Redana, "Dynamic relaying in 3GPP LTE-Advanced networks," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, July 2009.

[86] M. Dohler, E. Lefranc, and H. Aghvami, "Virtual antenna arrays for future wireless mobile communication systems," in *Proc. ICT 2002*, June 2002.

[87] M. Dohler, J. Dominguez, and H. Aghvami, "Link capacity analysis for virtual antenna arrays," in *Proc. VTC2002-Fall*, vol. 1, September 2002, pp. 440–443.

[88] M. Dohler, "Virtual antenna arrays," Ph.D. dissertation, University of London, London, November 2003.

[89] K. Hole and G. E. Oien, "Spectral efficiency of adaptive coded modulation in urban microcellular networks," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 205–222, January 2001.

[90] M.-S. Alouini and A. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1047–1066, July 1999.

[91] V. Erceg *et al.*, "Channel models for fixed wireless applications," IEEE, Tech. Rep. IEEE-C802.16.3c-01/29r4, July 2001.

[92] D. C. Schultz, R. Pabst, and T. Irnich, "Multi-hop based radio network deployment for efficient broadband radio coverage," in *Proc. WPMC'03*, Yokosuka, Japan, October 2003, pp. 377–381.

[93] L. J. Greenstein, S. Ghassemzadeh, V. Erceg, and D. G. Michelson, "Ricean K-factors in narrowband fixed wireless channels," in *Proc. WPMC-1999*, September 1999.

[94] F. Bohagen, P. Orten, and G. E. Oien, "Construction and capacity analysis of high-rank line-of-sight MIMO channels," in *Proc. Wireless Communications and Networking Conference, 2005 IEEE*, vol. 1, March 2005, pp. 432–437.

[95] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY: John Wiley and Sons, 1991.

[96] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41st Allerton Conf. on Comm., Control and Comp.*, vol. 1, October 2003.

[97] S. Toumpis, R. Muller, and J. Sayir, "On the transport capacity of a multiple access Gaussian channel," in *Proc. 2004 International Workshop on Wireless Ad-Hoc Networks*, June 2004, pp. 191–195.

[98] S. Toumpis and A. J. Goldsmith, "Performance bounds for large wireless networks with mobile nodes and multicast traffic," in *Proc. 2004 International Workshop on Wireless Ad-Hoc Networks*, June 2004, pp. 125–129.

[99] J. Jun and M. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 8–14, October 2003.

[100] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 736–748, July 2003.

[101] R. Nelson and L. Kleinrock, "Spatial TDMA: a collision-free multihop channel access protocol," *IEEE Trans. Commun.*, vol. 33, no. 9, pp. 934–944, September 1985.

[102] M. Nakagami, *The m distribution; a general formula of intensity distribution of rapid fading*, W. Hoffman, Ed. Pergamon Press, 1960.

[103] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: Mcgraw-Hill Higher Education, 2001.

[104] S. Zurbes, "Power control in simulcast digital cellular radio networks," in *Proc. PIMRC 1997*, vol. 3, September 1997, pp. 887–891.

[105] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," in *Proc. INFOCOM 2004*, vol. 1, March 2004, pp. 464–475.

[106] N. Esseling, R. Pabst, and B. Walke, "Delay and throughput analysis of a fixed relay concept for next generation wireless systems," in *Proc. 11th European Wireless Conference 2005*, vol. 1, Nikosia, Cyprus, April 2005, pp. 273–279.

[107] N. Bayer, D. Sivchenko, B. Xu, V. Rakocevic, and J. Habermann, "Transmission timing of signaling messages in IEEE 802.16 based mesh networks," in *Proc. European Wireless 2006*, April 2006.

[108] Y. Yao and J. Sun, "Study of UGS grant synchronization for 802.16," in *Proc. Ninth International Symposium on Consumer Electronics, 2005. (ISCE 2005)*, June 2005, pp. 105–110.

[109] S. Maheshwari, "An efficient QoS scheduling architecture for IEEE 802.16 wireless MANs," Master's thesis, Indian Institute of Technology Bombay, Bombay, India, 2005.

[110] K.-C. Chu and J.-W. Lin, "Pipelined multi-hop automatic repeat request mechanism in IEEE 802.16j," in *Proc. WiCOM '08.*, October 2008, pp. 1 –5.

[111] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774 –1786, May 2007.

[112] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [Coordinated and Distributed MIMO]," *IEEE Wireless Commun. Mag.*, vol. 17, no. 3, pp. 26 –34, June 2010.

[113] J. Salo, F. Mikas, and P. Vainikainen, "An upper bound on the ergodic mutual information of Ricean fading MIMO channels," in *Proc. 6th Nordic Signal Processing Symposium*, June 2004, pp. 288–291.

Appendix A

System Model Validation

A.1 Validation of the Salo Bound on Capacity of a Ricean Channel

A.1.1 Monte Carlo Simulation

A common method for finding the performance of a MIMO channel is Monte Carlo simulation in which a large number of samples of random variables are generated and calculations are made for each sample. Results are often displayed using cumulative distribution functions (CDFs) and means of the calculations. From theory we have the following formula for capacity in a $N_T \times N_R$ MIMO system (equal power case)

$$C_{EP} = \log_2 \left[\det \left(\mathbf{I}_{N_R} + \frac{E_x}{N_T N_0} \mathbf{H}\mathbf{H}^H \right) \right] \quad \text{b/s/Hz} \tag{A.1}$$

where E_x is the symbol energy, and N_0 is the noise variance. The channel matrix, $\mathbf{H}$, is a random variable, and thus capacity is also a random variable. A large number of channel matrices (samples) are generated, and the capacity is found for each sample using the above formula. From this set of values, we can determine the mean, variance, and CDF of the capacity. The mean capacity is called ergodic capacity, $\bar{C}_{EP}$.

$$\bar{C}_{EP} \triangleq \mathcal{E}_{\mathbf{H}}[C_{EP}] \tag{A.2}$$

where $\mathcal{E}_{\mathbf{H}}[*]$ is the average over random variable $\mathbf{H}$.

In order to make this calculation we must use a finite number of samples. As a result we do not obtain the exact mean, but rather an estimate of the mean. This method is very useful, but often a large number of samples may be necessary to obtain reliable results, which for a complicated system may require significant computation time. An alternative method finds an upper bound, as described in the next section.

A.1.2 Salo Upper Bound

Upper bounds for the average mutual information of Ricean MIMO channels are given in [113] and [60]. The general case and three different cases are given in these papers. The results assume that the receiver has perfect channel knowledge, and the transmitter does not have channel knowledge. Thus the transmitter allocates equal power to each of the transmit antennas.

The simplified three special cases are:

1. Ricean channel with independent and identically distributed (iid) Rayleigh $\mathbf{H}_{NLOS}$
2. Ricean channel with iid Rayleigh $\mathbf{H}_{NLOS}$, and $r_{LOS} = \text{rank}(\mathbf{H}_{LOS}) = 1$
3. semicorrelated (antenna array is uncorrelated at one end only) Rayleigh $\mathbf{H}$ (i.e. $K_r = 0$)

In our research we use the first two cases. Special case number 1 (Corollary 1) gives the upper bound for the average mutual information $\mathcal{E}_{\mathbf{H}}[I_{\mathbf{H}}]$ of a Ricean channel

$$\mathcal{E}_{\mathbf{H}}[I_{\mathbf{H}}] \leq \log_2 \left[\sum_{p=0}^{K} \left(\frac{\rho b^2}{N_T} \right)^p \sum_{j=0}^{p} K_r^j \left(L-p+1 \right)_{(p-j)} \binom{K-j}{p-j} \text{tr}_j(\mathbf{T}) \right] \tag{A.3}$$

where ρ is the SINR, $K = \min(N_R, N_T)$, $b = \sqrt{\frac{1}{K_r+1}}$, $L = \max(N_R, N_T)$, $(m)_n$ is the Pochhammer symbol given by

$$(m)_n = m(m+1)\ldots(m+n+1) \tag{A.4}$$

$\mathbf{T} = \mathbf{H}_{LOS}\mathbf{H}_{LOS}^H$, and $\text{tr}_j(\mathbf{T})$ is the j^{th} elementary symmetric function of $\mathbf{T}$ (see [60] and [69]).

Special case number 2 (Corollary 2) gives the upper bound for the average mutual information $\mathcal{E}_{\mathbf{H}}[I_{\mathbf{H}}]$ of a Ricean channel with rank 1 $\mathbf{H}_{LOS}$

$$\mathcal{E}_{\mathbf{H}}[I_{\mathbf{H}}] \leq \log_2 \left[1 + \sum_{p=1}^{K} \sum_{j=0}^{1} \left(\frac{\rho b^2}{N_T} \right)^p \left(K_r K L \right)^j \left(L-p+1 \right)_{(p-j)} \binom{K-j}{p-j} \right] \tag{A.5}$$

$\mathbf{H}_{LOS}$ is deterministic and must be normalized appropriately for fair comparison. According to [94] antenna arrays can be designed to make $\mathbf{H}_{LOS}$ full rank which creates orthogonal rows.

For a $N_T \times N_R$ MIMO system, the normalized $\mathbf{H}_{LOS}$ is

$$\mathbf{H}_{LOS} = C_{norm}\mathbf{H}'_{LOS} \tag{A.6}$$

where

$$C_{norm} = \frac{\sqrt{N_T N_R}}{\|\mathbf{H}'_{LOS}\|_F} \tag{A.7}$$

and $\| \star \|_F$ is the Frobenius norm.

For a 2×2 MIMO system, a full rank channel matrix is

$$\mathbf{H}_{LOS,full} = \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{A.8}$$

and a rank 1 channel matrix is

$$\mathbf{H}_{LOS,rank1} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{A.9}$$

For a 3×3 MIMO system, a full rank channel matrix is

$$\mathbf{H}_{LOS,full} = \sqrt{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.10}$$

and a rank 1 channel matrix is

$$\mathbf{H}_{LOS,rank1} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{A.11}$$

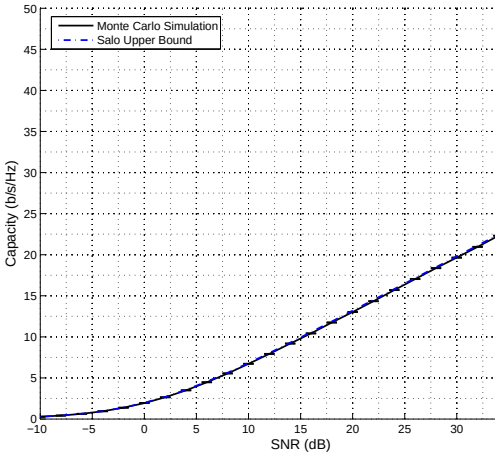
Full rank $\mathbf{H}_{LOS}$ can exist in short distance wireless links, such as encountered in wireless local area networks (WLANs), but for greater distances, such as those typical in cellular systems, $\mathbf{H}_{LOS}$ is usually rank 1. We will consider both full rank and rank 1 situations here.

A.1.3 Verifying the Salo Bound

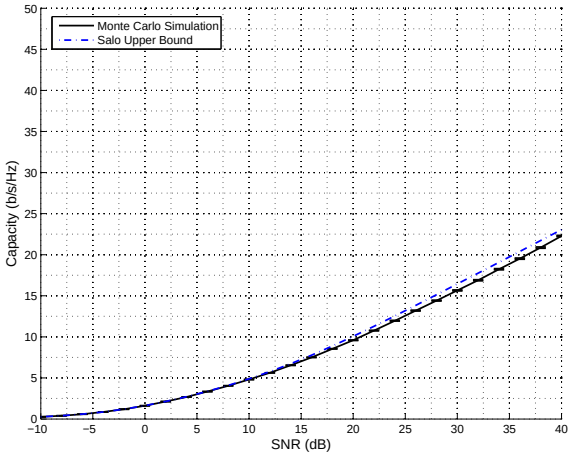
Since we wish to use these upper bounds (A.3) and (A.5) extensively in our work, it is useful to verify how accurate they are. In this section we compare the Salo bounds to Monte Carlo results.

Figs. A.1 and Figs. A.2 shows the average mutual information for 2x2, 3x3, 4x4 and 6x6 MIMO, comparing results obtained by both the Salo upper bound for special case 1 given by expression (A.3) and Monte Carlo simulation. It is clear that the results match very closely for all the scenarios, and thus we can conclude that the Salo bound is very tight.

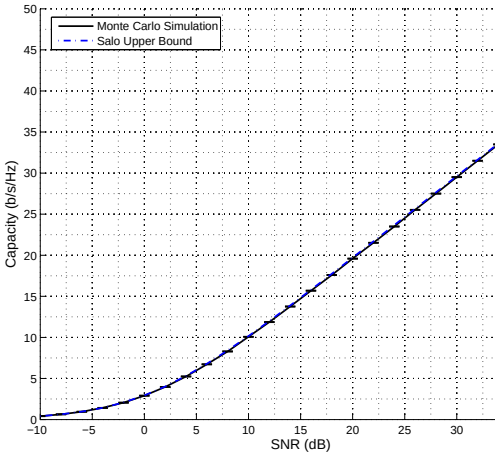
Using a Matlab program with 2000 samples, the Monte Carlo method can take up to 10 seconds to generate data for each of the figures. Calculating with the Salo upper bound is much quicker, and it is clear that we can obtain nearly identical results with it. Note that these calculations are for only one hop, so when performing these calculations for multiple relay hops and for numerous one dimensional and two dimensional configurations, the simulation time will become significant.



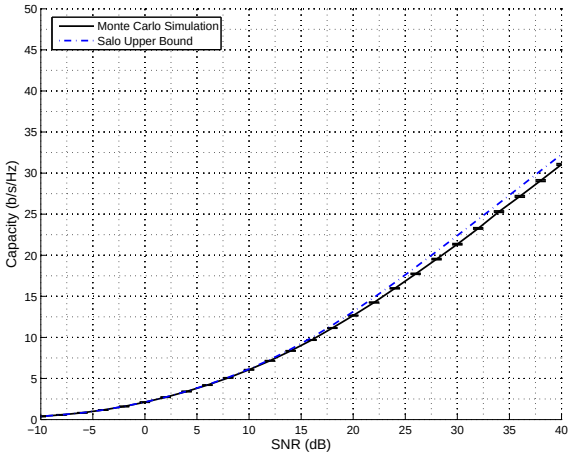
(a) 2x2 MIMO, full rank $\mathbf{H}_{LOS}$.



(b) 2x2 MIMO, rank 1 $\mathbf{H}_{LOS}$.

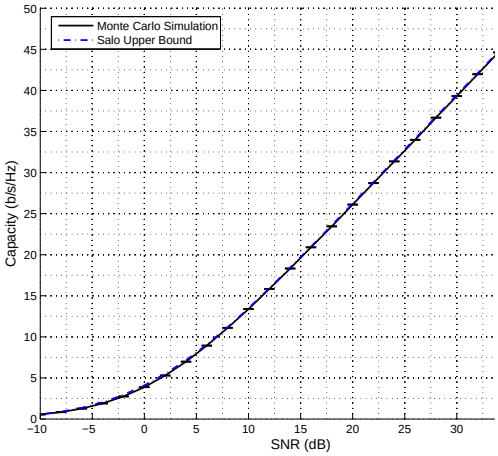


(c) 3x3 MIMO, full rank $\mathbf{H}_{LOS}$.

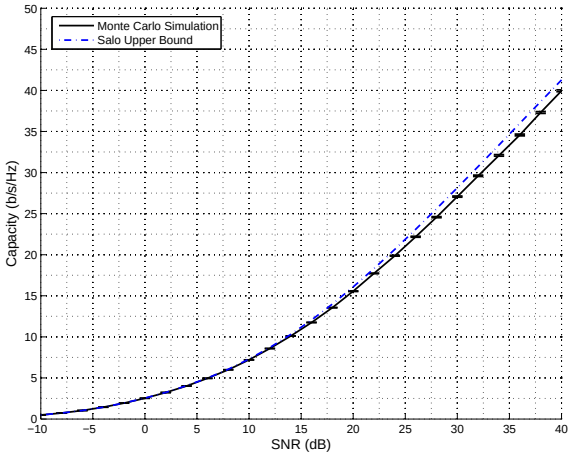


(d) 3x3 MIMO, rank 1 $\mathbf{H}_{LOS}$.

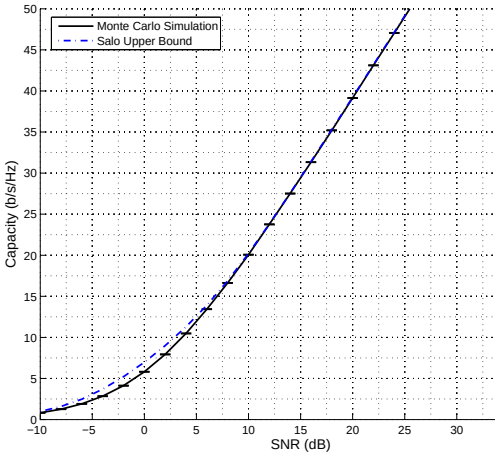
Figure A.1: Comparing the Salo bound to Monte Carlo simulation capacity results: $K_r = 10$, 2000 samples.



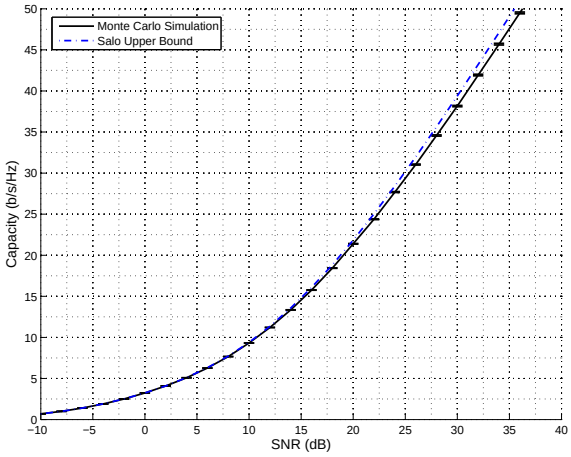
(a) 4x4 MIMO, full rank $\mathbf{H}_{LOS}$.



(b) 4x4 MIMO, rank 1 $\mathbf{H}_{LOS}$.



(c) 6x6 MIMO, full rank $\mathbf{H}_{LOS}$.



(d) 6x6 MIMO, rank 1 $\mathbf{H}_{LOS}$.

Figure A.2: Comparing the Salo bound to Monte Carlo simulation capacity results: $K_r = 10$, 2000 samples.

Table A.1: Simulation time and maximum $\tilde{S}_E$ for Monte Carlo sampling, 4x4 MIMO, $K_r = 10$, rank 1 $\mathbf{H}_{LOS}$.

$N_{samp} \rightarrow$	2000	200	100	20	10
Time (sec.)	7.0	0.87	0.53	0.26	0.23
Max. $\tilde{S}_E$ (%)	0.3	1.1	1.3	3.6	8.7

A.1.4 Effect of Number of Samples in Monte Carlo Simulation

In the previous section we used 2000 samples in the Monte Carlo simulation to guarantee accurate results, but we did not provide any justification for this. In this section we are interested in determining how accurate results are for various sample numbers, so that we can select the appropriate number of samples to use. Having verified in the previous section that the Salo bound is reliable, we show it in each of the following figures as a reference. Accuracy is shown by error bars in the figures, which are plotted using the *standard error of the mean* given by

$$S_E = \frac{\hat{\sigma}}{\sqrt{N_{samp}}} \tag{A.12}$$

where $\hat{\sigma}$ is the standard deviation resulting from a simulation and N_{samp} is the number of samples used in the simulation. Standard error of the mean tells us how close the estimated mean is to the actual mean. If we average over more samples, the calculated mean is more likely to be closer to the actual mean.

We normalize the standard error by dividing by the estimated mean from a simulation, $\hat{\mu}$, and express it as a percentage

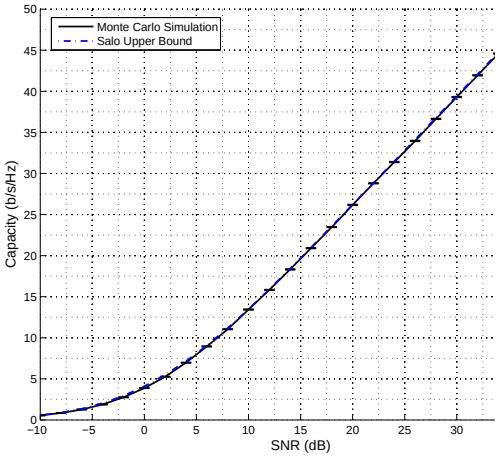
$$\tilde{S}_E = \frac{S_E}{\hat{\mu}} \times 100 \quad (\%) \tag{A.13}$$

For example, an $\tilde{S}_E$ of 1% implies that we are confident that the estimated mean is within 1% of the actual mean.

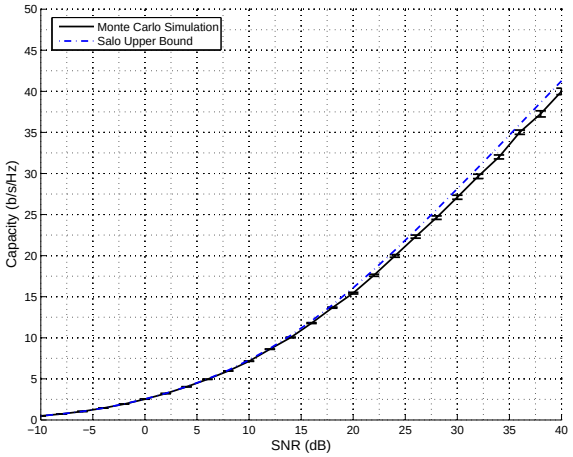
Figs. A.3 and A.4 show the results using 200, 100, 20 and 10 samples. As shown by the width of the error bars in in the figures, the standard error increases as we decrease the number of samples, and scatter becomes more apparent for 20 and 10 samples. However, the results for 100 samples appear to be quite reasonable. Greater inaccuracy occurs for the rank 1 $\mathbf{H}_{LOS}$ (which is deterministic), since in that case the NLOS component, the Rayleigh-distributed $\mathbf{H}_{NLOS}$, is a greater contributor to the capacity.

Table A.1 shows the simulation times (the times required to generate the the data plotted in each figure shown in this section) and the maximum $\tilde{S}_E$ for the rank 1 $\mathbf{H}_{LOS}$ simulations.

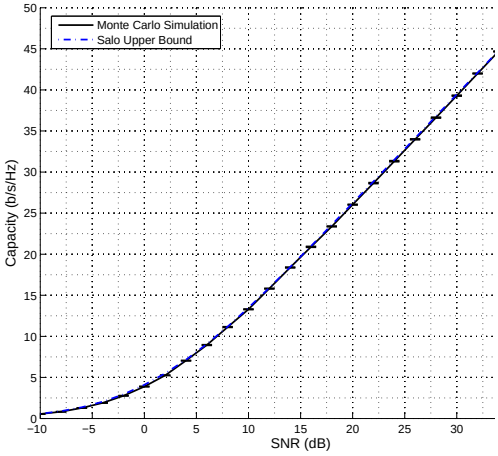
We can conclude that while 2000 samples gives very accurate results (within 0.3% of the actual mean), 200 samples still gives reasonable accuracy, about 1%, while greatly shortening



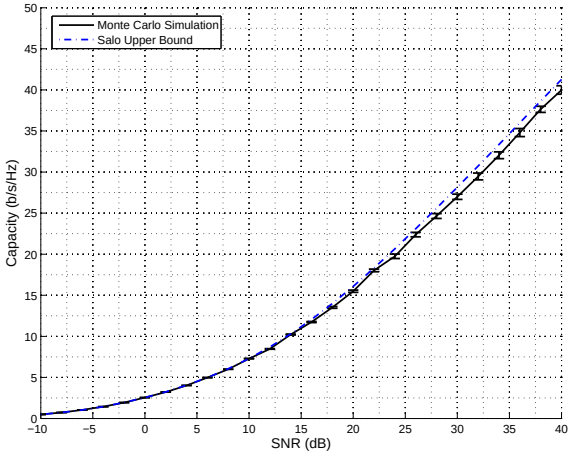
(a) Full rank $\mathbf{H}_{LOS}$, 200 samples.



(b) Rank 1 $\mathbf{H}_{LOS}$, 200 samples.

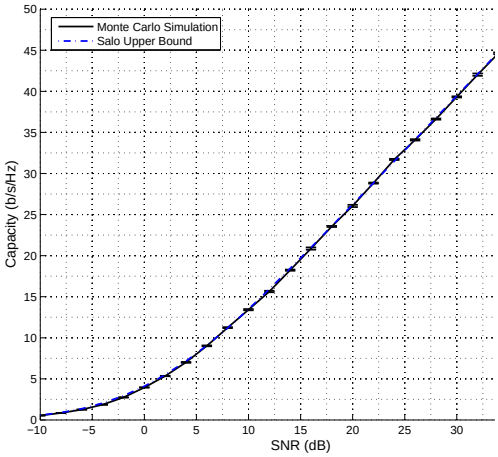


(c) Full rank $\mathbf{H}_{LOS}$, 100 samples.

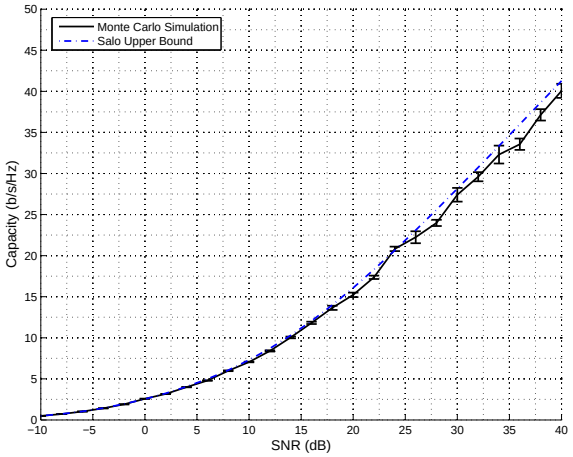


(d) Rank 1 $\mathbf{H}_{LOS}$, 100 samples.

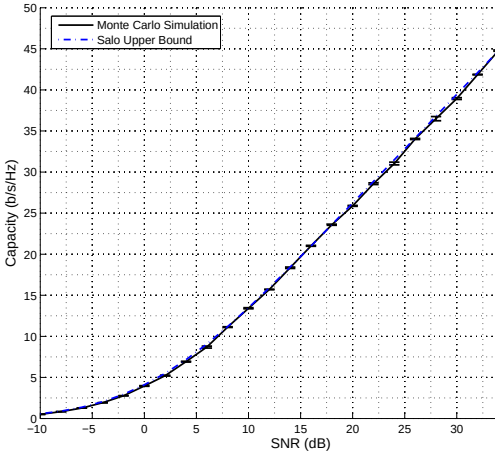
Figure A.3: The effect of the number of Monte Carlo samples, 4x4 MIMO, $K_r = 10$.



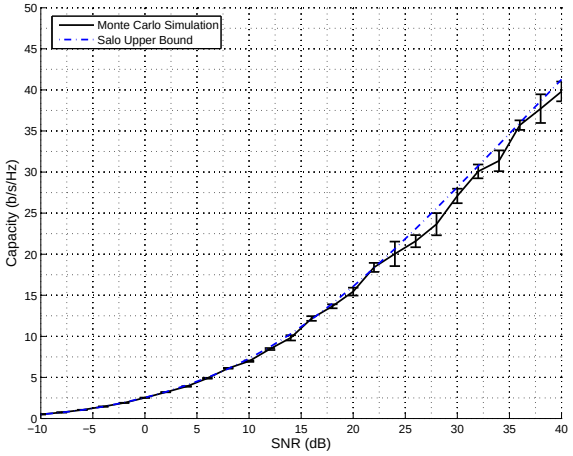
(a) Full rank $\mathbf{H}_{LOS}$, 20 samples.



(b) Rank 1 $\mathbf{H}_{LOS}$, 20 samples.



(c) Full rank $\mathbf{H}_{LOS}$, 10 samples.



(d) Rank 1 $\mathbf{H}_{LOS}$, 10 samples.

Figure A.4: The effect of the number of Monte Carlo samples, 4x4 MIMO, $K_r = 10$.

the simulation time. Thus we can use either the Salo bound or Monte Carlo simulation to quickly obtain good results.

A.1.5 Effect of Log-Normal Shadowing

We have up until now used a model incorporating bulk path loss, a deterministic LOS matrix, and a Rayleigh distributed random NLOS channel matrix. Now we include random shadowing effects by using a combined path loss-shadowing model and employing Monte Carlo simulation as before. We are interested in whether results still match up with previous results, and how many samples are necessary to obtain accurate results. Since we have two separate random variables involved (Rayleigh scattering and log-normal shadowing), we might expect to require more samples in order to obtain accurate results.

As given in Section 3.3.1, the following model can be used for path loss and shadowing at 5 GHz [1, 4], :

$$PL_{dB}(x) = -20\log_{10}[\gamma(x)] = \begin{cases} 42.5 + 38.0\log_{10}(x) + \psi_{dB} & b < x < 5000\text{ m}, \\ 38.2 + 26.0\log_{10}(x) + \psi_{dB} & 20\text{ m} < x < b \end{cases} \tag{A.14}$$

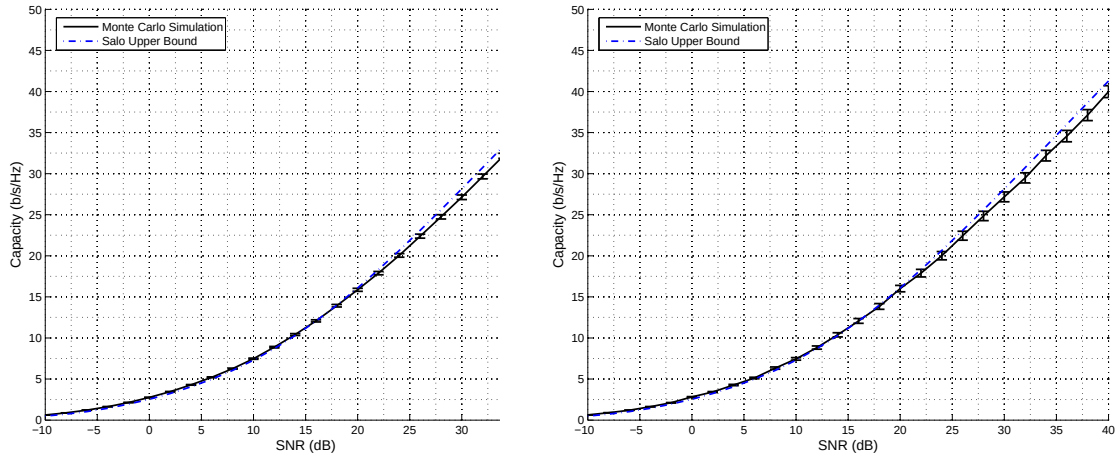
where x is the distance and ψ_{dB} is a Gaussian random variable with zero mean and standard deviation, $\sigma_{\psi_{dB}}$. Typically $\sigma_{\psi_{dB}} = 4$ dB in LOS urban microcells, and $\sigma_{\psi_{dB}} = 8$ dB in NLOS urban or suburban macrocells. Ergodic capacity is found using the link budget and calculation methods described in Section 3.3.1.

To simulate this in Matlab, we generate a large number of Gaussian samples, ψ_{dB}, with zero mean and standard deviation, $\sigma_{\psi_{dB}}$ (using `randn_matrix = randn(n_snrs, n_samp)*std_shad_log;`).

Simulation results for an urban microcell ($\sigma_{\psi_{dB}} = 4$ dB) using 4x4 MIMO, $K_r = 10$, and rank 1 $\mathbf{H}_{LOS}$, are shown in Fig. A.5. The Salo bound is included for reference. As expected, a larger number of samples is required for accurate results due to the existence of another random variable. In Fig. A.6 we show results for $\sigma_{\psi_{dB}} = 8$ dB, using different Rice factors, K_r. With a larger shadowing standard deviation, we require an even larger number of samples.

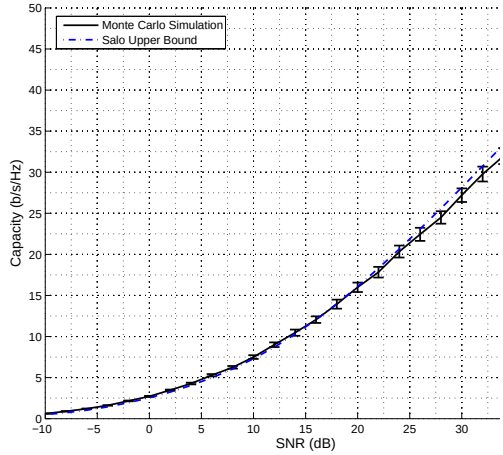
Table A.2 shows the simulation times (the times required to generate each figure shown in this section) and the maximum $\tilde{S}_E$ for the rank 1 $\mathbf{H}_{LOS}$ simulations including log-normal shadowing.

As expected, a larger number of samples are required to give accurate results when random shadowing effects are included. With $\sigma_{\psi_{dB}} = 4$ dB, we require at least 10000 samples to obtain better than 1% accuracy, and greater than 20000 samples with $\sigma_{\psi_{dB}} = 8$ dB. The simulation time is obviously longer, which leads us to select the Salo bounding technique to obtain quick and accurate results when including random shadowing in the model.

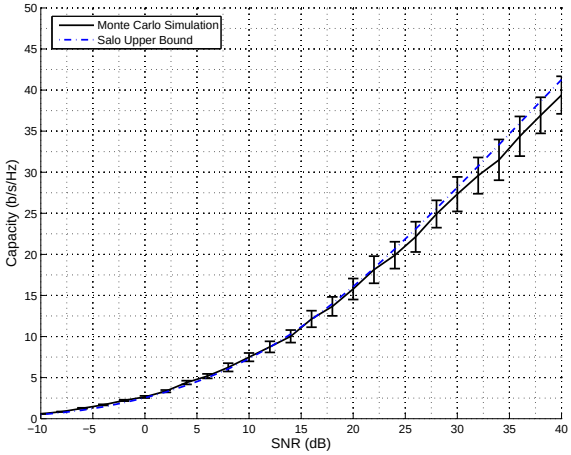


(a) 10000 samples.

(b) 2000 samples.



(c) 1000 samples.

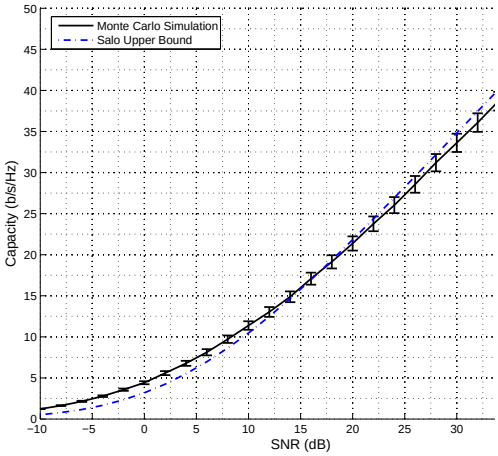


(d) 200 samples.

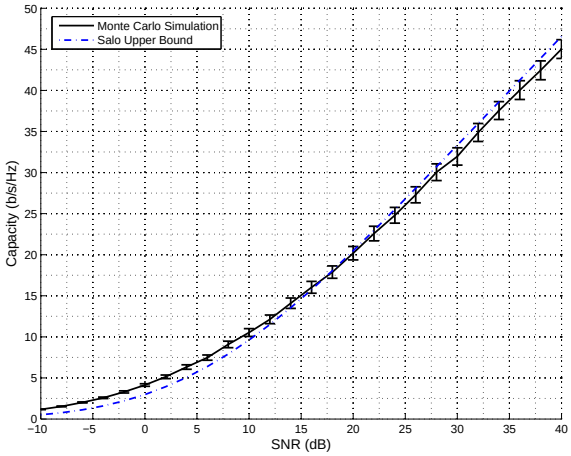
Figure A.5: Effect of including log-normal shadowing with various N_{samp}: 4x4 MIMO, $K_r = 10$, Rank 1 $\mathbf{H}_{LOS}$, $\sigma_{\psi_{dB}} = 4$ dB.

Table A.2: Monte Carlo simulation time and maximum $\tilde{S}_E$, log-normal shadowing, 4x4 MIMO, $K_r = 10$, rank 1 $\mathbf{H}_{LOS}$.

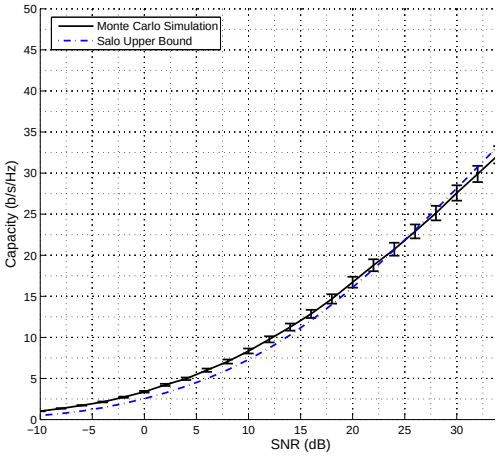
$N_{samp} \rightarrow$	20000	10000	2000	1000	200
$\sigma_{\psi_{dB}} = 4$ dB					
Time (sec.)	73	34.3	6.71	3.54	0.93
Max. $\tilde{S}_E$ (%)	0.81	1.1	2.6	3.9	8.8
$\sigma_{\psi_{dB}} = 8$ dB					
Time (sec.)	71.1	34.0	6.87	3.54	0.88
Max. $\tilde{S}_E$ (%)	2.7	4.0	8.9	13	34



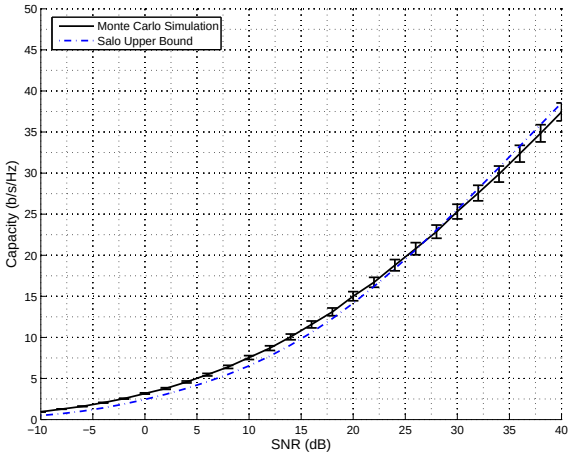
(a) $K_r = 1$.



(b) $K_r = 2$.



(c) $K_r = 10$.



(d) $K_r = 20$.

Figure A.6: Effect of including log-normal shadowing with various K_r: 4x4 MIMO, Rank 1 $\mathbf{H}_{LOS}$, $\sigma_{\psi_{dB}} = 8$ dB, 10000 samples.

Another observation can be made from these experiments. The ergodic capacity, obtained by averaging over the Monte Carlo samples, remains very close whether we use 4 dB shadowing, 8 dB shadowing, or simply use the path loss model without log-normal shadowing. This is not unexpected, since the bulk path loss is in fact the mean over random shadowing.

Appendix B

One Dimensional Relaying Topologies

The following figures show the topologies for one dimensional relay networks for one to five hops. A one dimensional "cell" is the subject of study, and the cells on either side create interference into this cell. The downlink (forward) and uplink (reverse) links are shown in blue and red, respectively, and the distances to the relevant interfering stations in other cells can be observed. Cluster sizes of $N = 1, 2, 3, 4$ were studied, and the frequency reuse patterns are shown in the figures.

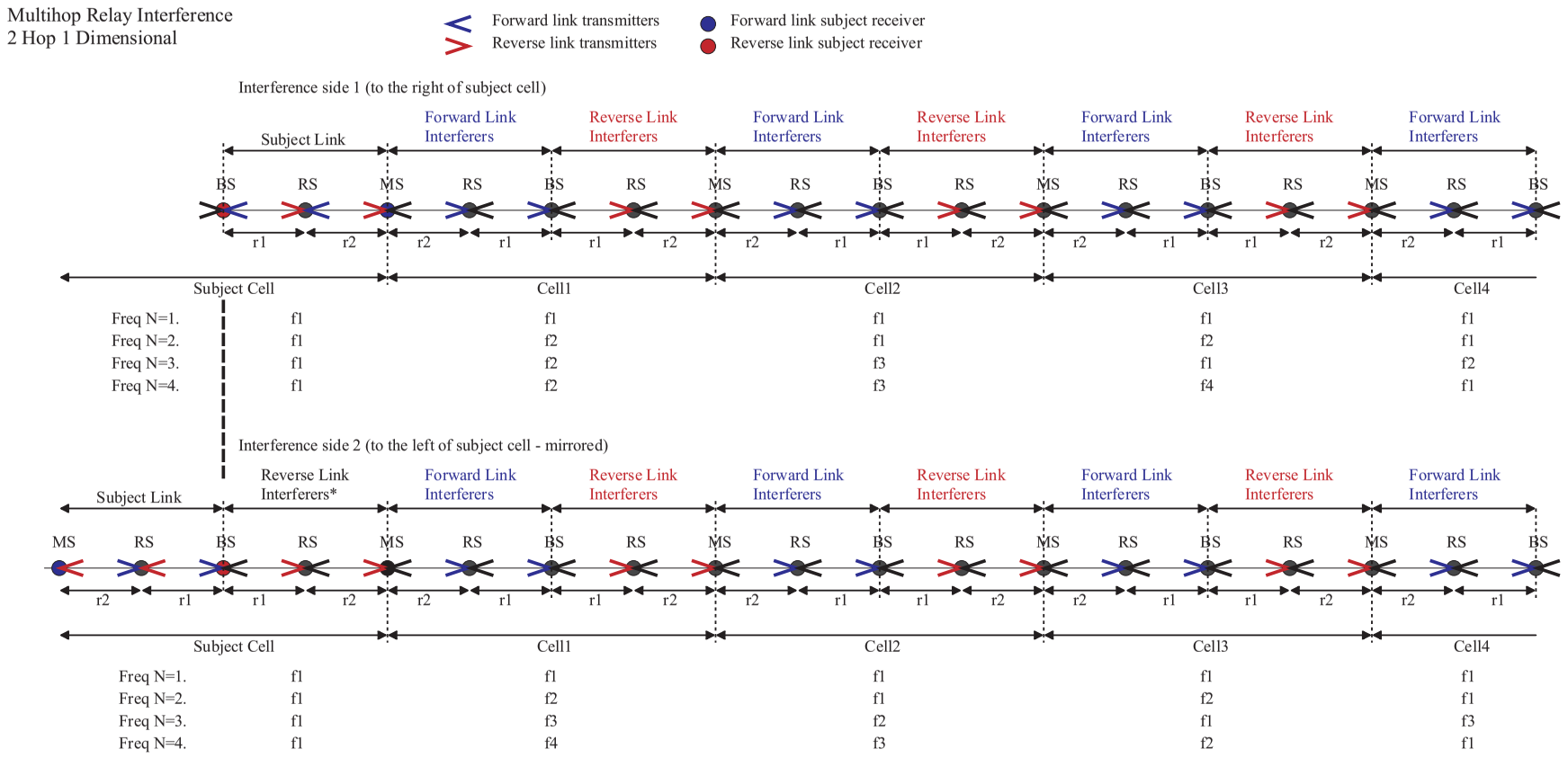


Figure B.1: A two-hop one-dimensional linear multihop relay system, showing four tiers of interference.

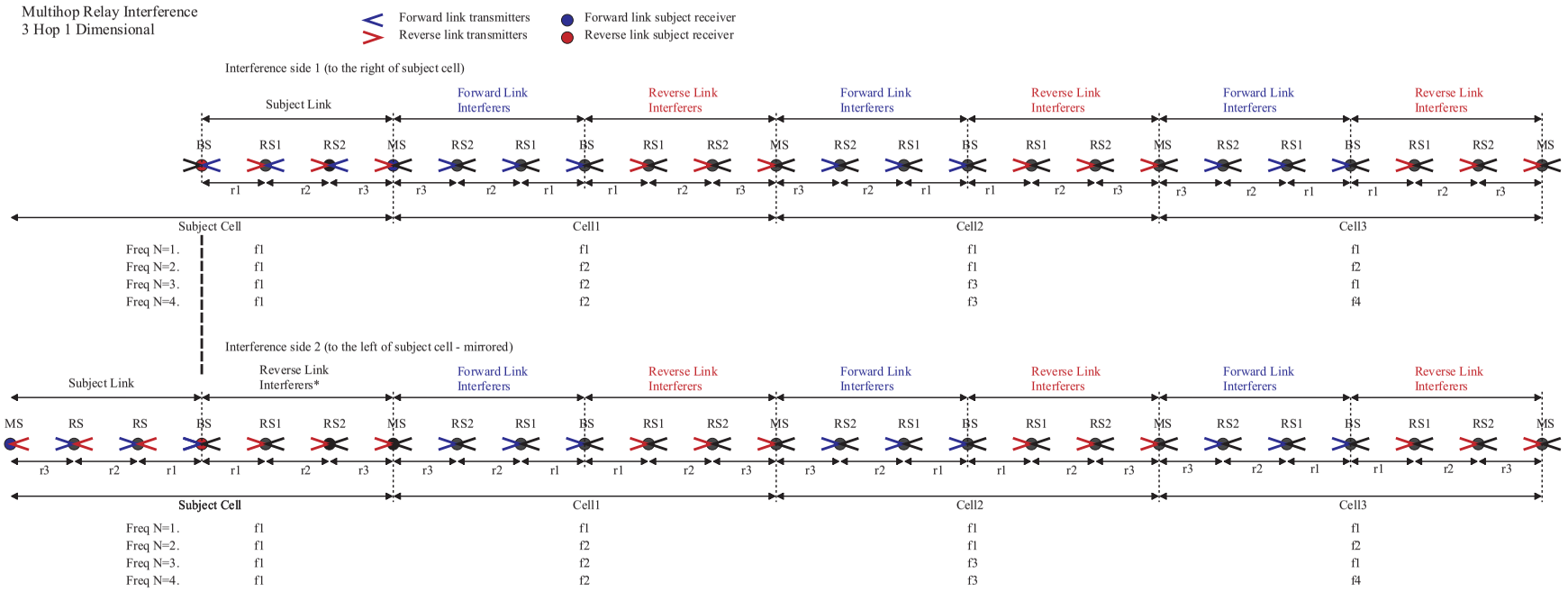


Figure B.2: A three-hop one-dimensional linear multihop relay system, showing three tiers of interference.

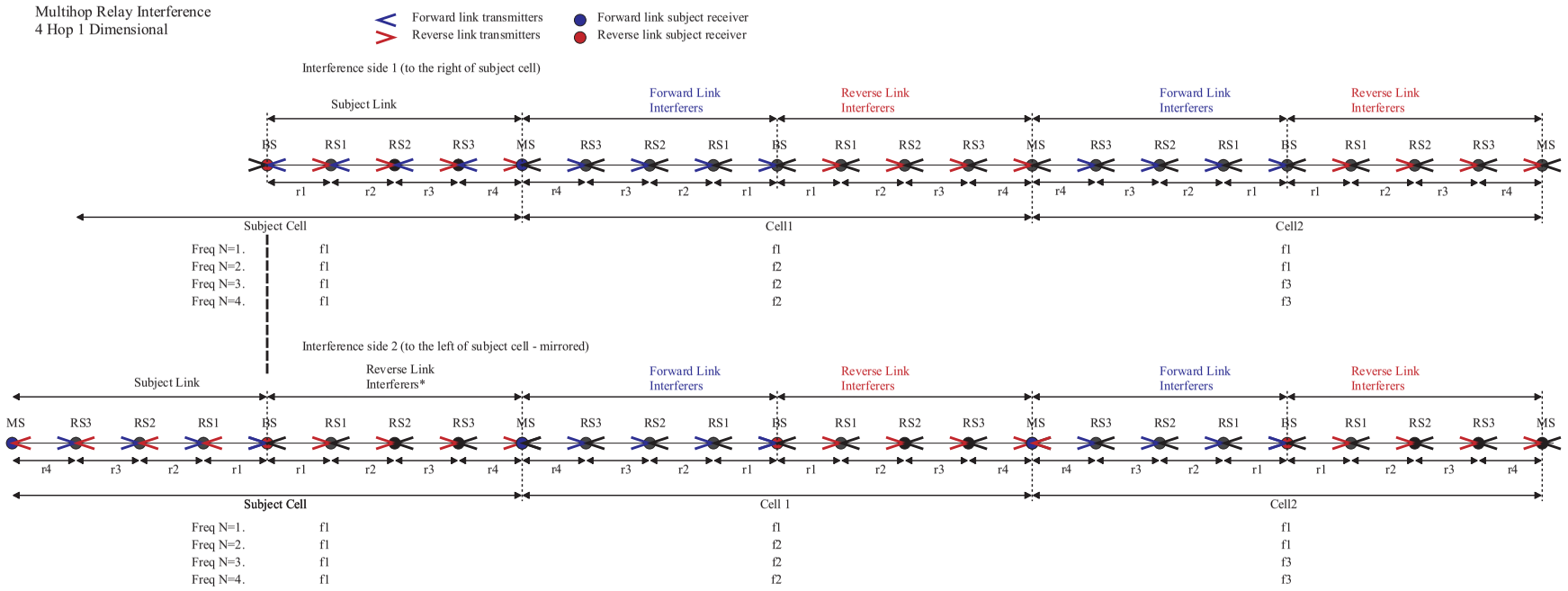


Figure B.3: A four-hop one-dimensional linear multihop relay system, showing two tiers of interference.

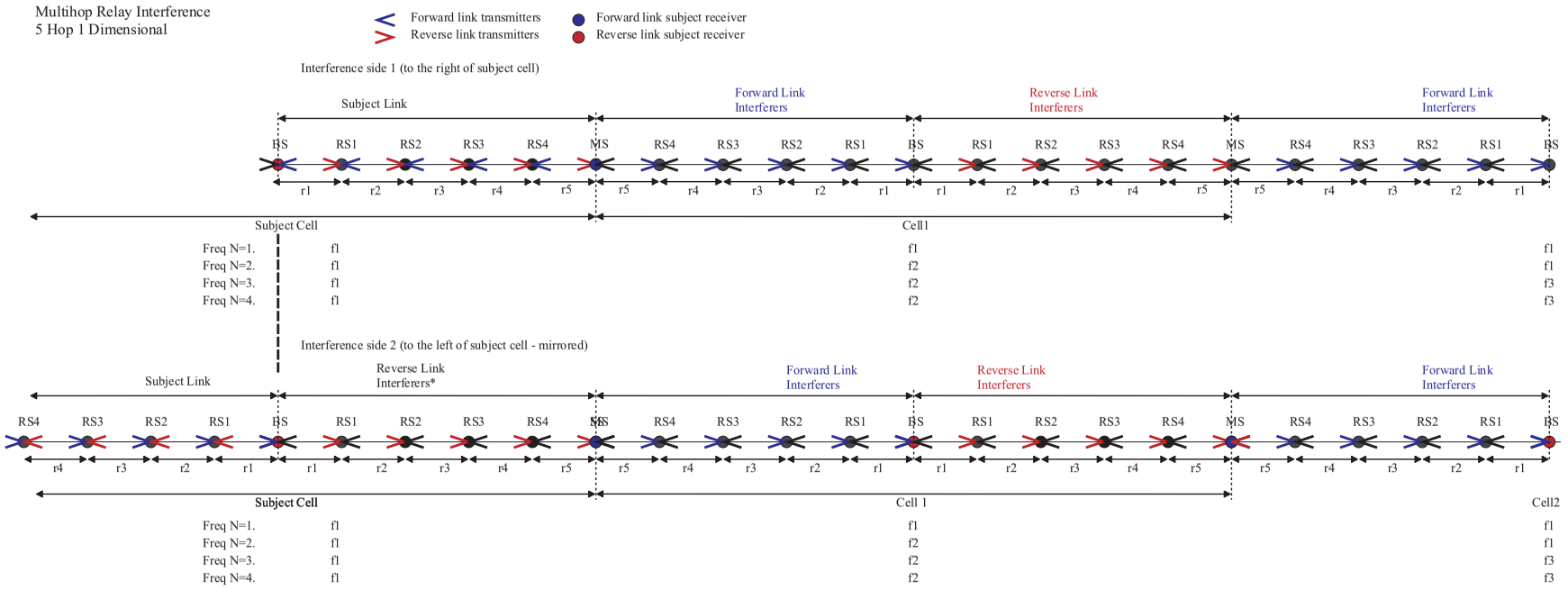
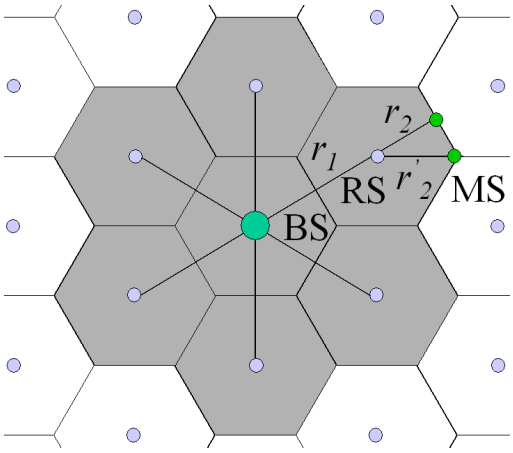


Figure B.4: A five-hop one-dimensional linear multihop relay system, showing two tiers of interference.

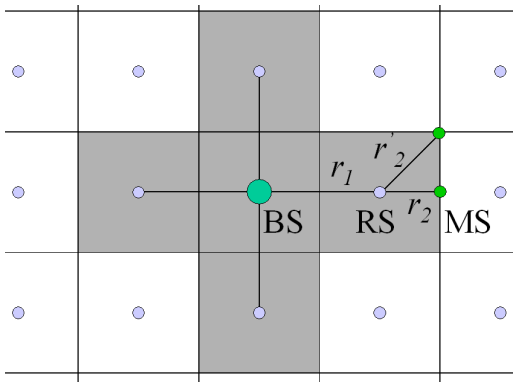
Appendix C

Two Dimensional Relaying Topologies

The one dimensional systems shown in Appendix B can be easily extended to the hexagonal and Manhattan cellular layouts shown in the following figures. Relays are deployed in the centres of the subcells, and distances, r_k, are determined by the specific geometry used (details given in the main body of thesis). Cluster sizes of $N = 1, 2, 3, 4$ were again studied, and interfering cells were considered in the simulations, but these details are not shown in the figures for simplicity.

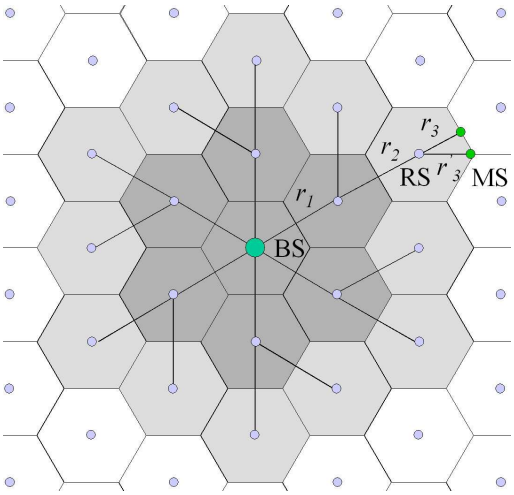


(a) Hexagonal topology: 7 shaded subcells comprise one cell.

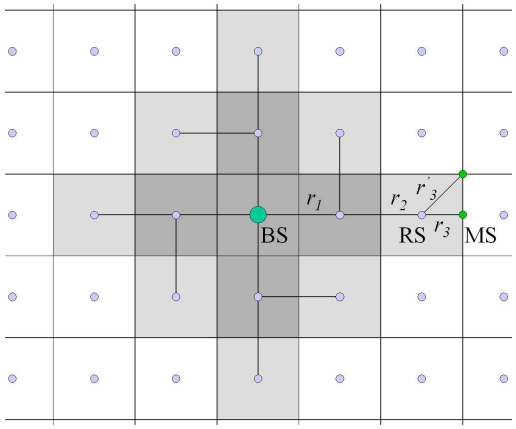


(b) Manhattan topology: 5 shaded subcells comprise one cell.

Figure C.1: Two hop relay cellular topologies.

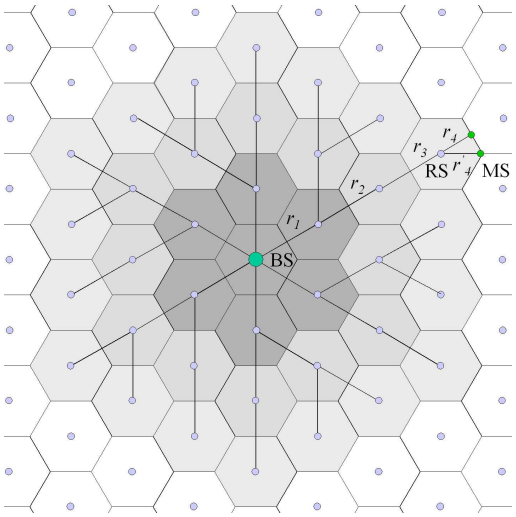


(a) Hexagonal topology: 19 shaded subcells comprise one cell.

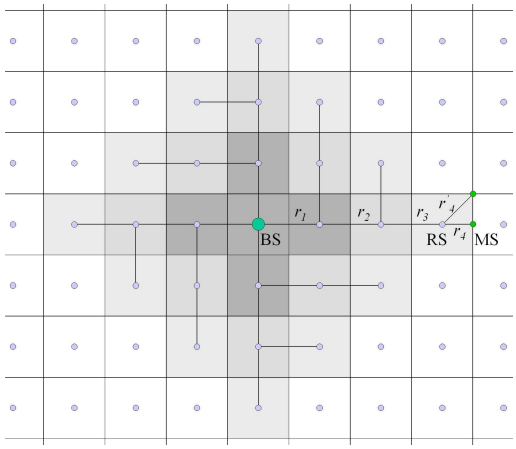


(b) Manhattan topology: 13 shaded subcells comprise one cell.

Figure C.2: Three hop relay cellular topologies.

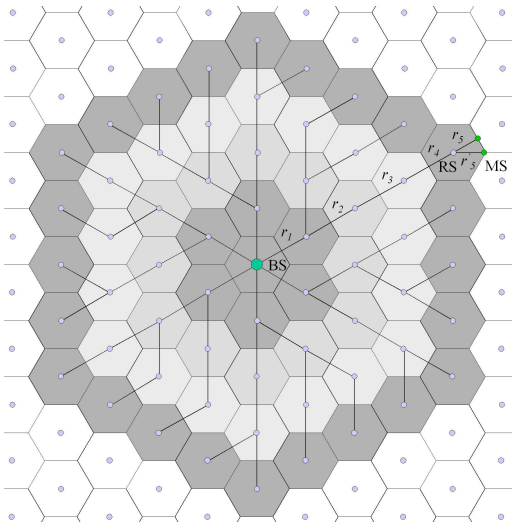


(a) Hexagonal topology: 37 shaded subcells comprise one cell.

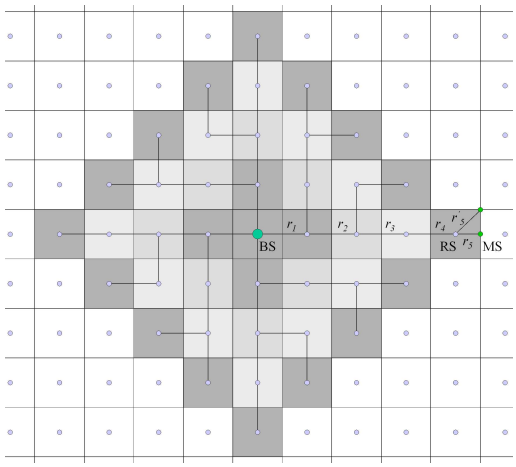


(b) Manhattan topology: 25 shaded subcells comprise one cell.

Figure C.3: Four hop relay cellular topologies.



(a) Hexagonal topology: 61 shaded subcells comprise one cell.



(b) Manhattan topology: 41 shaded subcells comprise one cell.

Figure C.4: Five hop relay cellular topologies.