

DIVERSITY AND SPECTRAL EFFICIENCY
IN WIRELESS RELAY NETWORKS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Deqiang Chen, B.S., M.S.

J. Nicholas Laneman, Director

Department of Electrical Engineering

Notre Dame, Indiana

December 2007

DIVERSITY AND SPECTRAL EFFICIENCY
IN WIRELESS RELAY NETWORKS

Abstract

by

Deqiang Chen

Motivated by the increasing demand for higher data rate, broader coverage, and lower infrastructure cost in wireless systems, a major effort is being made to study the use of relay stations in wireless networks. In this dissertation, we propose efficient relaying protocols for two types of wireless networks, *i.e.*, networks with one destination, one relay and multiple sources, and networks with one destination, one source and multiple relays. We model the former as a multi-access relay channel (MARC) and refer the latter as a multihop network.

For a MARC, we propose the multi-access relay amplify-forward (MAF) protocol. The proposed protocol allows for a low-complexity relay, allows the users to operate as if in a normal (non-cooperative) multi-access channel (MAC), and allows multiple users to achieve significant gains from *sharing* a single amplify-forward relay in slow fading environments. MAF achieves the optimal diversity-multiplexing trade-off at high multiplexing gains. Analysis of the protocol reveals that it outperforms the compress-forward strategy at low multiplexing gains and further outperforms the dynamic decode-forward protocol at high multiplexing gains.

We also propose different routing protocols for wireless multi-hop networks. An information theoretic analysis suggests that there is an optimum number of hops in maximizing the end-to-end spectral efficiency for a multihop network with

AWGN channels. Motivated by this observation, this dissertation takes the trade-off between power and bandwidth efficiency into account and proposes two routing schemes, namely, approximately-ideal-path routing (AIPR) and distributed spectrum-efficient routing (DSER). AIPR finds a path to approximate an optimum regular path and requires location information. DSER is more amenable to distributed implementations based on Bellman-Ford or Dijkstra's algorithms. We show by simulations that the spectral efficiencies of AIPR and DSER for random networks approach that of nearest-neighbor routing in the low signal-to-noise ratio (SNR) regime and that of single-hop routing in the high SNR regime. In the moderate SNR regime, DSER offers significant gains compared with nearest-neighbor or single-hop routing.

We further demonstrate the trade-off between the number of hops and spectral efficiency in a broadband multipath fading system by considering resource allocation in an orthogonal frequency division multiplexing (OFDM) multihop network. We formulate a power and subcarrier allocation problem for a network with one destination and multiple sources and relays and discuss the properties of the optimum solution for a convex relaxation of the problem. We then focus on low-complexity, efficient algorithms for a multihop network with one destination, one source and multiple relays. We develop an algorithm that is optimum at high SNR for a two-hop network. For a general multihop network, we propose a greedy approach to subcarrier allocation. We show by simulations that judicious allocation of power and subcarriers offers significant gains compared to a fixed assignment of subcarriers in a multi-hop network. Furthermore, our results demonstrate that, even for a broadband system with frequency selective fading, there exists an optimum number of hops that maximizes the end-to-end spectral efficiency.

To Li Xie and my parents.

CONTENTS

FIGURES	vi
ACKNOWLEDGEMENTS	ix
CHAPTER 1: INTRODUCTION	1
1.1 Wireless Relay Networks	1
1.2 Challenges	3
1.2.1 Trade-off between Throughput and Reliability	3
1.2.2 Bandwidth and Power Efficiency	5
1.2.3 Routing	6
1.2.4 Compatibility	6
1.3 Dissertation Scope and Outline	7
CHAPTER 2: BACKGROUND	9
2.1 Channel Model and Performance Metrics	10
2.1.1 Wireless Channels	10
2.1.2 Performance Metrics	15
2.2 The Relay Channel and Its Extensions	17
2.2.1 Relay channel	18
2.2.2 User Cooperation	20
2.2.3 MARC	21
2.2.4 Large Scale Wireless Ad-Hoc Networks	23
2.3 Routing	25
2.3.1 Routing Protocols in Wireless Networks	26
2.3.2 Routing Metrics in Wireless Networks	29
2.3.3 Algebraic Theories for Routing	30
2.3.4 Information Theory and Routing	31
2.4 Optimization	32
2.4.1 Complexity Theory	33
2.4.2 Convex Optimization	33
CHAPTER 3: RELAYING IN THE BLOCK-FADING MULTI-ACCESS CHANNEL	37
3.1 Background	38

3.1.1	Channel	38
3.1.2	Analysis Method	39
3.2	Multi-Access Amplify-Forward	40
3.3	DMT of MAF	41
3.3.1	Symmetric Two-User MARC	42
3.3.2	Symmetric N -User MARC, $N > 2$	45
3.4	Numerical Simulation	50
3.4.1	Symmetric Two-User MARC	51
3.4.2	Symmetric N -User MARC, $N > 2$	53
3.5	Conclusion	57
CHAPTER 4: SPECTRUM-EFFICIENT ROUTING IN MULTIHOP NETWORKS		59
4.1	System Model	60
4.2	Optimal Routing	61
4.3	Approximately Ideal Path Routing (AIPR)	63
4.4	Distributed Spectrum-Efficient Routing (DSER)	65
4.4.1	Values of the Routing Coefficient	66
4.4.2	Properties	67
4.5	Extensions	69
4.5.1	Spatial Reuse	69
4.5.2	Relation of DSER to Other Protocols	70
4.5.3	Finite Input Alphabet	72
4.6	Simulation Results	76
4.6.1	Gaussian Inputs	77
4.6.2	Finite Input Alphabet	87
4.7	Summary	93
CHAPTER 5: RESOURCE ALLOCATION FOR BROADBAND MULTI-HOP NETWORKS		94
5.1	Background	95
5.2	System Model	96
5.3	Problem Formulation	97
5.4	KKT Conditions After the Relaxation	99
5.5	Low-Complexity Algorithms	103
5.5.1	High-SNR Approximation Algorithm for Two-Hop Network	103
5.5.2	Greedy Algorithm	105
5.6	Simulation Results	106
5.7	Conclusion	111
CHAPTER 6: CONCLUSIONS		113
6.1	Contributions	113
6.2	Future Research	115

APPENDIX A: PROOF OF THEOREM 1	119
A.1 Preliminary	119
A.1.1 Notation	119
A.1.2 Exponential Equality	120
A.1.3 Codebooks	121
A.2 Lower Bound	121
A.2.1 Bounding Single User Error Event	122
A.2.2 Bounding Two User Error Event	123
A.3 Upper Bound	127
A.3.1 Bounding Single User Outage	128
A.3.2 Bounding Two User Outage	128
 APPENDIX B: PROOF OF THEOREM 2	 131
B.1 Proof of Lower Bound	131
B.2 Proof of Upper Bound	133
 APPENDIX C: PROOF OF PROPOSITION 1	 135
 BIBLIOGRAPHY	 136

FIGURES

1.1	A typical infrastructure network architecture. BS and MS correspond to base station and mobile station, respectively.	2
1.2	Example wireless relay networks. MARC stands for multi-access relay channel. BS, MS and RS correspond to base station, mobile station, and relay station, respectively.	4
2.1	Relay Channel	18
2.2	Cooperative diversity	21
2.3	Multi-access relay channel (MARC)	22
3.1	The Multi-access relay channel (MARC) with two users, multiplicative fading, and additive noise.	39
3.2	DMT of the different protocols for the two-user symmetric MARC.	46
3.3	Bounds on the DMT of 3-user MARC	49
3.4	Bounds on the DMT of 4-user MARC	50
3.5	Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the two-user MARC. The multiplexing gain is $r = 0.4$. Note that $R = r \log(1 + \rho)$	52
3.6	Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the two-user MARC. The multiplexing gain is $r = 0.8$. Note that $R = r \log(1 + \rho)$	53
3.7	Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the 3-user MARC. The multiplexing gain is $r = 0.4$. Note that $R = r \log(1 + \rho)$	54
3.8	Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the 3-user MARC. The multiplexing gain is $r = 0.8$. Note that $R = r \log(1 + \rho)$	55
3.9	Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the 4-user MARC. The multiplexing gain is $r = 0.4$. Note that $R = r \log(1 + \rho)$	56
3.10	Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the 4-user MARC. The multiplexing gain is $r = 0.8$. Note that $R = r \log(1 + \rho)$	57
4.1	Illustration of the first step in AIPR.	64

4.2	Spectral efficiency of a single-input single-output AWGN channel with finite size alphabets	74
4.3	Spectral efficiency of linear regular multihop networks with Gaussian input	75
4.4	Spectral efficiency of linear regular multihop networks with BPSK input	76
4.5	Average spectral efficiencies of different routing schemes for uniformly random linear networks with 5 nodes and $\alpha = 4$	81
4.6	Average spectral efficiencies of different routing schemes for uniformly random linear networks with 9 nodes and $\alpha = 4$	82
4.7	Average spectral efficiencies of different routing schemes for 2-D random networks with 9 nodes, $\alpha = 4$ and $\phi = \pi/2$	83
4.8	Sample DSER paths in a linear network.	84
4.9	Average spectral efficiencies of the optimal routing with bandwidth optimization (ORBO) and DSER for uniformly random linear networks with 5 nodes and $\alpha = 4$	85
4.10	Average spectral efficiencies as a function of path-loss exponent for uniformly random linear networks with 9 nodes.	86
4.11	Average spectral efficiencies versus network SNR for uniformly random linear networks with 9 nodes and $\alpha = 4$. The dashed lines correspond to TDMA without spatial reuse and the solid lines correspond to modulo- K scheduling.	87
4.12	Average spectral efficiencies of different routing schemes for uniformly random linear networks with BPSK, 5 nodes and $\alpha = 4$	89
4.13	Average spectral efficiencies of different routing schemes for uniformly random linear networks with BPSK, 9 nodes and $\alpha = 4$	90
4.14	Average spectral efficiencies of different routing schemes for uniformly random linear networks with 8-PSK, 5 nodes and $\alpha = 4$	91
4.15	Average spectral efficiencies of different routing schemes for uniformly random linear networks with 8-PSK, 9 nodes and $\alpha = 4$	92
5.1	Average rate versus SNR for a two-hop wireless network with 8 subcarriers.	108
5.2	Average rate versus SNR for a three-hop wireless network with 8 subcarriers.	109
5.3	Average rate versus SNR for a three-hop wireless network with different number of subcarriers. The greedy algorithm is used.	110

5.4	Average rate versus SNR of networks with different number of hops, 16 subcarriers. The greedy algorithm is used.	111
A.1	The shadowed regions are feasible regions for $(v_{\Theta}, v_{r,d})$ on the plane $v_{\Omega} = C \geq 1 - r$ when $0 \leq \min\{v_{1,r}, v_{2,r}\} \leq 1 - r$ for $0 \leq r \leq 1$	126
A.2	The shadowed regions are feasible regions for $(v_{\Theta}, v_{r,d})$ on the plane $v_{\Omega} = C \geq 1 - r$ when $\min\{v_{1,r}, v_{2,r}\} \geq 1 - r$ for $0 \leq r \leq 1$	126

ACKNOWLEDGEMENTS

I am grateful to my advisor Dr. J. Nicholas Laneman. I have benefited tremendously from countless interactions and discussions with him. His emphasis on the fundamental nature of techniques and problems has not only inspired many aspects of this dissertation, but also instilled me with a better understanding of research. His precise and thorough approach to research has set up an example that I wish to follow in my future career. He also deserves many, many thanks for carefully reading my manuscripts and patiently improving my English writing.

I wish to thank Prof. Costello and Fuja for their willingness to serve in my committee and for a critical reading of the dissertation and many suggestions. I also want to thank Prof. Haengi for many useful interactions and for contributing his broad perspectives to this dissertation.

Thanks go to all present and former members of the NCIP group: Kambiz Azarian, Shivaprasad Kotagiri, Wenyi Zhang, Brian Dunn and Michael Dickens for the time that we spent both at work and after work. Especially I want to thank Kambiz and Shiva for intensive technical discussions on various parts of the dissertation. Thanks also go to all my friends both at Notre Dame and elsewhere who made the last five years full of fun.

I wish to thank my wife, Li Xie, and my parents for their patience and constant support. My wife's understanding and encouragement has meant more than what I can express in a few words here. I am forever indebted to my parent for their support throughout these many years. I hope to have something better than this dissertation to dedicate to them in the future.

CHAPTER 1

INTRODUCTION

1.1 Wireless Relay Networks

Recent years have witnessed the explosive growth of personal wireless communication systems. The most notable examples are cellular voice networks, *e.g.*, IS-54, IS-95, cdma2000 and GSM [55], whose pervasive use impacts all walks of life. The use of wireless data networks, *e.g.*, IEEE 802.11 networks, is also quickly picking up speed. The basic architecture of these wireless networks can be seen in Fig. 1.1. These networks consist of base stations or access points communicating with end users, and a wired backbone, typically a telephone or data network, that provides a fast connection among base stations and a connection to the Internet. These networks have been intensively studied from different perspectives, *e.g.*, the capacity of the multi-access channel (MAC) [22] and routing algorithms for the Internet [66]. Hence, trade-offs involved in designing this type of networks are relatively well understood.

Exploiting the rapid progress in integrated circuits, signal processing, and networking, many applications of wireless communications are also emerging that suggest different architectures from those described above. The standard task group developing IEEE 802.16j Mobile Multihop Relay (MMR) suggests extending the coverage of a base station by deploying several relay stations around the base station [1]. In contrast to a regular cellular network, the connection between relay

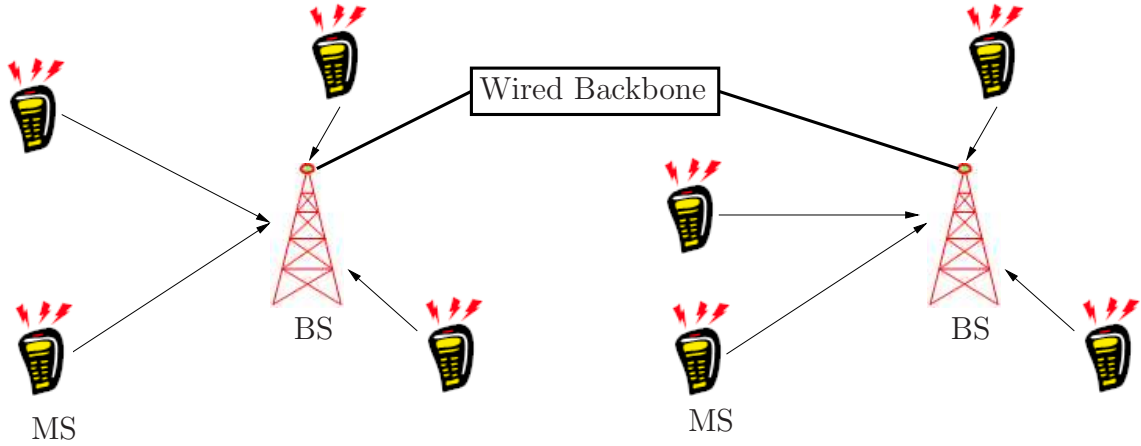


Figure 1.1. A typical infrastructure network architecture. BS and MS correspond to base station and mobile station, respectively.

stations and the base station is through wireless communication, instead of wired communication. Wireless Mesh Networks (WMN) [19, 26, 47, 13], an attractive alternative solution for broadband wireless access, also extend the coverage to users that are not in the direct transmission range of access points by allowing peer nodes to relay information for each other. Finally, deploying sensor networks in environments lacking infrastructure support often requires that sensors communicate with each other through the wireless medium.

These emerging networks differ from traditional wireless networks by utilizing wireless transmission to relay information, hence the name *wireless relay networks*. More specifically, a station is called *a relay station* if it retransmits signals received from some stations to other stations. A *wireless* relay station is a relay station that receives and retransmits through the wireless medium. A wireless relay network is a wireless network in which there exists at least one wireless relay station. Fig. 1.2 demonstrates some potential configurations of wireless relay networks that are particularly related to this dissertation, including examples such as user cooperation or cooperative diversity, multi-access relay channel, and multihop transmission. In

cooperative diversity, peer users cooperate to relay information for each other. In a multi-access relay channel, multiple sources share a designated relay station to communicate with a base station. In multihop transmission, a message is transmitted to the base station through a sequence of wireless relay stations.

Extending wireless communication beyond last hop via wireless relaying can potentially help extend the coverage, reduce the infrastructure costs and expedite deployment time. However, the addition of wireless relay stations also takes the design challenge of wireless networks to a new level as we will see in the next section.

1.2 Challenges

Wireless relay networks distinguish themselves from the existing infrastructure networks in that wireless transmission is no longer limited between the base station and users, but is utilized to relay information among users and relays as well. Even though the extension of wireless transmission from the last hop to intermediate hops brings many unique advantages, it also presents many fundamental challenges. This section discusses at a high level some of these challenging issues, including trade-off between throughput and reliability, bandwidth versus power efficiency, efficient routing, and compatibility with legacy systems. Many of these issues are intertwined. The discussion here does not intend to cover all aspects of wireless relay networks, but rather those issues that are particularly related to this dissertation.

1.2.1 Trade-off between Throughput and Reliability

Wireless networks are designed to transmit as much information as possible given certain constraints. For sensor networks, efficiency of transmission also translates into longer lifetime of the networks. For mesh networks, higher throughput could mean the efficient use of limited bandwidth resources. However, wireless networks, in particular data-service oriented networks, also require high reliability. Unfortu-

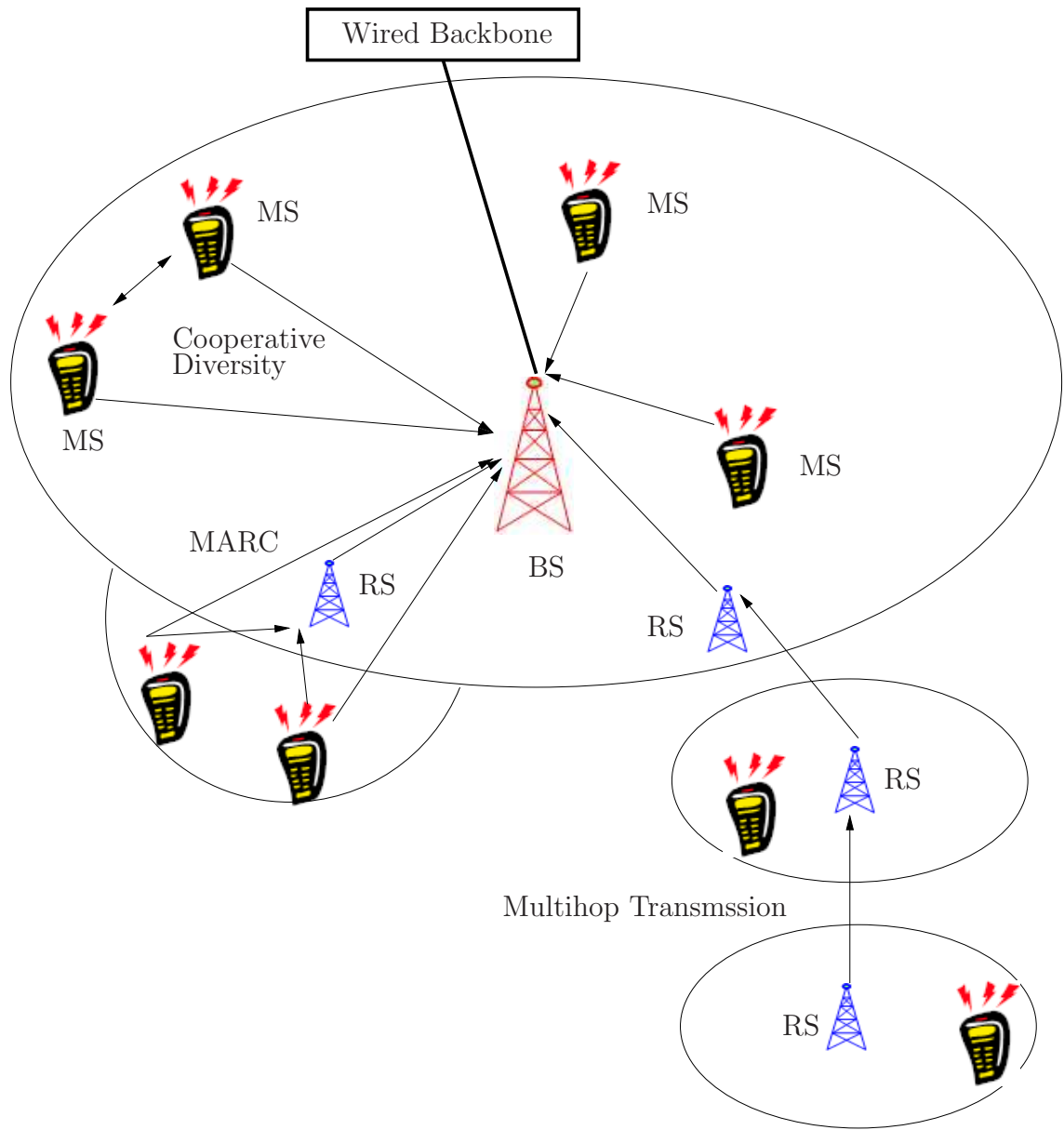


Figure 1.2. Example wireless relay networks. MARC stands for multi-access relay channel. BS, MS and RS correspond to base station, mobile station, and relay station, respectively.

nately, wireless transmissions are corrupted by path-loss, shadowing, fading, interference and noise; typically, the higher the transmission rate, the more resources are required, and the higher the probability of error.

In wireless relay networks, the trade-off between throughput and reliability is particularly intriguing. In traditional wireless networks, only the last hop is wireless, and the wired backbone network can be regarded as essentially error-free. Therefore, it is sufficient to study the trade-off of the last hop. For wireless relay networks, the probability of error in one of the intermediate hops is not negligible and could impact the design of the whole network.

Chapter 3 discusses the trade-off between throughput and reliability for a multi-access relay channel (MARC) using the tool of the diversity-multiplexing trade-off (DMT). The trade-off between system throughput and per-link reliability also motivates the routing protocols that we propose in Chapters 4 and 5, though less directly than in Chapter 3.

1.2.2 Bandwidth and Power Efficiency

Unlike a copper wire or optical fiber connecting the transmitter and receiver in wired communication, the medium for wireless communication is inherently shared. Regulatory agencies, *e.g.*, the Federal Communications Commission (FCC), have established strict rules regarding the use of frequency bands to avoid excessive interference. Wireless service providers, in particular, cellular companies, sometimes pay billions of dollars for the license of tens of megahertz. Thus, transmitting as much information as possible in a limited frequency band, *i.e.*, operating at high bandwidth efficiency, is a top priority.

The FCC also has strict regulations with regard to the maximum transmission power allowable in each frequency band. This translates into an average power

constraint for the wireless devices. Moreover, due to mobility requirements, wireless devices often rely on batteries as their power supply. Therefore, increasing the power efficiency means longer life time for the wireless devices.

Unfortunately, the aim to boost the bandwidth efficiency and power efficiency often results in conflicting requirements. The trade-off between the bandwidth efficiency and power efficiency for point-to-point channels has been widely studied since the seminal work of Shannon [61]. However, trade-offs between the bandwidth and power efficiency for most multiuser networks are still open problems [22], particularly for wireless relay networks.

The trade-off between the bandwidth and power efficiency in multihop networks motivates the routing protocols in Chapter 4 and power and subcarrier allocation algorithms in Chapter 5.

1.2.3 Routing

Due to the dynamic nature of wireless network topology and volatile wireless channels, routing information to the desired destination in wireless networks is another challenging task. To further complicate the situation, routing schemes have a significant impact on the performance of wireless relay networks, as we will see in Chapter 2. Thus, routing schemes must be carefully designed to fully exploit wireless resources, *e.g.*, power and bandwidth. Chapters 4 and 5 focus on the impact of power and bandwidth efficiency on selection of multihop routes.

1.2.4 Compatibility

Currently, there are millions of wireless service users. If the migration from the current wireless system to the next generation of wireless relay networks requires replacement of all user devices, cost could become a huge obstacle for the adoption of the new technology. Therefore, the development of wireless relay networks should

consider how legacy devices can be integrated into new systems. Throughout this dissertation, we design relaying protocols that are, at least in principle, compatible with existing operations of current wireless networks.

1.3 Dissertation Scope and Outline

This dissertation discusses the impact of relays on different layers, demonstrates different trade-offs unique to wireless relay networks, and suggests significant gains provided by efficient relaying protocols. In particular, we propose relaying or routing protocols that improve performance of wireless networks by judiciously utilizing relay stations. In contrast with previous work that focused on proposing and studying complicated relaying protocols, we focus on proposing simple, yet efficient, relaying protocols. The design of these protocols must consider different trade-off in wireless relay networks, *e.g.*, power efficiency versus bandwidth efficiency, reliability versus throughput, and complexity versus performance.

This dissertation focuses on two type of architectures, namely, the multi-access relay channel (MARC) and the multihop network. The MARC captures scenarios in which only one relay is available for multiple users, *e.g.*, extending the coverage of cellular networks or wireless local area network (WLAN) with a single relay station. By contrast, the multihop network models scenarios in which multiple relay stations are available, *e.g.*, a wireless backbone network for wireless service providers for an area in which the wired backbone network does not justify the cost. Although these two architectures do not capture all possible application scenarios of wireless relay networks, we are able to get some valuable insights about unique difficulties and trade-offs facing design of wireless relay networks from the study of these two particular cases.

The rest of the dissertation is outlined as follows. Chapter 2 surveys develop-

ment in the study of wireless relay networks and reviews some key results in detail. Emphasizing the importance of graceful transition from traditional cellular networks to wireless relay networks, Chapter 3 demonstrates the significant gains that multi-access users can achieve from sharing a *single* amplify-forward relay in slow fading environments. Chapter 3 focuses on the trade-off between reliability and throughput through the study of the DMT. In contrast with previous works, our proposed protocol in Chapter 3 enjoys low complexity at the relay station. Surprisingly, our proposed protocol is DMT-optimal for a multiplexing regime in which several more complicated protocols are not DMT-optimal. On the other hand, focusing on the scenario of *multiple* relays, Chapter 4 combines different perspectives from networking and information theory in the design of routing schemes that can be implemented via distributed algorithms. Chapter 4 proposes two different routing protocols, namely, approximately ideal path routing (AIPR) and distributed spectrum efficient routing (DSER). Through simulation results, Chapter 4 demonstrates the gains in spectral efficiency offered by our proposed protocols relative to direct communication and nearest neighbor routing. As both Chapter 3 and Chapter 4 focus on *narrowband* systems, Chapter 5 considers orthogonal frequency division multiplexing (OFDM) systems and studies the problem of power and subcarrier allocation in a multihop wireless network. Furthermore, Chapter 5 shows that, similar to the narrowband case, increasing the number of hops does not always improve the spectral efficiency of a *broadband* multihop network. Chapter 6 concludes the dissertation by summarizing the contributions and discussing potential directions for future work.

CHAPTER 2

BACKGROUND

This chapter reviews important results from a broad array of related references and discusses many considerations involved in formulating the problems studied by the dissertation. The motivation of this chapter is to familiarize the readers with the field of study and lay the foundation for the rest of the dissertation.

The rest of the chapter is organized as follows. Section 2.1 first reviews basic wireless channel models and performance metrics. Section 2.2 summarizes important results on wireless relay networks, *e.g.*, the relay channel, from an information theoretic perspective. Although traditional information theoretic analyses offer valuable insights about fundamental trade-offs, *e.g.*, the trade-off between bandwidth and power efficiency in the relay channel, parameters of a large scale network form a gigantic design space and precise analysis proves to be extremely difficult. As a result, the focus on analysis for large scale wireless ad-hoc networks is shifted from precise characterization of the capacity region to the scaling behavior of rates with respect to network size. Section 2.2 also reviews related results from the study of scaling behavior on large scale networks with an emphasis on their implications for routing. Routing has traditionally been studied within the networking community, but it can be viewed as a practical alternative to more complicated relaying protocols. Section 2.3 briefly reviews some important routing protocols and metrics from the networking community. Finally, Section 2.4 provides necessary background for

optimization techniques that are utilized.

2.1 Channel Model and Performance Metrics

2.1.1 Wireless Channels

The performance of wireless communication systems is fundamentally limited by the wireless propagation channels. Even though wireless communication can occur across a wide range of frequencies, and radio channels exhibit different features depending on frequency, we focus on the complex baseband model and consider wireless signal impairments caused by three nearly-independent effects [55]: noise, path-loss, and fading. The first component is modeled as additive, and the rest are modeled as multiplicative. As an illustrative example, a general discrete-time point-to-point baseband-equivalent model that captures these three effects is given by

$$y[n] = \sqrt{G} \sum_{m=0}^{M-1} h_m[n] x[n-m] + z[n], \quad (2.1)$$

where:

- n is the discrete time index
- m is the tap index of the linear filter that models fading effect and M is the maximum number of taps or the maximum delay spread
- $x[n]$ and $y[n]$ represent the channel input and output, respectively.
- $z[n]$ represents the additive white Gaussian noise (AWGN). Typically, it is modeled as a white zero-mean circularly symmetric complex Gaussian random process with variance N_0 . The additive noise mainly arises from thermal noise and other interference in receiver circuits
- G is the path-loss factor between the transmitter and receiver and captures the loss of signal power due to free-space propagation. In general, G is not a ran-

dom variable and can be determined by the distance between the transmitter and receiver, antenna gains, antenna height, *etc*

- $h_m[n]$, often modeled by a random process, captures the effects of reflections, scattering and diffraction through multipath propagation and results in SNR fluctuation in time and frequency domain called fading. We provide a more detailed description about the classification and stochastic characterization of fading in the sequel.

Path-Loss

For different environments, different models can be used to predict the loss given the transmit distance [55]. In this dissertation, we use a standard path-loss model, *i.e.*, the path-loss factor from the transmitter to the receiver is given by

$$G = c[\max(D, D_f)]^{-\alpha}, \quad (2.2)$$

where D is the Euclidean distance between the transmitter and the receiver, D_f is the far-field distance [55], α is the path-loss exponent (typically taking values between 2 and 4), and c is a constant that can be determined by wavelengths, antenna gains, *etc*. For most practical scenarios, D is much larger than D_f ; thus, (2.2) can be approximated as

$$G \approx cD^{-\alpha}. \quad (2.3)$$

For simplicity of presentation, we mainly use (2.3) in this dissertation. Also, after appropriately normalizing the transmission power, which is sufficient for relative comparison, we will assume that $c = 1$ in (2.3).

In some scenarios, *e.g.*, free-space communications with strong line-of-sight components, fading can be neglected. For those scenarios, we can simplify (2.1) to an AWGN channel, *i.e.*,

$$y[n] = \sqrt{G}x[n] + z[n]. \quad (2.4)$$

Fading

In general, fading could arise due to a scatter-rich environment or the movement of transceivers. In a scatter-rich environment like offices, wireless signals could go through different paths, as seen in (2.1), before they arrive at the receiver and add up constructively or destructively. Multipath fading results in significant variations in the received signal power over a short distance. The moving of devices causes fading as well. Fading can be classified as slow or fast depending upon the coherence time and coding interval. If the coding interval is small compared to the channel coherence time, fading is regarded as slow fading. If the coding interval is large relative to the channel coherence time, fading is regarded as fast fading. Intuitively, for slow fading, the fluctuation of fading during one coding interval is small. In this dissertation, we adopt the block fading channel in [81] as a first-order approximation to the slow fading channel. In the block fading channel, the fading coefficients remain the same within one coding block, and change independently between blocks. Hence, during one coding block, the channel in (2.1) can be simplified to:

$$y[n] = \sqrt{G} \sum_{m=0}^{M-1} h_m x[n-m] + z[n]. \quad (2.5)$$

We adopt (2.5) as the fading channel model throughout this dissertation.

Depending on the signal bandwidth and channel coherence bandwidth, fading can be further classified as frequency-selective or frequency-nonselective. For frequency-selective fading, the signal bandwidth is large relative to the channel coherent bandwidth, hence, different frequency components of signals experience different fading. By contrast, for frequency-nonselective fading, different frequency components of signals experience the same fading at any given time.

To model frequency-nonselective slow fading channels, we can assume that the

maximum delay spread is unity and simplify (2.5) to

$$y[n] = \sqrt{G}h_0x[n] + z[n]. \quad (2.6)$$

For ease of exposition, we often combine $\sqrt{G}h_0$ as one element \tilde{h} and rewrite (2.6) as

$$y[n] = \tilde{h}x[n] + z[n]. \quad (2.7)$$

For $M > 1$, (2.5) models frequency-selective slow fading channels. Frequency-selective channels often cause severe inter-symbol interference (ISI) [54]. To simplify signal processing, orthogonal frequency division multiplexing (OFDM) can be employed to turn the channel in (2.5) to a set of parallel channels [46], *i.e.*,

$$\hat{y}_k[n] = \sqrt{G}\hat{h}_k\hat{x}_k[n] + \hat{z}_k[n], k = 1, \dots, K, \quad (2.8)$$

where K is the total number of subcarriers; \hat{x}_k , \hat{y}_k , and \hat{z}_k denotes the channel input, output, and noise, respectively, at subcarrier k ; \hat{h}_k denotes the fading coefficient at subcarrier k . We note that \hat{h}_k of (2.8) and h_m of (2.5) are connected by [46]

$$\hat{h}_k = \sum_{m=0}^{M-1} h_m e^{-2\pi mk/K}. \quad (2.9)$$

OFDM has been adopted widely in broadband wireless networks, *e.g.*, IEEE 802.11 and 802.16. Interesting readers are referred to [46] for more detail description of OFDM. In this dissertation, we adopt (2.8) as the channel model when OFDM is employed.

Stochastic property of fading can be modeled using different probability distributions, *e.g.*, Rayleigh and Ricean distributions [55]. In this dissertation, we focus on the Rayleigh fading, modeled by assuming $h_m, i = 0, \dots, M - 1$ as independent circularly symmetric complex Gaussian random variables. However, for multipath fading, the strengths of different fading paths may be different, *i.e.*, the variances

of $h_m, i = 0, \dots, M - 1$ may not be the same. Power delay profile (PDP) characterizes the relative fading strength among different paths. If the variances of $h_m, i = 0, \dots, M - 1$ are the same, the PDP is said to be a uniform PDP. Another typical PDP is the exponential PDP [65], *i.e.*,

$$E[|h_m|^2] = E[|h_0|^2]e^{-\zeta m} \quad (2.10)$$

where $\zeta > 0$ is the power decay factor. Note that $E[|h_0|^2]$ can be determined by assuming $\sum_{m=0}^{M-1} E[|h_m|^2] = 1$.

More detailed descriptions about channel models are available in [55, 54, 52, 65].

Power Constraint and Signal-to-Noise Ratio

The FCC also has strict regulations with regard to the maximum transmission power allowable in each frequency band. This translates into an average power constraint in the wireless devices. In this dissertation, we focus on symbol-wise average transmit power P , *i.e.*,

$$E[|x[n]|^2] \leq P \quad (2.11)$$

Throughout this dissertation, we consider the setting in which all transmit devices are constrained by the same symbol-wise average transmit power P and assume all devices transmit with power P when transmitting. This assumption is justified by the fact that, for low-power transceivers, local oscillators and bias circuitry dominate energy consumption [34]. Moreover, the radio frequency (RF) power amplifier (PA) should mostly operate close to its saturated power for the most energy efficient operation, as this is when the power added efficiency (PAE) is largest [32]. We also assume that the variance of AWGN at all receivers are the same as N_0 . With a system bandwidth W , the signal-to-noise ratio (SNR) is then defined by $\rho := P/N_0W$.

2.1.2 Performance Metrics

Capacity

For an AWGN channel, the well-known Shannon capacity characterizes the trade-off between the power and bandwidth for a given data rate. More specifically, for a single-input single-output (SISO) AWGN channel in (2.4), the Shannon capacity is

$$C = \log(1 + G\rho). \quad (2.12)$$

Throughout this dissertation, \log represents the logarithm with base 2 unless otherwise specified. Therefore, the units of (2.12) are bits per complex channel uses.

For a fading channel, the notion of Shannon capacity applies as a useful performance measure if the fading process is stationary and ergodic and coding is performed over long enough intervals. When the fading process is available to the receiver, but not the transmitter, the Shannon (or ergodic) capacity of a fading channel in (2.6) is

$$C = E[\log(1 + G|h_0|^2\rho)], \quad (2.13)$$

where the expectation is with respect to the fading distribution.

Outage

If the fading process is non-ergodic or the system's delay constraint prevents the fading process from revealing its ergodic structure within the coding interval, we cannot guarantee reliable communication of any fixed, non-zero rate *a priori* since the fading coefficient might be so small that the realized SNR is insufficient to support that rate. Hence, the Shannon capacity is often zero and therefore not a useful performance measure. For such scenarios, the capacity-vs.-outage and outage probability are more relevant performance metrics.

Given a fixed rate R , we refer to the event $\log(1+G|h|^2\rho) \leq R$ as an *outage* event. The probability of this event is referred to as the *outage probability* of the channel. Outage probability reflects the reliability of a communication system operating at a given rate. The capacity-versus-outage is defined to be the maximum rate with outage probability less than a given level. In general, the outage probability increases as the rate requirement increases and the capacity-versus-outage decreases as the required outage probability level decreases.

For more detailed discussions about capacity, we refer readers to [29, 22]. A comprehensive summary about capacity, outage-probability, and capacity-versus-outage in fading channels is available in [9].

Diversity-Multiplexing Trade-off

Based upon error probability, or outage probability, the diversity-multiplexing trade-off (DMT) characterizes the trade-off between reliability and rate in the high SNR regime. We motivate the concept of DMT through a multiple-input multiple-output (MIMO) channel with n_T transmit antennas and n_R receive antennas. For a fixed rate R , the error or outage probability of MIMO can be approximated as

$$P_e \approx \frac{c}{\rho^{n_T n_R}}, \quad \rho \gg 1,$$

where: the constants c and $n_T n_R$ are referred to as the *coding gain* and *diversity gain*, respectively. If we plot this probability as a function of SNR on a log-log scale, the diversity gain $n_T n_R$ represents the high-SNR slope of the error probability curve, revealing how fast the error probability decays as SNR increases. As the coding gain in general does not increase as SNR grows, the diversity gain is a particularly useful measure of reliability in the high SNR regime. However, the diversity gain $n_T n_R$ is achieved when the system operates with a *fixed* rate, regardless of SNR. In practice, communication systems typically operate at higher rates if SNR

increases. To characterize how the achievable diversity gain changes if the operating rate changes with SNR, Zheng & Tse proposes the DMT [81].

The DMT provides the largest diversity gain that can be achieved if the system operates at a fixed fraction of the corresponding AWGN capacity, which increases as the SNR increases. Specifically, let $R(\rho) = r \log \rho$ and denote $P_e(\rho, R(\rho))$ as the error probability of the ML decoder at the receiver, the diversity gain

$$d := \lim_{\rho \rightarrow \infty} -\frac{\log P_e(\rho, R(\rho))}{\log \rho}$$

is a function of the multiplexing gain r . A protocol's DMT provides the diversity gain d as a function of multiplexing gain r , *i.e.*, $d(r)$. We note that the multiplexing gain is analogous to the spectral efficiency normalized by the channel capacity [81, 43].

Intuitively, the outage probability of a protocol decays at a slope of $d(0)$ on a log-log scale for a given rate, *i.e.*, $r = 0$. On the other hand, if the outage probability is fixed, for every 3 dB gain in SNR, the rate can be increased by the maximal multiplexing gain. This intuitive interpretation is not entirely accurate due to the approximation of high SNR [6]. However, given two protocols operating at the same SNR and multiplexing gain, the DMT can be a useful tool in qualitatively comparing their performance.

2.2 The Relay Channel and Its Extensions

Depending on the network architecture, wireless relay networks can be modeled in different ways, *e.g.*, the relay channel, the multi-access relay channel (MARC), user cooperation, and ad-hoc networks. Section 2.2.1 and Section 2.2.2 review some results on the relay channel and user cooperation respectively. Section 2.2.3 provides a brief summary of related results on MARC. Focusing on the scaling behavior of large scale wireless ad-hoc networks, Section 2.2.4 outlines several key results that highlight the importance of efficient routing in wireless relay networks.

2.2.1 Relay channel

Fig. 2.1 shows a block diagram of relay channel. In this channel, the source transmits its messages to the destination with the help of a relay station. In general, the relay does not have its own information to send. The relay channel can be viewed as a special case of the more general framework of channels with “generalized” feedback [70, 71].

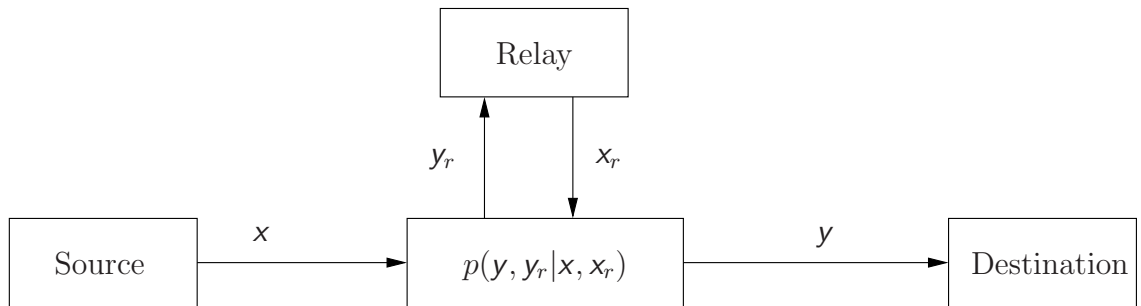


Figure 2.1. Relay Channel

Cover and El Gamal [21] study the relay channel and provide lower and upper bounds. More specifically, [21] proposes three different coding strategies: decode-forward (DF), compress-forward (CF), and superposition of DF and CF. Both DF and CF coding strategies are based on the principle of block-Markov coding. Essentially, the encoding process is to superimpose a new encoded message with the bin index of the previous message in a block-Markov way. A summary of different block-Markov coding and decoding process is available in [41].

The DF coding strategy in [21] achieves the capacity of the degraded relay channel, *i.e.*, if $p(y, y_r | x_r, x) = p(y_r | x_r, x)p(y | y_r, x_r)$ in Fig. 2.1. Furthermore, when the relay channel is not degraded, [41] shows that a DF strategy achieves the capacity for AWGN channels with phase fading if the source-relay SNR is sufficiently strong.

Unfortunately, the capacity-achieving strategy is generally unknown for the general relay channel.

In certain circumstances, as we will see in Chapter 3, instead of decoding the whole message at the relay, it is better for the relay to just quantize or estimate its received signal, or decode only part of the messages. These are the ideas behind the CF strategy [21] and the partial decode-forward (PDF) strategy [2], respectively. In the CF strategy, the relay uses Wyner-Ziv source coding [74] to exploit side information at the destination. In the PDF strategy, the relay only decodes part of the messages.

Most of the above discussions on the relay channel are under the assumption of full-duplex relay, *i.e.*, the relay can transmit and receive at the same time at the same frequency. In traditional wired networks, devices capable of listening and transmitting at the same time at the same frequency can be easily built. However, wireless transmission is broadcast in nature, and a node that is transmitting causes its own near-far problem if it tries to listen to the same frequency on which it transmits. To avoid this effect, we can consider wireless devices being half-duplex, *i.e.*, not transmitting and receiving at the same time on the same frequency. The information theoretic theorems on capacity of full-duplex relay channel can be extended to a relay channel with half-duplex constraint using Kramer's framework [40]. In [40], the relay state, *i.e.*, whether the relay is listening, transmitting or idle, is represented as a random variable. By considering the relay state as part of the channel input, the achievable rates for various relaying schemes can be derived based on the results from the full-duplex relay channel.

The DMT of the half-duplex relay channel has been studied [43, 5]. In [5], a dynamic decode-forward (DDF) strategy is proposed. In DDF, the relay does not transmit until it collects enough information to decode and the source keeps

on transmitting all the time. It is shown that DDF offers the best performance comparing with other schemes in terms of DMT [5]. Moreover, DDF achieves the cut-set upper bound in the low multiplexing gain regime. A non-orthogonal amplify-forward (NAF) strategy [5] is also developed. The NAF is similar with the orthogonal amplify-forward (AF) with the difference that the source in the NAF keeps on transmitting through the whole coding interval. Not surprisingly, NAF offers better performance than AF, though not as good as DDF.

2.2.2 User Cooperation

When there is no designated relay, wireless terminals possessing generalized feedback can cooperate to relay for each other to improve the overall performance as depicted in Fig. 2.2. This scheme has been referred to as “user cooperation” or “cooperative diversity” depending upon the context [43, 60]. Sendonaris *et. al* consider the capacity and bit error rate of cooperative diversity assuming each node with full-duplex capability [60]. In [44, 43], a new perspective on user cooperation based upon outage probability is studied assuming half-duplex terminals and quasi-static fading channels. A variety of algorithms, *e.g.*, amplify-forward and selection decode-forward, are proposed and shown to achieve full diversity order [44], motivating the term “cooperative diversity” for such approaches. However, those algorithms use the channel orthogonally and therefore cause the loss of degrees of freedom.

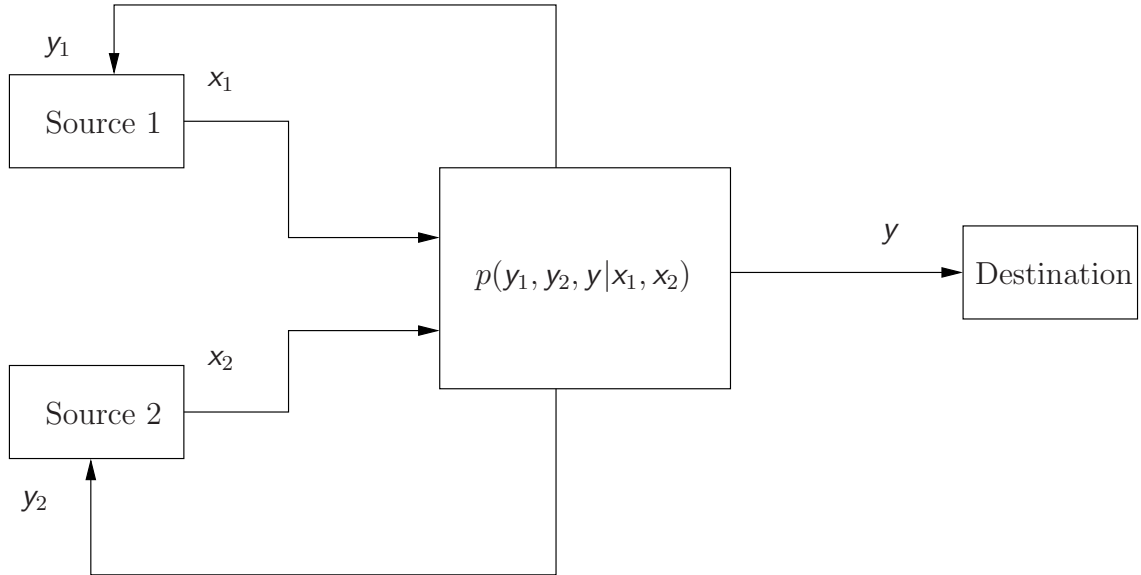


Figure 2.2. Cooperative diversity

In [5, 10], strategies based upon nonorthogonal channel use have been studied, *e.g.*, NAF and DDF. A surprising observation is that insights offered in the study of the relay channel cannot be carried over to user cooperation without some caution. This point can be illustrated from one example in [5]. Even though DDF is better for both AF and NAF in the half-duplex relay channel, when multiple users cooperatively transmit information to the central station, it turns out that a strategy based upon amplify-forward is optimal.

2.2.3 MARC

Kramer and Wijnngaarden [42] first proposed a multi-access relay channel (MARC) model in which sources share a common relay between themselves and the common destination. Fig. 2.3 shows a typical MARC. The MARC can be viewed as a combination of the multi-access channel (MAC) and relay channel.

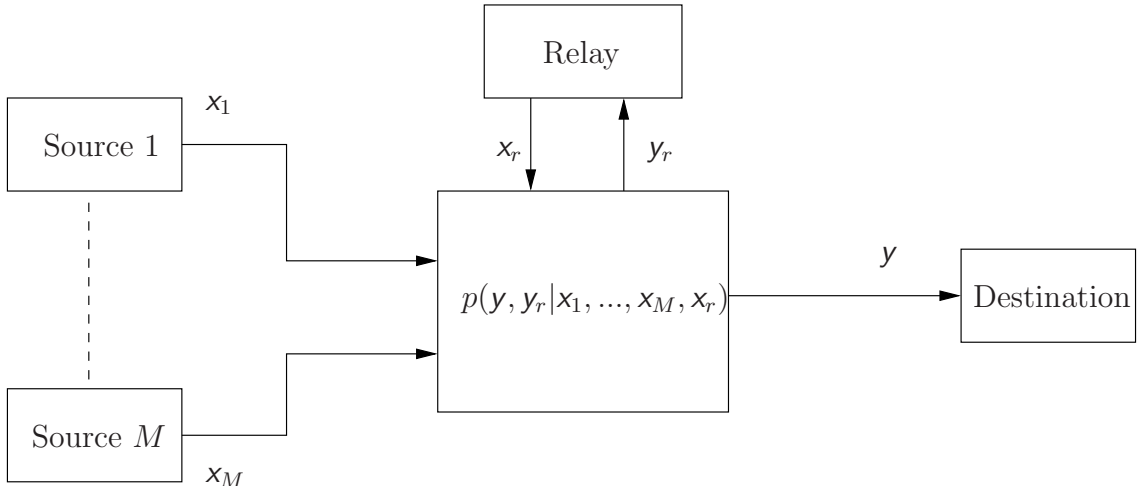


Figure 2.3. Multi-access relay channel (MARC)

Building upon [21], [42] provides upper and lower bounds on the capacity of AWGN MARC assuming a full-duplex relay. Subsequently, [57, 41, 59, 58] have also developed capacity bounds employing different coding strategies. Unfortunately, the capacity of the general MARC is still unknown.

From the DMT perspective, [7] studies the performance of DDF for MARC. It is shown that DDF achieves the optimal DMT in the low multiplexing regime for two-user MARC. However, the optimal DMT in the high multiplexing regime is not achieved by DDF. On the other hand, [79] shows that the CF strategy can be extended to the MARC and provides the optimal DMT in the high multiplexing regime. However, CF suffers from a significant performance loss in the low multiplexing regime. Part of the dissertation, *i.e.*, Chapter 3, develops strategies to achieve the optimal DMT in a regime where neither DDF nor CF is optimal. As we will argue in Chapter 3, our proposed strategy also enjoys the advantage of low complexity.

2.2.4 Large Scale Wireless Ad-Hoc Networks

It is extremely difficult to precisely characterize the capacity region of multi-terminal networks [22]. Therefore, research on large scale networks has focused on how the performance scales with the size. This section summarizes some key results in this field. Discussions of these results demonstrate the importance of efficient routing in wireless relay networks.

Knuth's notation [18] is used to describe the scaling behavior. For a given function $g(n)$, the tight bound in Knuth's notation is

$$\Theta(g(n)) = \{f(n) : \exists c_1, c_2, n_0, 0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n), \forall n \geq n_0\}. \quad (2.14)$$

We often abuse the notation to denote $f(n) \in \Theta(g(n))$ as $f(n) = \Theta(g(n))$, and we often use the notation $\Theta(1)$ to mean either a constant or a constant function with respect to some variables.

If we only have an asymptotic upper bound, we use O -notation, which is defined as

$$O(g(n)) = \{f(n) : \exists c, n_0, 0 \leq f(n) \leq c g(n), \forall n \geq n_0\}. \quad (2.15)$$

An asymptotic lower bound is expressed as Ω -notation, which is

$$\Omega(g(n)) = \{f(n) : \exists c, n_0, 0 \leq c g(n) \leq f(n), \forall n \geq n_0\}. \quad (2.16)$$

We consider a naive example of the application of Knuth's notation by assuming $f(n) = 10n^2$. By definition, it can be easily verified that $f(n) = \Theta(n^2)$, $f(n) = O(n^3)$ and $f(n) = \Omega(n)$.

The seminal work in [31] studies the scaling of network transport capacity, *i.e.*, the maximum summation of the product of information rate and its corresponding transmit distance, with respect to the number of nodes. Two channel models, namely, the Protocol model and the Physical model, are considered in [31]. The

Protocol model reflects scenarios where the medium access control policy ensures that there are no transmitting nodes around the scheduled receiver. The Physical model captures the fact that a minimum signal-to-interference ratio (SIR) is necessary for successful receptions. In both models, it is assumed that each node can transmit at most W bits per second. The scaling behavior of transport capacity also depends on the ratio between the number of nodes and the area size of the network, *i.e.*, the node density of the network. Depending on how the node density varies as the number of nodes changes, ad-hoc networks can be further classified into “dense” networks or “extended” networks. In a dense network, the node density increases as the number of nodes increases, *e.g.*, the number of nodes increases in an fixed area. By contrast, in an extended network, the node density remains constant; hence, the area of the network increases as the number of nodes increases.

For a dense wireless network consisting of n nodes in a disk of area A , [31] states that the total bit-distance product per second that can be supported by the network is $\Theta(W\sqrt{n})$, assuming that the traffic patterns and node locations are optimally assigned and each transmission’s range is optimally chosen. If nodes are randomly located, the throughput in the Protocol model is $\Theta(W/\sqrt{n \log n})$ [31]. The proof of this lower bound provides us an achievable scheme and suggests that multihop is order optimal. This work has been used as a justification for favoring multihop strategies. However, we have a more conservative interpretation of the results. Several important factors in communication and information theory are not considered here. As mentioned before, all nodes are assumed to transmit only at W . From an information theoretic perspective, this means all nodes are using fixed rate single user coding. More sophisticated coding and decoding might help and impact the behavior of the scaling law. Link failures are also neglected. For scenarios in which link failure is not negligible, the multi-hop strategy might require

a significant number of retransmissions. Finally, delay associated with multi-hop is not quantified.

Furthermore, for the Physical model in the above dense wireless network, the upper bound $O(W/\sqrt{n})$ and the lower bound $\Omega(W/\sqrt{n \log n})$ do not coincide. Similarly, for extended networks, we have an $O(1)$ upper bound and an $\Omega(1/\sqrt{\log n})$ lower bound [75]. The intriguing fact is that, there is always a $\sqrt{\log n}$ factor difference between the upper and lower bound when the Physical model is considered.

These gaps between the lower and upper bounds are later closed in [23] using a more structured routing scheme. More specifically, [23] provides a much stronger achievable bound for the random extended network. Assuming the path-loss model as $\min\{1, e^{-\gamma d_{ij}}/d_{ij}^\alpha\}$ with $\alpha > 0, \gamma > 0$ or $\alpha > 2, \gamma = 0$, a per-node throughput rate $\Omega(1/\sqrt{n})$ is achievable in a random extended network, with a corresponding average packet delay upper bounded by $E(D(n)) = O(\sqrt{n})$. It is assumed that all nodes transmit at constant power P . Again, the physical channel is modeled as an AWGN channel treating all other interference as noise and only point-to-point coding is used. Fading is not considered. The per-packet delay is regarded as the number of hops needed to reach its destination. The most intriguing fact is that the proof suggests a hierarchical routing structure. One of the key step in the proof of [23] is to identify a set of nodes to construct a wireless backbone network to relay information for all other nodes. The proof highlights the importance for an efficient routing in multihop networks and partly motivates our study for efficient routing in Chapter 4.

2.3 Routing

Routing can be viewed as a special relaying protocol, *i.e.*, in routing, the relay only stores and forwards the message it receives and the destination typically does

not combine signals from different transmitters. Compared with many other relaying protocols, routing is simpler in terms of complexity at both relays and destinations. This section reviews some important routing literature related to this dissertation.

2.3.1 Routing Protocols in Wireless Networks

Distance Vector and Link State Routing

Two of the most popular routing algorithms are distance vector routing and link state routing [66]. In distance vector routing, each router maintains a table, *i.e.*, a vector, giving the best known distance or cost to each destination and which neighbor to use to get there. These tables are updated by exchanging information with the neighbors. More specifically, each node periodically broadcasts its distance vector, *i.e.*, its estimated cost of traveling to each destination. The neighbors, upon reception of a new distance vector from node u , check to see if a new route going through node u can reduce the cost. The corresponding entry of table is updated if a new route brings lower cost. Distance vector routing of [66] is essentially a distributed Bellman-Ford routing algorithm [18]. Distance vector routing converges slowly. In particular, it reacts rapidly when a new router joins the network, but reacts slowly when an existing router goes off-line. In fact, the well-known “count-to-infinity” problem [66] shows that it takes infinite time for nodes in a linear network to realize one of the neighbors is down.

Link state routing is introduced to overcome the slow convergence of the distance vector algorithm. Link state routing relies on Dijkstra’s algorithm to find the shortest path. Given a weighted, directed graph and a specific source, Dijkstra’s algorithm maintains two sets of vertices \mathcal{S} and \mathcal{Q} . Set \mathcal{S} contains all vertices to which we know the shortest path from the source and set \mathcal{Q} contains all other vertices. In each step, a vertex with the lowest cost is moved from \mathcal{S} to \mathcal{Q} . When a vertex u is moved to \mathcal{S} , for each neighbor v of the vertex u , the algorithm checks to see if

the shortest known path to v can be improved by first following the shortest path from the source to u and then to v . If this new path is better, the algorithm updates the cost of v with the new smaller value. Compared with the Bellman-Ford algorithm, the Dijkstra's algorithm has a smaller running time. However, as we can see, the Dijkstra's algorithm requires the complete topology and cost of all links in the network. Hence, link state routing protocols require measurement and distribution of the complete topology and link cost. Link state routing typically consists of the following steps [66]: learn about the neighbors when a router is first booted; measure line cost to neighbors by sending probe packets; build link state packets; distribute the link state packets by flooding; compute new routes. We note that the Open Shortest Path First (OSPF) protocol, widely used in Internet, uses a link state algorithm.

Source and Hop-by-Hop Routing

Depending on how the route is specified, routing can also be classified into source routing or hop-by-hop routing. In source routing, the source determines the complete route and attaches this information into packets. Intermediate routers simply read the routing information from the packets and execute it. By contrast, in hop-by-hop routing, the source only specifies the destination and the next hop. Intermediate relay nodes only have the address of the destination and determine independently the best next hop for each packet.

Proactive and Reactive Routing

Compared with routing protocols in wired networks, wireless routing protocols face many unique challenges, *e.g.*, they must provide scalable routing in the presence of mobile nodes. Hence, routing protocols from wired networks often cannot be applied directly to wireless networks, and a large body of research effort has been

devoted to designing new routing protocols that can adapt to topology change in wireless networks [37, 45]. Note that most wireless routing protocols are still based on the principle of either link state routing or distance vector routing, but differ from wired routing protocols in implementation details, *e.g.*, the content of distance vector or link state packets.

Both distance vector routing and link state routing require information exchange among nodes. Depending on how frequent the information is exchanged, routing protocols for wireless networks can be further classified as proactive or reactive [37, 45].

In proactive routing protocols, mobile nodes exchange routing information in the background and maintain a routing table regardless of communication needs. Many proactive protocols are adopted from traditional routing protocols used in wired networks. Proactive routing protocols in general have the desired properties of having low latency in establishing routes and Quality-of-Service (QoS) guarantees. However, maintaining a routing table in each mobile node is expensive as it requires constant exchange of distance vector or link state packets. Furthermore, many entries of the stored routing table might never be used. Hence, exchange of information about these entries is a waste of precious wireless resources, *i.e.*, power and bandwidth.

Reactive or on-demand routing protocols reduce routing overhead by sending enquiry packets to discover the route only when there is a communication request. Reactive routing protocols can scale well to large networks when traffic is mainly directed to a few destinations and the network has low mobility [37]. Reactive routing protocols, in particular, Ad Hoc On Demand Distance Vector Routing (AODV) and Dynamic Source Routing (DSR), have been extensively evaluated and are the leading candidates for standardization [37]. In AODV, intermediate nodes learn

the path to the source by receiving the query packet from the source and enter the route in the forwarding table. In DSR, the query packet sent out by the source stores the ID of intermediate nodes that it goes through before reaching the destination. DSR also allows nodes to keep multiple routes to a destination, hence reducing the demand of route rediscovery in the case of link breakage. The Location-Aided Routing (LAR) protocol, another on-demand protocol, operates similar to DSR, but reduces routing overhead by utilizing location information to limit the flooding area of enquiry packets.

2.3.2 Routing Metrics in Wireless Networks

Routing metrics assess the quality of a route from a source to a sink based on one or more criteria. Routing metrics are typically, but not always, the sum of the individual link metrics. Widely cited wireless routing metrics include Hop Count (HC) and Expected Number of Transmissions (ETX) [25, 24, 20].

We refer to routing based on minimizing HC as minimum hop-count routing. Minimum hop-count routing is particularly appealing due to its simplicity. The link metric for minimum hop-count routing is binary: either the link exists or it does not. Hence, computing the hop count requires no additional measurements, unlike the other metrics we will describe in this section. However, as shown in [20], in low-mobility networks, minimum hop-count route does not necessarily maximize the end-to-end flow because it does not take into account the effect of link qualities, *e.g.*, packet loss rate or bandwidth constraints.

The ETX routing metric estimates the total number of data transmissions required to send a packet from end to end, including retransmissions. In the presence of an ARQ mechanism, the expected number of transmissions and retransmissions on a link between node i and j is $1/P_{i,j}^e$, where $P_{i,j}^e$ represents an estimate of the

probability that node i receives an acknowledgment sent by node j . In [20], the acknowledgment probability is obtained by measuring loss rate of periodically sent probes. ETX depends on a link layer retransmission mechanism to estimate the packet loss rate by broadcasting probe packets. Hence, it increases routing overhead. ETX also assumes that radios have a fixed transmit power level and the probability that a given packet is lost in transmission is independent of the packet size.

2.3.3 Algebraic Theories for Routing

Recent research [63, 64] has developed algebraic theories for studying the properties of routing protocols, *e.g.*, convergence. These algebraic theories of [63, 64] work on a weighted directed graph. The weight function of link L , *i.e.*, $W(L)$, can be a function of one or more parameters, *e.g.*, delay and channel quality of a link. It turns out that the convergence and correctness of Dijkstra’s algorithm and link state routing depend on the *isotonicity* of the weight function [63]. In a network setting, the isotonicity of the weight function means that the order between the weight of two paths does not change if they are extended with the same new link. More specifically, the weight function is isotonic if $W(L_1) \leq W(L_2)$ implies both $W(L_1 \oplus L_3) \leq W(L_2 \oplus L_3)$ and $W(L_3 \oplus L_1) \leq W(L_3 \oplus L_2)$, where $L_1 \oplus L_2$ denotes the concatenation of two paths L_1 and L_2 . The weight function is strictly isotonic if $W(L_1) < W(L_2)$ implies both $W(L_1 \oplus L_3) < W(L_2 \oplus L_3)$ and $W(L_3 \oplus L_1) < W(L_3 \oplus L_2)$. We note that the isotonic property suggests any subpath of an optimal path is an optimal path, hence leading to the correctness of Dijkstra’s algorithm [18, 63]. In [63], it is further shown that isotonicity is necessary and sufficient for hop-by-hop routing. However, without strict isotonicity, not every implementation of Dijkstra’s algorithm results in loop-free hop-by-hop routing.

The algebraic theories are further extended to distance vector routing in [64] where it is shown that *monotonicity* is the property that guarantees the protocol convergence in every network. In a network, the weight function is monotonic if for any paths L_1, L_2 , we have $W(L_1) \leq W(L_1 \oplus L_2)$. Intuitively, monotonicity implies that the weight of a path does not decrease when it is extended by a new link. Furthermore, isotonicity assures convergence onto optimal paths for distance vector routing [64].

2.3.4 Information Theory and Routing

Previous work on routing within the networking community, *e.g.*, [20, 25], mainly studies how to design new routing metrics to improve the throughput, and how to modify existing routing protocols to incorporate new metrics. These models are often built on link-level abstractions of the network without fully considering the impact of the physical layer. There is little if any discussion about the fundamental performance limits, such as Shannon capacity or spectral efficiency.

On the other hand, as Section 2.2.4 summarizes, the study of wireless networks using information theory [31, 30, 41, 75] has led to several relaying protocols that are asymptotically order-optimal as the number of nodes goes to infinity. However, all practical networks have a finite number of terminals. Furthermore, relaying protocols derived from information theory often involve complicated multiuser coding techniques, such as block-Markov coding and successive interference cancellation, which are often too complex to implement in practical systems. Moreover, information-theoretic relaying strategies may not be easily implementable in a distributed manner. The gap between information-theoretic analyses and practical implementations has inspired us to study influence of different routing schemes on spectral efficiency and designs distributed routing schemes based on insights from

an information-theoretic analysis in Chapter 4.

The work that are most related with Chapter 4 are those in [62, 49, 50]. They provide important guidelines for designing spectrum-efficient networks. Assuming a one-dimensional linear network with AWGN channels, [62, 49, 50] show that there is an optimum number of hops in terms of maximizing end-to-end spectral efficiency. The results challenge the purely signal-to-noise ratio (SNR) guided traditional wireless routing paradigm of “the more hops the better”. Similar observations are also made for non-ergodic Rayleigh fading channels in [49] by considering how the end-to-end outage probability varies as a function of the number of hops. The trade-off between the delay, *e.g.*, length of coding intervals, and reliability is also studied through the perspective of error exponent in [80]. However, [62, 49, 50, 80] assume that the number of relays and their locations are design parameters. In practice, the network geometry changes as the network operates and evolves; thus, neither the number of available relay nodes nor their locations between a source and destination are design parameters. By contrast, Chapter 4 considers choosing a route in a network comprised of an arbitrary number of randomly located nodes.

2.4 Optimization

Optimization has been widely used in information theory, communications, and networking, *e.g.*, the celebrated Shannon channel coding theorem requires maximization of mutual information over all possible input distributions and Bellman-Ford is a version of dynamic programming to find the shortest path in a weighted graph. The increasing demand for wireless communications has made wireless resources, *e.g.*, power and bandwidth, scarce and made the potential benefits from optimizing the resource allocation particularly important. Although optimization techniques have been a subject of intensive study, it has not always been possible to solve an

optimization problem efficiently as we will see in the next section.

2.4.1 Complexity Theory

The concept of P and NP can be useful in classifying optimization problems and in helping us determine the strategy for tackling them.

An algorithm is called a polynomial-time algorithm if, given inputs of size n , its worst case running time is $O(n^k)$ for some constant k . Both Bellman-Ford and Dijkstra's algorithms belong to the class of polynomial-time algorithms. The class of problems that can be solved by polynomial-time algorithms are called P problems. NP problems refer to the class of problems whose solutions can be *verified* by polynomial-time algorithms. Hence, a P problem is also a NP problem, but not vice versa. A problem is NP-hard if any algorithm for solving it can be translated into one for solving any NP-problem. Hence, NP-hard is at least as hard as any NP problem. A problem is called NP-complete if it is both NP and NP-hard. No polynomial-time algorithm has yet been discovered for the class of NP-complete problems. In fact, if any NP-complete problem can be solved in polynomial time, every NP-complete problem has a polynomial-time algorithm [18].

Although it has not been formally proven, NP-hard and NP-complete problems are generally considered intractable problems. Hence, if a problem is NP-complete or NP-hard, we in general do not further pursue an optimum solution. Instead, we focus on efficient methods to provide suboptimum solutions. Many combinatorial optimization problems are NP-hard. It is also widely known that scheduling transmission in wireless networks with interference is NP-hard [3, 35].

2.4.2 Convex Optimization

Convex optimization problems are a class of optimization problems that are well behaved and lend themselves to both analytic tools and efficient numerical methods.

A standard formulation of optimization problems is as follows:

$$\min f_0(\mathbf{x}),$$

subjected to constraints

$$f_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, m$$

$$g_j(\mathbf{x}) = 0 \quad j = 1, \dots, n$$

where \mathbf{x} is a vector of finite dimension. This problem is a convex optimization if the objective function $f_0(\mathbf{x})$ is convex ¹ and the constraint functions $f_i, i = 1, \dots, m$ and $g_j, j = 1, \dots, n$ are also convex. A particular important property of convex optimization problems is that any locally optimal point is also globally optimal [11]. This property allows numerical methods to stop at a locally optimal point.

Duality theory, particularly, the Lagrangian duality theory, is widely used in analyzing optimization problems and can lead to efficient numerical methods [11]. The basic idea in Lagrangian duality is to augment the objective function with a weighted sum of the constraint functions, leading to the Lagrangian

$$L(\mathbf{x}, \lambda_1, \dots, \lambda_m, \nu_1, \nu_n) = f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^n \nu_j h_j(\mathbf{x}).$$

We can define the Lagrangian dual function as

$$g(\lambda_1, \dots, \lambda_m, \nu_1, \nu_n) = \inf_{\mathbf{x}} \{f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^n \nu_j h_j(\mathbf{x})\},$$

and define the Lagrangian dual problem as

$$\max g(\lambda_1, \dots, \lambda_m, \nu_1, \nu_n)$$

subjected to

$$\lambda_i \geq 0, \quad i = 1, \dots, m.$$

¹A function $f(x)$ is convex if $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$ for $0 \leq \lambda \leq 1$.

By the weak duality [11], the solution to the Lagrangian dual problem yields lower bounds on the optimal value of the original problem. Note that the Lagrangian dual function is always concave even if the original problem is not convex. For a convex optimization problem, strong duality holds, *i.e.*, the infimum of the Lagrangian function yields the optimal value for the original problem. Furthermore, if the objective and constraints functions are differentiable, we can further exploit the Karush-Kuhn-Tucker (KKT) conditions, *i.e.*, the primal and dual optimal points \mathbf{x}^* , λ_i^* , $i = 1, \dots, m$ and ν_j^* , $j = 1, \dots, n$ satisfy the following conditions:

$$\begin{aligned}
 f_i(\mathbf{x}^*) &\leq 0 \quad i = 1, \dots, m \\
 g_j(\mathbf{x}^*) &= 0 \quad j = 1, \dots, n \\
 \lambda_i^* &\geq 0 \quad i = 1, \dots, m \\
 \lambda_i^* f_i(\mathbf{x}^*) &= 0 \quad i = 1, \dots, m \\
 \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{j=1}^n \nu_j^* \nabla g_j(\mathbf{x}^*) &= 0
 \end{aligned}$$

where ∇ denotes the gradient operation. For a convex optimization problem, the KKT conditions provide both necessary and sufficient conditions for the solutions to be both primal and dual optimal [11]. Although convex optimization problems can in principle be solved by existing numerical methods, in practice, straightforward application of general numerical methods often encounters difficulties if the constraint functions are non-linear or the problem size is large. For such scenarios, the KKT conditions are particularly useful as they often lead to insights about the optimum solution and result in more efficient numerical methods.

In many cases, the optimization problem that we formulate based on practical constraints, *e.g.*, one variable must be an integer, often does not allow the application of existing optimization techniques. A typical approach to avoid this difficulty is to *relax* non-convex constraints, *e.g.*, treat the integer variable as a real variable, to

allow some analysis. The solution from the relaxation is then modified to satisfy the original constraint, *e.g.*, rounding the real number solution to an integer.

More comprehensive discussions of optimization techniques are available in [11, 8].

CHAPTER 3

RELAYING IN THE BLOCK-FADING MULTI-ACCESS CHANNEL

Emphasizing a graceful transition from existing wireless networks to wireless relay networks, this chapter demonstrates the significant gains that multi-access users can achieve from *sharing* a single amplify-forward relay in slow fading environments. In this system, the users need not be aware of the existence of the relay, *i.e.*, all cost and complexity is placed in the relay and destination. Such an architecture may be suitable for infrastructure networks, in which the relay and destination correspond respectively to a relay station and a base station deployed and managed by the service provider. It is worth noting that since a single relay is *shared* by multiple users in the MARC, the extra cost of adding the relay is amortized across many users and may thus be more acceptable, especially as the number of users in the system grows. In contrast with some other alternative wireless relay networks, *e.g.*, cooperative diversity, the MARC requires little, if any at all, effort of users to benefit from relaying. Thus, this approach facilitates a graceful transition, both technologic as well as economic, from existing systems to wireless relay networks, *e.g.*, cooperative diversity.

The remainder of the chapter is organized as follows. Section 3.1 describes the channel model and defines the performance metric, namely, diversity-multiplexing trade-off. Section 3.2 proposes the multi-access amplify-forward (MAF) protocol. Section 3.3 presents the diversity-multiplexing trade-off (DMT) of MAF and com-

compares the performance of MAF with several other alternative protocols to highlight the advantage of MAF. Section 3.4 presents simulation results to support our DMT analysis. Section 3.5 concludes this chapter.

The results of much of this chapter, particularly for the case of $N = 2$ users, are to be published in [15].

3.1 Background

3.1.1 Channel

The MARC is distinguished from the standard multi-access channel by the existence of one or more relays solely intended to facilitate communication between the users and the destination. Fig. 3.1 shows the MARC with two users and one relay.

All wireless links are assumed to be frequency non-selective and Rayleigh block fading, *i.e.*, the fading coefficients remain constant within a block of l symbols, but change independently from one block to the other according to a circularly symmetric complex Gaussian distribution with zero mean and unit variance. The terminals are assumed to be far enough from one another such that the fading coefficients of different links are independent. Moreover, the block length l is assumed to be long enough such that the channel state information (CSI) can be tracked at the receiving end, though it is not available to or otherwise not exploited by the transmitting end. It is assumed that the destination has knowledge of all CSI, including those of the user-relay links. The variance of the AWGN is taken to be unity.

Since we focus on the block fading scenario in this chapter, the term “MARC” therefore refers to the “block-fading MARC” in the rest of this chapter.

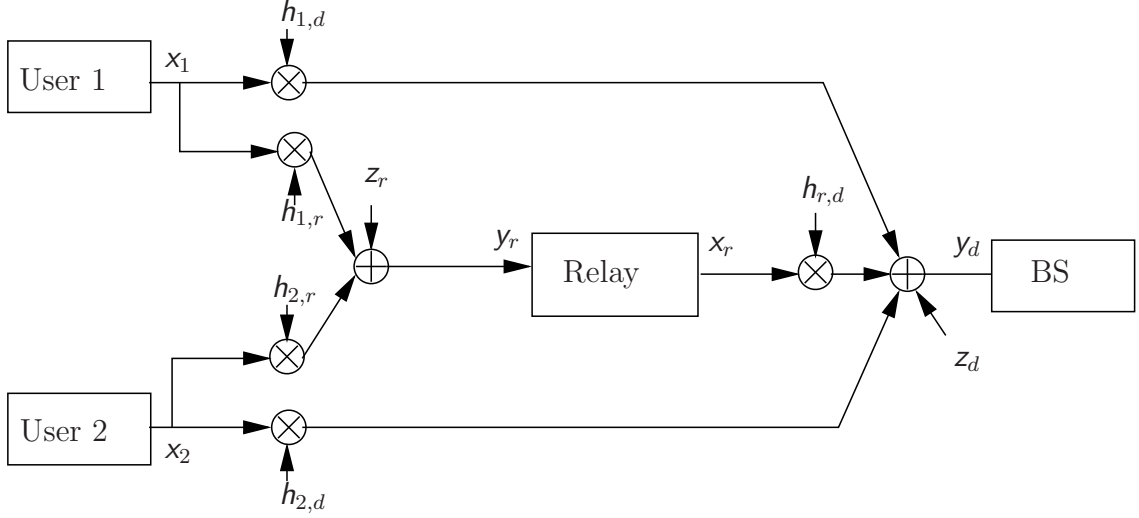


Figure 3.1. The Multi-access relay channel (MARC) with two users, multiplicative fading, and additive noise.

3.1.2 Analysis Method

In order to characterize the performance of the proposed protocol in the high SNR regime, DMT is adopted as the performance metric [81]. In Chapter 2, we have discussed the DMT for point-to-point channels. In this chapter, we extend the DMT to the MARC. More specifically, this chapter focuses on the symmetric case, *i.e.*, the users transmit their messages at the same data rate of R/N bits per channel use (bpcu), where N is the number of users in the system. Furthermore, the users and the relay use the same transmission power ρ since power allocation in the absence of transmitter CSI does not improve a protocol's DMT performance. Let $\mathcal{C}(\rho) = \{\mathcal{C}_1(\rho), \mathcal{C}_2(\rho), \dots, \mathcal{C}_N(\rho)\}$ denote a family of codes indexed by SNR ρ , such that user i 's codebook $\mathcal{C}_i(\rho)$ has data rate $R(\rho)/N$ and codelength l . As in the case of point-to-point channels, we let $P_{\mathcal{E}}(\rho, R(\rho))$ denote the error probability of the joint ML decoder at the base station and define the multiplexing gain r and

the diversity gain d as

$$r := \lim_{\rho \rightarrow \infty} \frac{R(\rho)}{\log \rho}, \quad d := \lim_{\rho \rightarrow \infty} -\frac{\log P_{\mathcal{E}}(\rho, R(\rho))}{\log \rho}.$$

3.2 Multi-Access Amplify-Forward

This section describes the proposed MAF protocol. In MAF, the relay listens to all users during the first half of the block; then, in the second half of the block, it simply amplifies and broadcasts the signal it received in the first half. All users continue transmitting their messages throughout the entire block. During the first half of the block, the equivalent channels seen by the destination and the relay are

$$y_d[j] = \sum_{i=1}^N h_{i,d} x_i[j] + z_d[j], \quad (3.1)$$

$$y_r[j] = \sum_{i=1}^N h_{i,r} x_i[j] + z_r[j], \quad (3.2)$$

respectively, where $1 \leq j \leq l/2$ denotes the time index; $h_{i,d}$ and $h_{i,r}$ denote the fading coefficients of the user i -destination and user i -relay links, respectively; and x_i denotes the signal transmitted by user i . Likewise, the equivalent channel seen by the destination during the second half of the block is

$$y_d[j] = \sum_{i=1}^N h_{i,d} x_i[j] + h_{r,d} x_r[j] + z_d[j] \quad \text{for } l/2 < j \leq l, \quad (3.3)$$

where $h_{r,d}$ denotes the fading coefficient of the relay-destination link, and x_r denotes the signal transmitted by the relay. Note that

$$x_r[j] = b y_r[j - l/2] \quad \text{for } l/2 < j \leq l,$$

where b denotes the relay's amplification coefficient. In practice, b should be chosen to, *e.g.*, minimize the outage probability at the target data rate and SNR, subject to the relay's transmission power constraint, *i.e.*,

$$|b|^2 \leq \frac{\rho}{\sum_{i=1}^N |h_{i,r}|^2 \rho + 1}. \quad (3.4)$$

However, as shown in [5], the particular value of \mathbf{b} does not affect the DMT of the protocol as long as its exponential order remains zero (which is guaranteed by (3.4)). Therefore, the optimization of \mathbf{b} is not further pursued in this chapter. Note that the base station needs to know the value of \mathbf{b} to decode the two messages. As the last comment, in the single user scenario, the MAF protocol reduces to the non-orthogonal amplify-forward (NAF) protocol of [5].

As mentioned earlier, one of the features of MAF is that it allows the users to operate as if in a normal MAC. In fact, the users may not be aware of the existence of the relay. Therefore, they simply use the capacity-achieving codebook for a standard MAC, *i.e.*, each codebook consists of i.i.d complex Gaussian random variables. Such inputs may not be optimal in terms of capacity or outage probability for MAF, due to the correlation that exists between the relay's signal and those of the users. However, as Theorem 1 shows, Gaussian inputs are optimal in terms of DMT at high multiplexing gains for the two-user MARC.

3.3 DMT of MAF

As introduced in Chapter 2, several other alternative relaying protocols, *e.g.*, DDF and CF have been proposed and studied for MARC [5, 79]. This section compares MAF with DDF and CF.

We first discuss different factors that impact these protocols based on intuition. DDF requires the relay to wait until it can fully decode both users' messages, hence removing the impact of fading and noise on source-relay links from the destination. However, as a result, the listening interval for the relay depends on channel realizations and rate requirements. In practice, the dynamic interval requirement might increase the difficulty of coding and scheduling. Moreover, when MARC operates with a high rate requirement, *e.g.*, a high multiplexing gain, DDF might require a

large portion of listening interval to decode at the relay such that the relay might not have enough time to transmit its decoded information. The CF strategy employs Wyner-Ziv source coding at the relay to compress its receive signal. The compressed signals are then coded using channel codes and are sent to the destination. The destination needs to first decode the compressed signals correctly and then decodes the sources' messages using the compressed signals and the signals directly transmitted from the sources jointly. Obviously, a high compression rate leads to low distortion, but also results in a higher rate requirement for the relay-destination link. And a higher rate for the relay-destination link might results in more interfere with signals from the sources. At last, for MAF, amplifying the signal at the relay will also amplify the noise as well. Another disadvantage of MAF is that it treats signals from both users unanimously by amplifying them with the same amplification coefficients even though one of the users might suffer from a deeper fade than the other.

The above comments highlight the different factors impacting the performance of DDF, CF and MAF. However, from these qualitative discussions, it is not immediately clear how the performance of MAF compares with that of CF and DDF given that they all have unique trade-offs. This section collects our results on the DMT for the symmetric MARC to allow a quantitative comparison between the performance of MAF, CF and DDF. Although the proofs for these DMT consist of an important part of this dissertation, due to the length of the proofs, we collect them at Appendices and focus on discussing insights offered by our DMT results in this section.

3.3.1 Symmetric Two-User MARC

The following theorem demonstrates the optimality of MAF protocol in the high multiplexing regime for the symmetric MARC with two users.

Theorem 1 *For the symmetric MARC with two users and one relay, the DMT of the MAF protocol is given by*

$$d_{MAF}(r) = \begin{cases} 2 - \frac{3r}{2}, & \text{for } 0 \leq r \leq \frac{2}{3} \\ 3(1 - r), & \text{for } \frac{2}{3} \leq r \leq 1 \end{cases}. \quad (3.5)$$

Proof: See Appendix A. ■

For purposes of comparison, an upper bound on the achievable DMT for the symmetric two-user MARC is first provided, along with the DMT's of the DDF and CF protocols. For the symmetric MARC with two users and one relay, an upper bound on the achievable DMT for $0 \leq r \leq 1$ is [7]

$$d_{MARC}(r) \leq \begin{cases} 2 - r, & \text{for } 0 \leq r \leq \frac{1}{2} \\ 3(1 - r), & \text{for } \frac{1}{2} \leq r \leq 1 \end{cases}. \quad (3.6)$$

On the other hand, the DMT of DDF for $0 \leq r \leq 1$ is [7]

$$d_{DDF}(r) = \begin{cases} 2 - r, & \text{for } 0 \leq r \leq \frac{1}{2} \\ 3(1 - r), & \text{for } \frac{1}{2} \leq r \leq \frac{2}{3} \\ \frac{2(1-r)}{r}, & \text{for } \frac{2}{3} \leq r \leq 1 \end{cases}, \quad (3.7)$$

and that of CF for $0 \leq r \leq 1$ is [79]

$$d_{CF}(r) = \begin{cases} 1 - \frac{r}{2}, & \text{for } 0 \leq r \leq \frac{4}{5} \\ 3(1 - r), & \text{for } \frac{4}{5} \leq r \leq 1 \end{cases}. \quad (3.8)$$

To highlight the advantage gained from adding a single relay, the DMT of a symmetric MAC with two users is also provided [68], *i.e.*,

$$d_{MAC}(r) = \begin{cases} 1 - \frac{r}{2}, & \text{for } 0 \leq r \leq \frac{2}{3} \\ N(1 - r), & \text{for } \frac{2}{3} \leq r \leq 1 \end{cases}. \quad (3.9)$$

Fig. 3.2 shows these trade-offs, along with the trade-off for MAF as given by (3.5). Inspecting these results and Fig. 3.2, the following observations for the symmetric two-user MARC can be made:

1. The MAF protocol achieves the optimal DMT for $2/3 \leq r \leq 1$. In fact, over this range of multiplexing gains, MAF behaves like a MISO system with three transmit antennas and one receive antenna. This observation together with our proof in Appendix A suggests that in the high multiplexing regime, the dominant error event is that both users cannot be decoded.
2. MAF outperforms the CF protocol in terms of DMT, *i.e.*, $d_{MAF}(r) \geq d_{CF}(r), \forall r$. Relative to MAF, CF appears to suffer from a significant loss in diversity gain at low multiplexing gains. In particular, MAF achieves the full diversity gain of 2 as r vanishes to 0; in contrast, CF only achieves a diversity gain of 1 as r vanishes to 0. Compared to CF, MAF enjoys another advantage of lower complexity at the relay.
3. It is interesting to observe that for $2/3 \leq r \leq 1$, MAF outperforms DDF in terms of DMT, despite the fact that AF protocols generally suffer from a significant performance loss relative to DDF in the single user relay channel [44, 5]. The advantage of MAF over DDF at high multiplexing gains can be attributed to the fact that in this operating regime, DDF requires the relay to spend a large percentage of time on decoding the two users' messages. As a result, the relay may not have enough time to re-transmit the decoded information. On the other hand, it has been observed that the performance of AF protocols improves in certain multi-user scenarios, *e.g.*, for the cooperative multi-access channel, the nonorthogonal amplify-forward (CMA-NAF) protocol achieves the optimal DMT [5]. Therefore, the advantage of MAF over DDF at high multiplexing gains should not be surprising.
4. In the regime of $2/3 \leq r \leq 4/5$, neither DDF nor CF is optimal, but MAF is. To the best of our knowledge, MAF is the only protocol that achieves the

optimal DMT in this regime.

5. Even over the range of multiplexing gains for which MAF becomes suboptimal, *i.e.*, $0 \leq r \leq 2/3$, the achieved DMT is identical to that of the NAF relay [5] with a single user. In other words, for low multiplexing gains, each user benefits from the relay as if it is the only user present. Our proof in Appendix A further suggests that in the low multiplexing regime, the dominant error event is that only one user cannot be decoded. Also, in this regime, the DMT gap between DDF and MAF is much smaller compared to the gap between DDF and CF.
6. For all multiplexing gains, MAF outperforms MAC in terms of DMT. This observation reveals the significant advantage that users can potentially gain from a single MAF relay. The DMT of DDF approaches that of the MAC in the high multiplexing regime. Thus, the gain of a complicated DDF relay diminishes for high multiplexing gains. The DMT of CF overlaps with that of the MAC for low multiplexing gains. This suggests that there may be little advantage of employing a CF relay for multiple users when the multiplexing gain is small.

Although the above discussions are limited to the two-user MARC, a lot of insights also hold for the MARC with more than two users as we will see in the sequel.

3.3.2 Symmetric N -User MARC, $N > 2$

The following theorem provides the upper and lower bounds on DMT for the symmetric MARC with N users, $N > 2$.

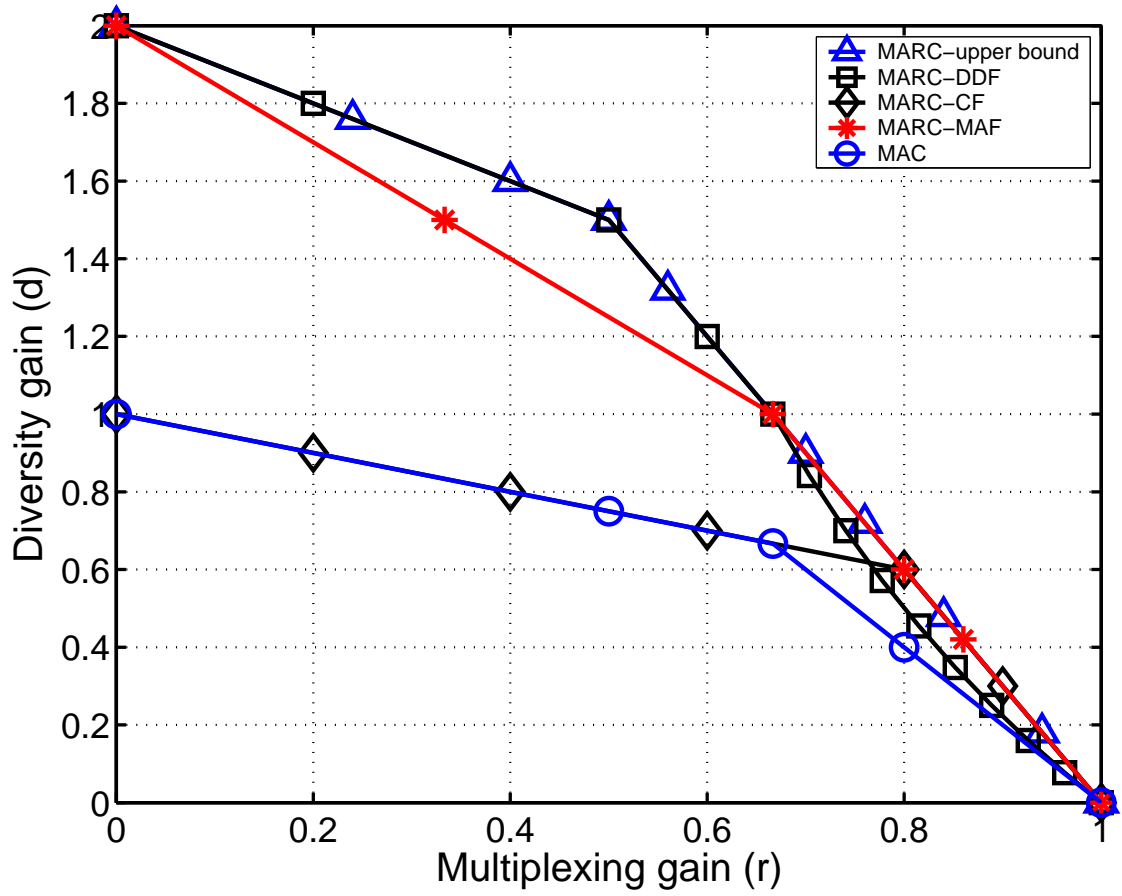


Figure 3.2. DMT of the different protocols for the two-user symmetric MARC.

Theorem 2 For the symmetric MARC with N users, $N > 2$, and one relay, the DMT of the MAF protocol satisfies

$$d_{MAF}(r) \geq \begin{cases} 2 - \frac{1+N}{N}r & \text{for } 0 \leq r \leq \frac{N(N-2)}{N^2-3} \\ N(1-r) & \text{for } \frac{N(N-2)}{N^2-3} \leq r \leq 1 \end{cases}, \quad (3.10)$$

and

$$d_{MAF}(r) \leq \begin{cases} 2 - \frac{1+N}{N}r & \text{for } 0 \leq r \leq \frac{N(N-1)}{N^2+N-3} \\ (N+1)(1-r) & \text{for } \frac{N(N-1)}{N^2+N-3} \leq r \leq 1 \end{cases}. \quad (3.11)$$

Proof: See Appendix B. ■

We also provide a cut-set bound on the DMT of the symmetric MARC with N users,

$$d_{MAF}(r) \leq \begin{cases} 2 - r & \text{for } 0 \leq r \leq \frac{N(N-1)}{N^2+N-3} \\ (N+1)(1-r) & \text{for } \frac{N(N-1)}{N^2+N-3} \leq r \leq 1 \end{cases}. \quad (3.12)$$

Fig. 3.3 and Fig. 3.4 provide these trade-offs for the symmetric MARC with 3 and 4 users, respectively. Inspecting these plots together with Theorem 2 and (3.12) suggests that

- The DMT lower bound for MAF (3.10) overlaps with the upper bound (3.11) in the low multiplexing regime, *i.e.*, $r \leq \frac{N(N-2)}{N^2-3}$. Hence, the upper bound (3.11) is tight in the low multiplexing regime. As a result, the observation that, in the low multiplexing regime, MAF allows each user benefit from the relay as if it is the only user present, holds for the more general N -user MARC.
- Note that (3.11) provides a tighter upper bound on the DMT of MAF in the low multiplexing gain regime than the cut-set bound (3.12). However, both upper bounds, *i.e.*, (3.11) and (3.12) overlap in the high multiplexing gain regime.
- The multiplexing gain regime in which the upper bound (3.11) is tight grows as the number of user increases. Moreover, as $N \rightarrow \infty$, the upper bound (3.11),

the lower bound (3.10), and the cut-set bound (3.12) all approaches $2 - r$ for $0 \leq r \leq 1$, indicating that, as the number of users grows, the performance gap between MAF and an optimum protocol becomes smaller.

- For $N = 2$, (3.11) is the same as (3.5), *i.e.*, (3.11) is also a tight bound for the two-user MARC in the whole multiplexing regime. This is surprising because (3.11) and (3.5) are proved via different methods. This observation leads us to conjecture that (3.11) is tight in the whole multiplexing regime. If our conjecture holds, MAF is DMT-optimal in the high multiplexing regime for the N -user MARC due to the fact that the curve of (3.11) overlaps with that of (3.12) in the high multiplexing regime. Although we have tried to prove this conjecture, it appears rather difficult and is not further pursued in this dissertation.

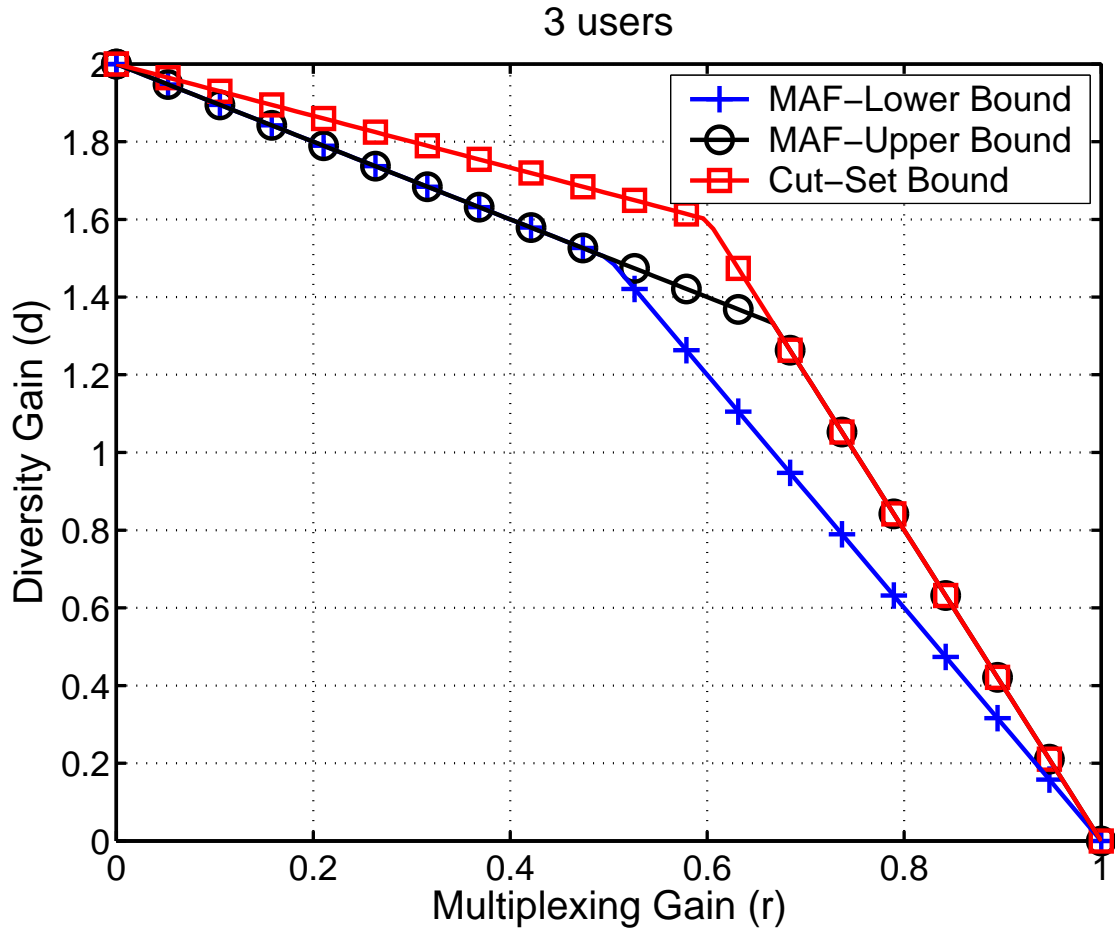


Figure 3.3. Bounds on the DMT of 3-user MARC

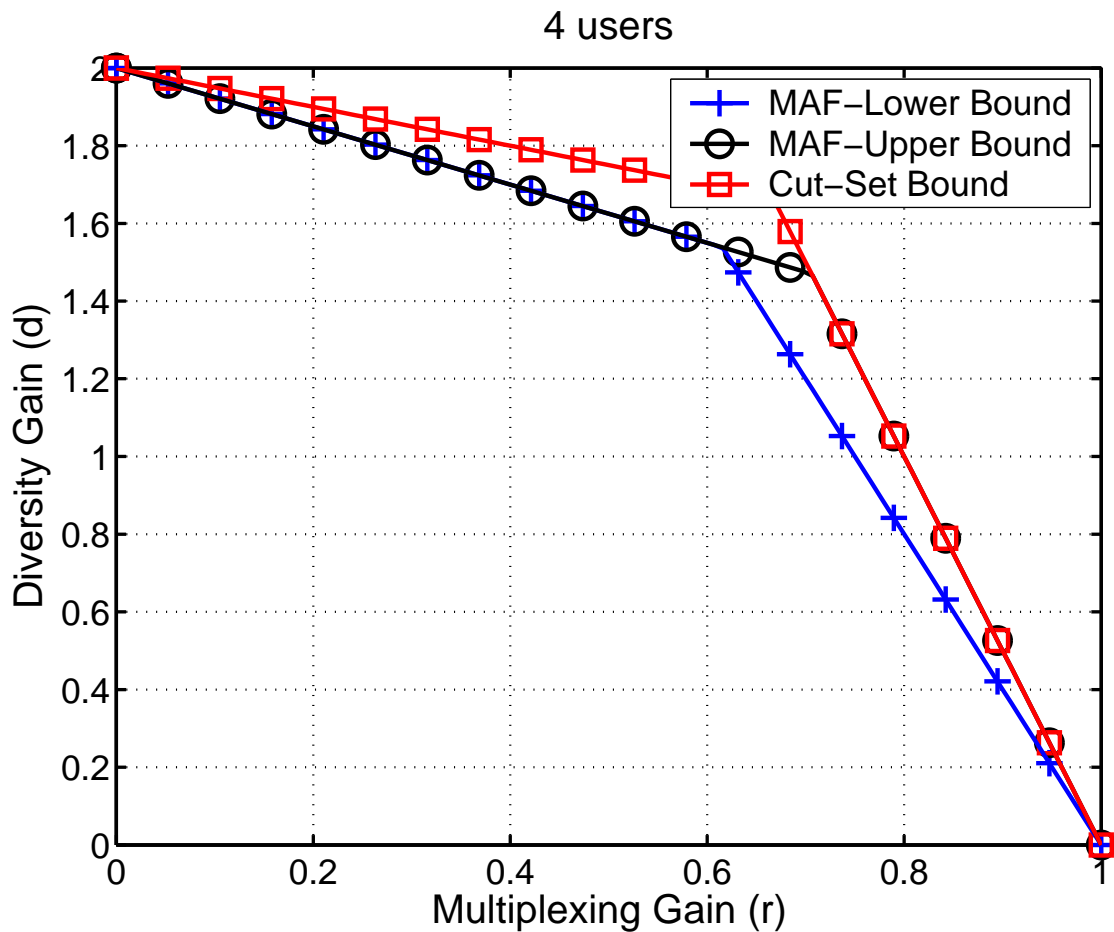


Figure 3.4. Bounds on the DMT of 4-user MARC

3.4 Numerical Simulation

Section 3.3 compares the performance of MAF, DDF and CF in terms of DMT. The analysis of DMT relies on taking the limit of SNR to infinity. This section provides simulation results for high but finite SNR to show that many insights offered by the DMT comparisons hold for finite SNR. The simulation results also highlight the importance of system parameters that are neglected by the DMT analysis, *e.g.*, coding gain, for a comprehensive system design.

3.4.1 Symmetric Two-User MARC

This section presents some simulation results for CF, MAF and DDF for the two-user symmetric MARC. For CF, it is assumed that the relay listens to the source for the first half of the block, and then re-encodes and transmits the messages in the second half of the block. This assumption may not be optimum in terms of minimizing the outage probability of CF, but it does not affect the DMT of the protocol [79]. For MAF, the amplification coefficient b is chosen such that (3.4) holds with equality. This choice of amplification coefficient may not minimize the outage probability of MAF, but it again does not affect the DMT. For a multiplexing gain of r , the total data rate R is chosen to be $r \log(1 + \rho)$.

Fig. 3.5 and Fig. 3.6 show the outage probabilities $P_o(R)$ of CF, DDF and MAF for $r = 0.4$ and 0.8 , respectively. As can be seen in Fig. 3.5, for $r = 0.4$, the outage probability curve of DDF enjoys a steeper slope compared to that of MAF and CF, indicating a higher diversity gain for DDF. Fig. 3.5 also shows that for $r = 0.4$, MAF offers a better diversity gain relative to CF. For $r = 0.8$, the intersection between the curve of MAF and that of DDF in Fig. 3.6 suggests that MAF has a higher diversity gain. These observations are in line with the DMT analysis, *i.e.*, the diversity gain of MAF is higher than that of DDF at high multiplexing gains (*e.g.*, $r = 0.8$), but smaller at low multiplexing gains (*e.g.*, $r = 0.4$).

It is interesting to notice that even though Fig. 3.2 suggests both MAF and CF provide a better DMT than DDF for $4/5 \leq r \leq 1$, DDF yields a better outage probability than both MAF and CF over a large range of SNRs as can be seen from Fig. 3.6. Furthermore, Fig. 3.6 demonstrates outage performance differences between MAF and CF, despite the fact that their DMT curves are identical for $4/5 \leq r \leq 1$. In particular, the slopes of the outage curves for CF and MAF appear to differ slightly. This small difference could be due to the finite SNR at which these

simulations have been carried out, and the suboptimum choice of the amplification coefficient b in MAF. These *discrepancies* between the outage simulation results and the DMT results suggest that the complete system design requires characterization of not only the DMT, which captures the *exponential behavior* of the error probability with SNR, but also the multiplicative coefficient that captures the *geometric dependence* and the *coding gain* of the relaying protocols.

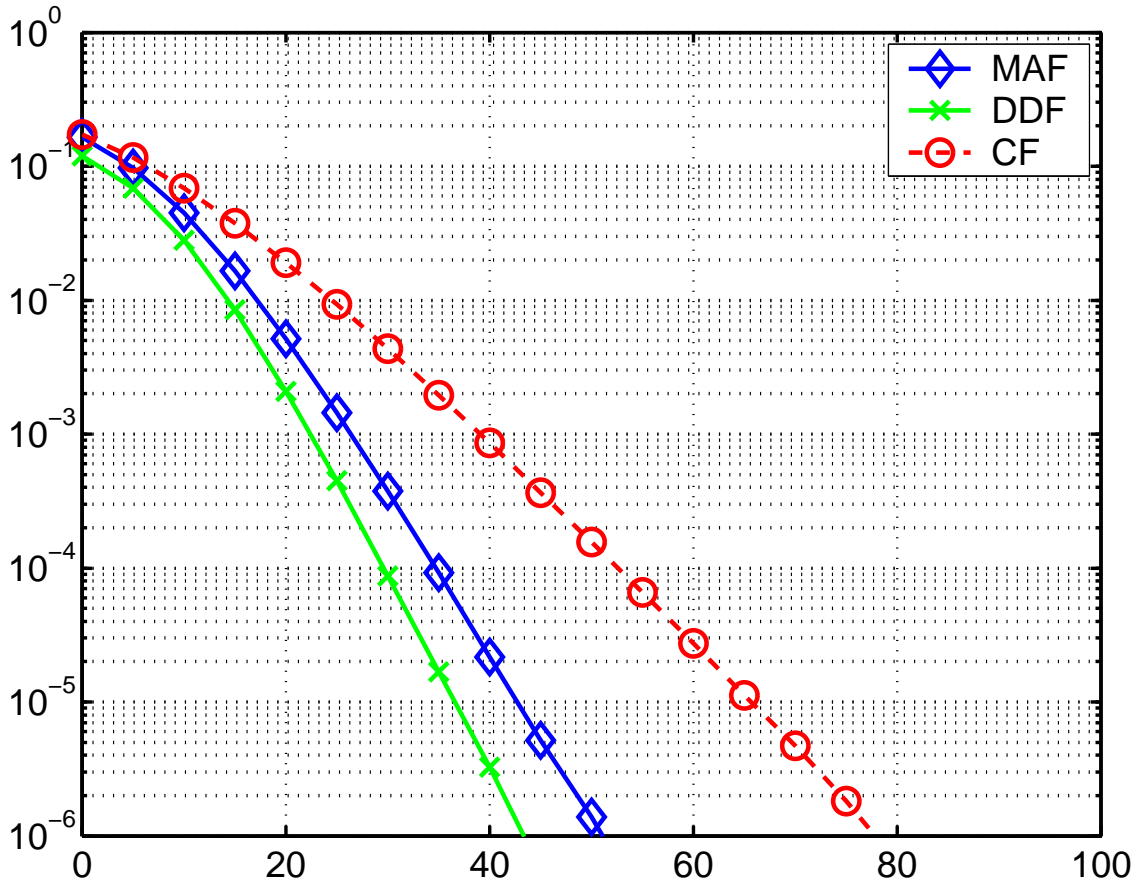


Figure 3.5. Outage probabilities $P_O(R)$ for CF, DDF and MAF in the two-user MARC. The multiplexing gain is $r = 0.4$. Note that $R = r \log(1 + \rho)$

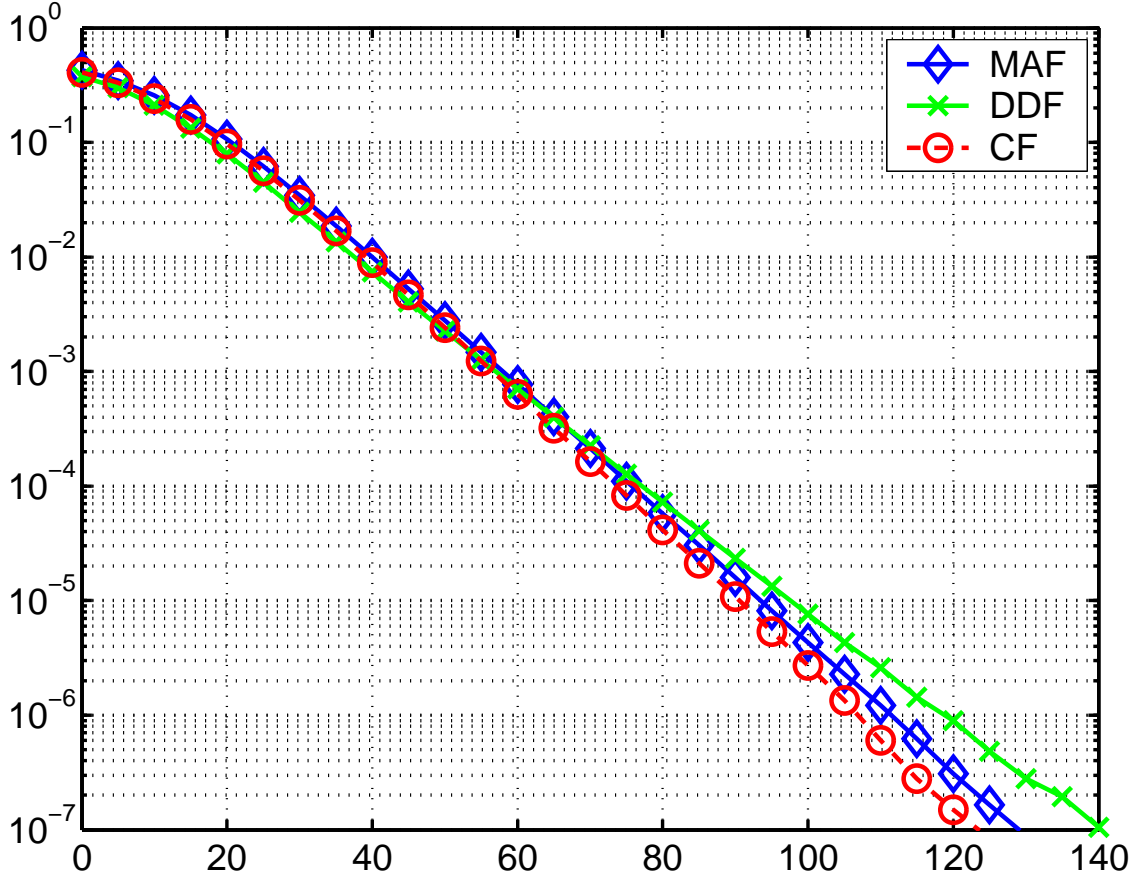


Figure 3.6. Outage probabilities $P_{\mathcal{O}}(R)$ for CF, DDF and MAF in the two-user MARC. The multiplexing gain is $r = 0.8$. Note that $R = r \log(1 + \rho)$

3.4.2 Symmetric N -User MARC, $N > 2$

This section provides simulation results for the N -user MARC for $N > 2$ to show that MAF also exhibits very good performance in MARC with more than two users. Other than the number of users, the simulation setup in this section is the same as those in Section 3.4.1.

Fig. 3.7 and Fig. 3.9 show the outage performance curves of MAF, DDF and CF with $r = 0.4$ for the MARC with 3 and 4 users, respectively. Similar with the case of two-user MARC, the outage performance of MAF is slightly inferior compared with DDF. However, a comparison of Fig. 3.5, Fig. 3.7 and Fig. 3.9 clearly indicates that

the slope difference between the outage curve of MAF and that of DDF decreases as the number of user increases. This observation from simulation is consistent with our previous prediction from analysis, *i.e.*, as the number of users grows, the performance gap between MAF and an optimum protocol becomes smaller. By contrast, the outage performance of CF is significantly worse than both MAF and DDF.

In the high multiplexing regime, as Fig. 3.8 and Fig. 3.10 clearly indicate, MAF enjoys a better diversity gain compared with DDF and CF, supporting our conjecture that MAF is optimal in terms of DMT for MARC with more than two users.

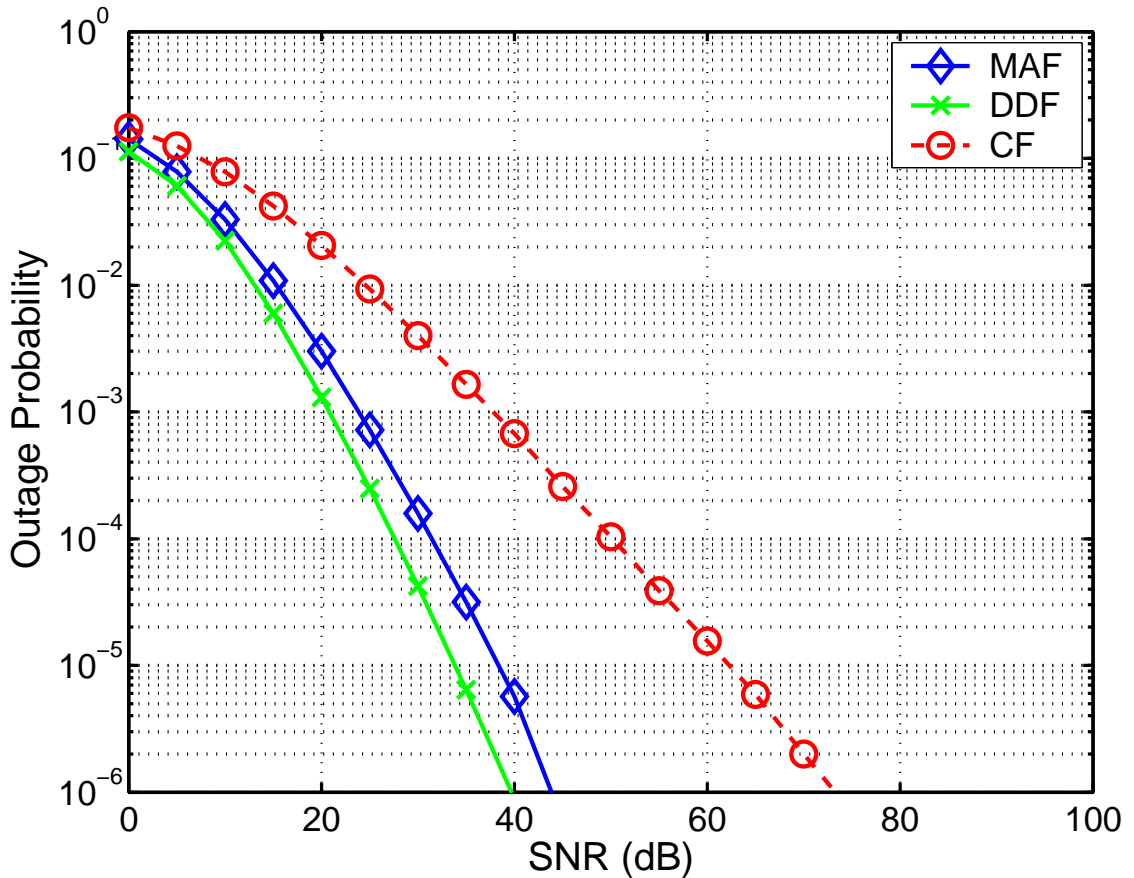


Figure 3.7. Outage probabilities $P_{\Theta}(R)$ for CF, DDF and MAF in the 3-user MARC. The multiplexing gain is $r = 0.4$. Note that $R = r \log(1 + \rho)$

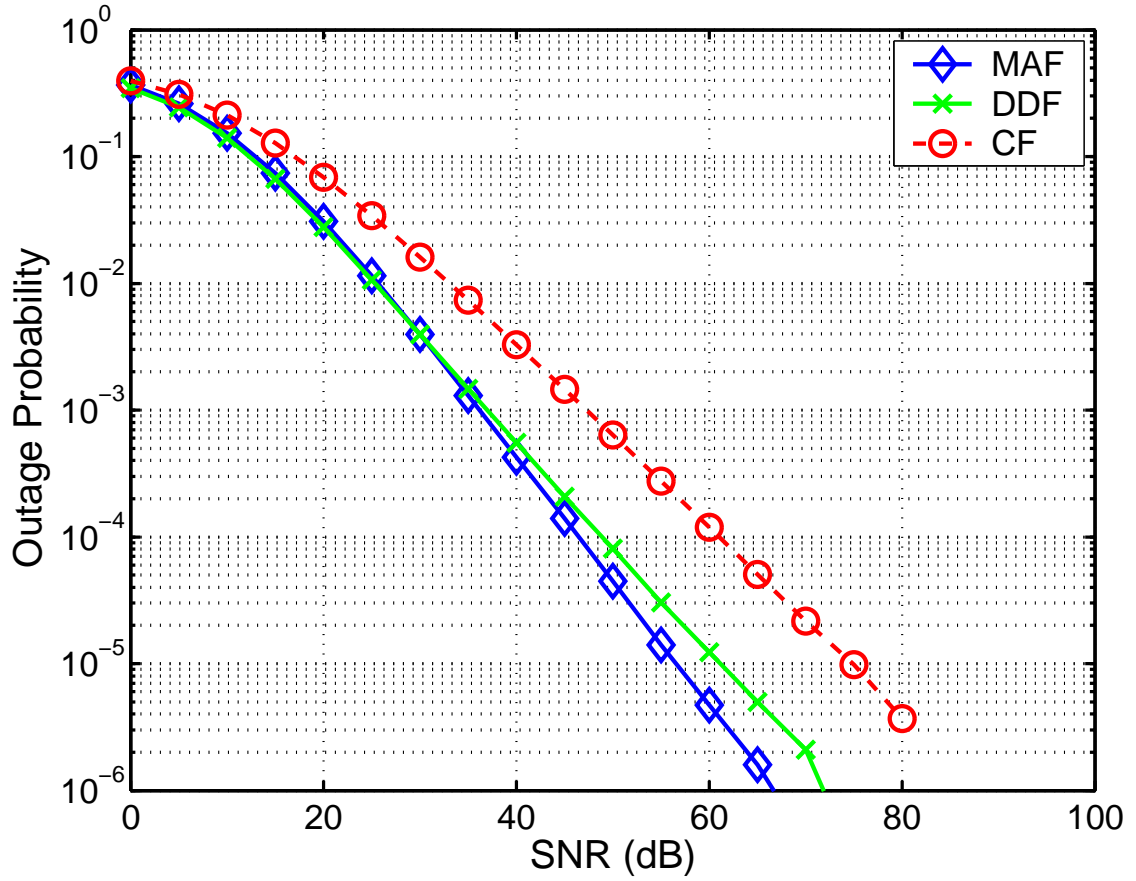


Figure 3.8. Outage probabilities $P_O(R)$ for CF, DDF and MAF in the 3-user MARC. The multiplexing gain is $r = 0.8$. Note that $R = r \log(1 + \rho)$

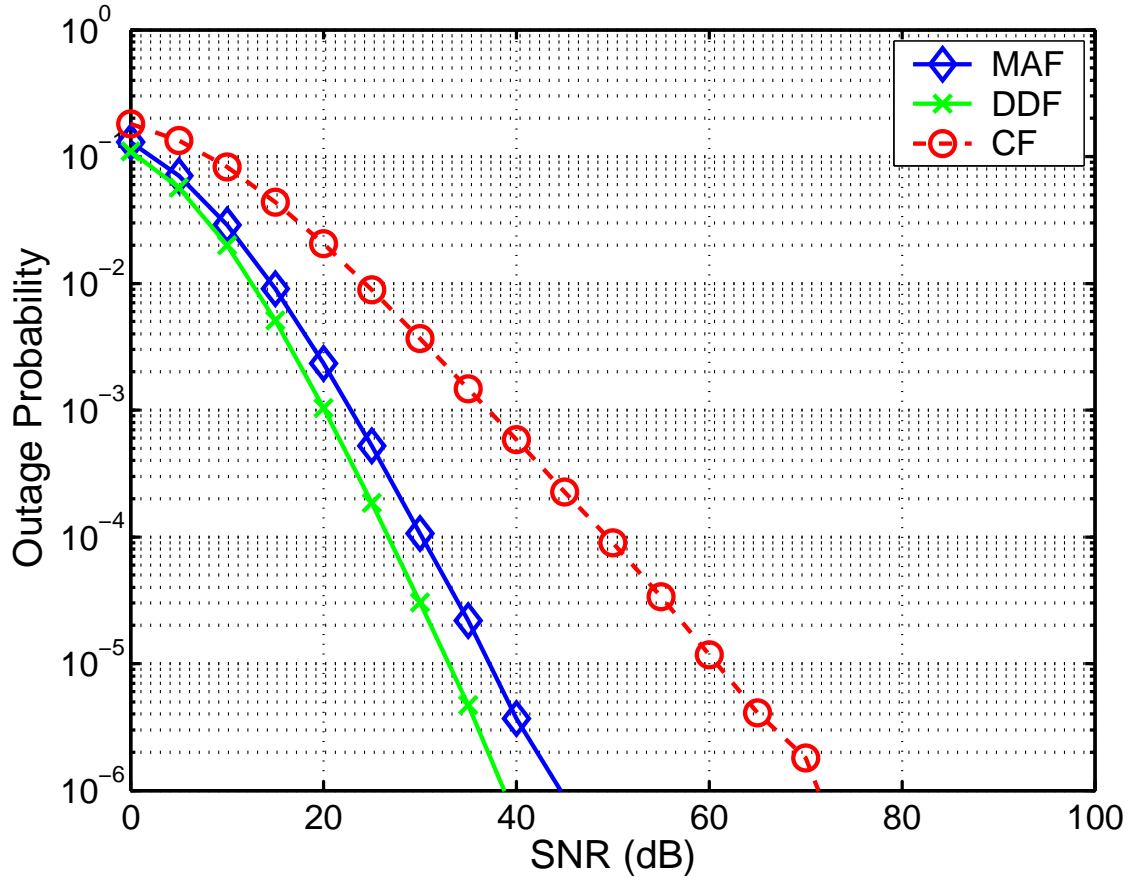


Figure 3.9. Outage probabilities $P_O(R)$ for CF, DDF and MAF in the 4-user MARC. The multiplexing gain is $r = 0.4$. Note that $R = r \log(1 + \rho)$

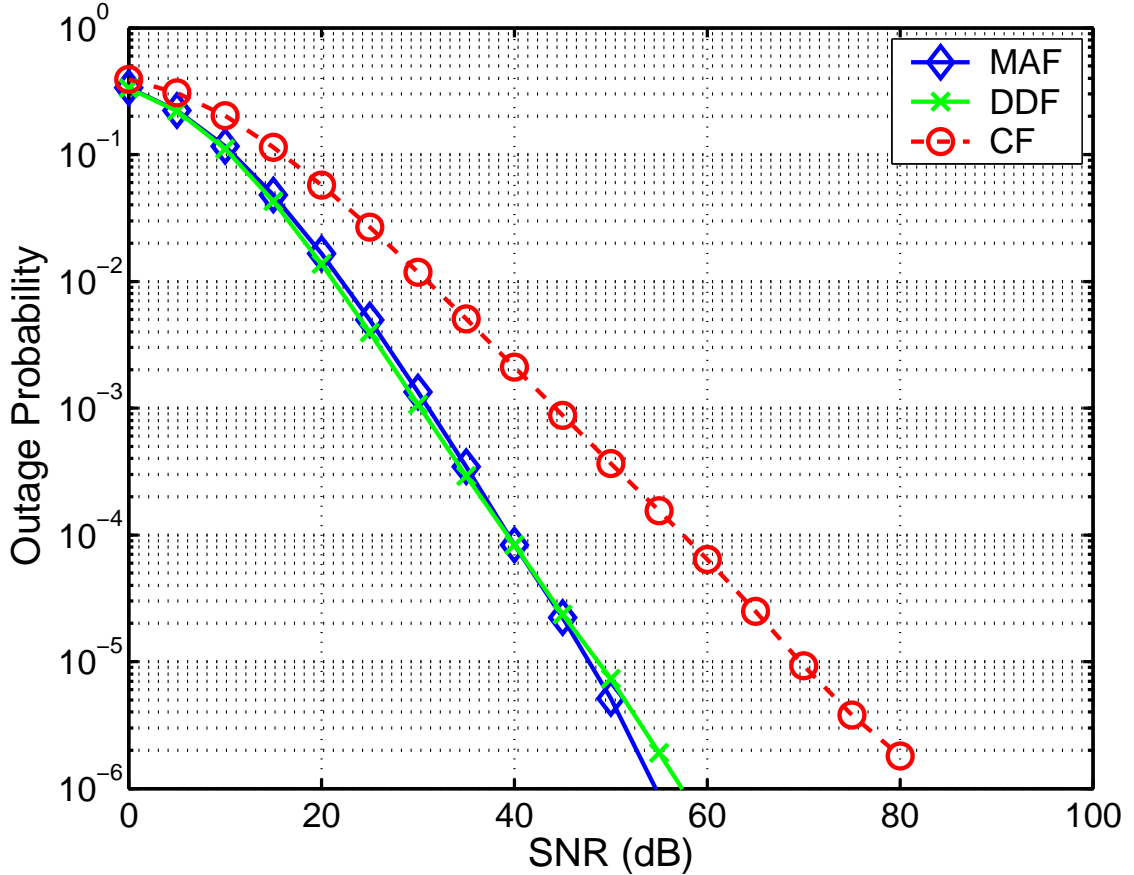


Figure 3.10. Outage probabilities $P_O(R)$ for CF, DDF and MAF in the 4-user MARC. The multiplexing gain is $r = 0.8$. Note that $R = r \log(1 + \rho)$

3.5 Conclusion

In contrast to the majority of previous works on multi-access relay channel (MARC), which focus on protocols that require complicated signal processing at the relay [7, 79], this chapter proposes a linear relaying protocol, *i.e.*, the multi-access amplify-forward (MAF), which not only enjoys a low complexity at the relay, but also exhibits a very good performance in slow-fading environments.

In particular, for the two-user MARC, MAF achieves the optimal diversity-multiplexing trade-off over the range $2/3 \leq r \leq 1$. Interestingly, for $2/3 \leq r \leq 4/5$

neither DDF nor CF, despite their complicated relaying algorithms, is optimal. In the low multiplexing gain regime, *i.e.*, $0 \leq r \leq 2/3$, MAF allows each user to gain cooperative diversity as if there is no interference from the other users and no contention for the relay.

For MARC with more than two users, this chapter provides upper and lower bounds on the DMT of MAF. The bound is tight in the low multiplexing regime. We further conjecture that the upper bound is tight in the high multiplexing regime as well. We support our conjecture through simulations.

In conclusion, MAF provides good performance at a low complexity, and is therefore an appealing architecture for wireless relay networks.

CHAPTER 4

SPECTRUM-EFFICIENT ROUTING IN MULTIHOP NETWORKS

Chapter 3 demonstrates a single relay station can benefit multiple sources in a multi-access channel. This chapter considers a different scenario in which multiple relay stations are available for relaying information from one source to the base station, *i.e.*, a multihop network. This scenario can appear for a wireless backbone network consisting of multiple wireless relay stations or for a wireless mesh network in which nodes rely on their peer to relay information to reach the access point. As Chapter 2 summarizes, research from different perspectives, *e.g.*, networking and information theory, often results in different relaying paradigms for wireless networks [31, 30, 41, 75, 37, 20, 25]. The goal of this chapter is to study the wireless routing problem combining networking and information-theoretic perspectives. We focus on routing, a practical relaying protocol, as analysis for other more complicated relaying protocols gets more difficult for multiple relays. We discuss the trade-off between the power and bandwidth for a multihop network and utilize the insight from an information theoretic analysis to design routing protocols.

The remainder of the chapter is organized as follows. Section 4.1 describes the system model and assumptions. Section 4.2 formulates the problems of finding a route with the maximum spectral efficiency assuming both the optimal bandwidth allocation and the equal bandwidth allocation. Since bandwidth allocation requires exchange of at least some global information, most of the chapter focuses on provid-

ing solutions for the case of equal bandwidth sharing. Specifically, Section 4.3 proposes approximately-ideal-path routing (AIPR), a location-assisted routing scheme, and Section 4.4 proposes distributed spectrum-efficient routing (DSER) scheme as another near-optimum solution. The spectral efficiency of AIPR and DSER closely follows the optimal spectral efficiency. Furthermore, DSER can be implemented with standard distributed algorithms that are guaranteed to converge and generate loop-free paths. Section 4.5 discusses the applications of AIPR and DSER in interference-limited networks, the connections between DSER and other well-known routing schemes, and the potential impact of finite input alphabets on our routing schemes. Section 4.6 presents simulation results, and Section 5.7 concludes the chapter.

Most results of this chapter are to be published in [16].

4.1 System Model

For simplicity, we only consider routing for one source-destination pair and limit our study to single-path routing. Also we do not allow the links to exploit cooperative diversity, *e.g.*, [60, 44]. We represent the nodes in a network and the possible transmissions between nodes by a complete graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} represents the set of nodes in the network and \mathcal{E} represents the set of edges (links). In general, nodes are arbitrarily located. For each link $e \in \mathcal{E}$, we use $t(e)$ to represent the transmit end of the link and $r(e)$ to represent the receive end. A path L from node s to node d , $s \neq d$, consists of an ordered sequence of unique links $l_1, l_2, l_3, \dots, l_n \in \mathcal{E}$ that satisfies the following: for each $1 \leq k \leq n - 1$, $r(l_k) = t(l_{k+1})$; $t(l_1) = s$; and $r(l_n) = d$. We also denote the source and destination of a given path L as $t(L) = t(l_1)$ and $r(L) = r(l_n)$, respectively. The length of the path $|L|$ is the number of links or hops in the path.

We assume all links in the network is corrupted with AWGN and path-loss. We adopt the path-loss model in (2.2) and (2.3). We denote the path-loss factor from node i to node j , $i, j \in \mathcal{V}$ as $G_{i,j}$ or $G_l, t(l) = i, r(l) = j$. And we define the SNR on link l as

$$\rho_l = \rho G_l. \quad (4.1)$$

Except Section 4.5.3 and Section 4.6.2, we assume Gaussian input throughout this chapter.

In general, optimum scheduling in networks is NP-hard [35]. To avoid the difficulty of jointly optimizing routing and scheduling, we first assume the network operates with time division multiple access (TDMA) without spatial reuse, *i.e.*, each node transmits in its own unique time slot. For a network with spatial reuse of bandwidth, it is important to design the medium access control (MAC) layer judiciously to mitigate interference. In this case, the interference stays approximately constant over time, and our framework is directly applicable by adding the interference to the noise. In Section 4.5.1, we will discuss extensions of our routing schemes to a simple scheduling scheme that allows for some spatial reuse.

4.2 Optimal Routing

This section discusses selection of routing paths that maximize the end-to-end spectral efficiency. The resulting optimization problem and its solution depends on whether the bandwidth or time slots can be optimally allocated across links.

Given optimum bandwidth allocation among links, [49] shows that the maximum spectral efficiency along a route L is

$$\frac{1}{\sum_{l \in L} \frac{1}{\log(1 + \rho_l)}}. \quad (4.2)$$

Since the denominator of (4.2) is additive, we can use Bellman-Ford or Dijkstra's algorithms with a link metric of $1/\log(1 + \rho_l)$ to find the route that maximizes the

spectral efficiency by minimizing $\sum_l (1/\log(1 + \rho_l))$ [49]. We refer to such a routing scheme as optimal routing with bandwidth optimization (ORBO). Although ORBO can be performed in a distributed fashion, allocating bandwidth requires propagating the value of (4.2) backwards to the nodes on the route [49]. As we will see, ORBO is most beneficial in the low SNR regime, where the power spent in distributing this knowledge may not be neglected. Another concern about bandwidth optimization is the issue of fairness, as one node with a larger share of the bandwidth might spend more energy than other nodes with a smaller share of the bandwidth. For these reasons, the rest of the chapter focuses on the case of equal bandwidth sharing.

Under the constraint of equal bandwidth sharing, the end-to-end spectral efficiency of a given path L is [22, 53]

$$R_L = \min_{l \in L} \frac{1}{|L|} \log(1 + \rho_l), \quad (4.3)$$

where the factor $1/|L|$ results from the sharing of bandwidth among relay links. For a path L , the signal quality is reflected by the worst link SNR $\rho_L^* = \min_{l \in L} \rho_l$, and the bandwidth use is characterized by inverse of the number of hops $|L|$. The spectral efficiency (4.3) increases as ρ_L^* increases or $|L|$ decreases. However, for routes connecting a given source and destination, if the number of links $|L|$ increases (or decreases), there are more (or less) relay nodes and ρ_L^* is more likely to increase (or decrease) due to shorter (longer) inter-relay distances. This trade-off can be seen by comparing the nearest-neighbor route and the single-hop route (the source directly transmits to the destination) in a linear network. Among all routes connecting a given source and destination, the nearest-neighbor route has the largest ρ_L^* but also the largest $|L|$. On the other hand, single-hop has the smallest ρ_L^* , but also has the smallest $|L|$. Therefore, there is a trade-off between physical layer parameters, *i.e.*, signal quality and bandwidth use, in selection of routes. The optimal routing scheme takes this trade-off into account by providing a solution to the optimization

problem

$$\max_{L:t(L)=s,r(L)=d} \min_{l \in L} \frac{1}{|L|} \log(1 + \rho_l), \quad (4.4)$$

where nodes s and d form the desired source-destination pair.

Unfortunately, generalized Bellman-Ford and Dijkstra's algorithms cannot be used to solve (4.4), because the routing metric (4.3) is neither isotonic nor monotone [63, 64]. In general, the computation of the spectral efficiency by (4.3) requires global information about a path. Therefore, the problem in (4.4) does not exhibit the optimal substructure that is necessary for the use of dynamic programming methods [18]. The solution to (4.4) can in principle be obtained by an exhaustive search method. However, for a network with n relays, there are at least 2^{n-1} different reasonable routes connecting the source and destination. This exponential growth makes exhaustive search unrealistic in practice if the network has a moderate to large number of relay nodes. More importantly, an exhaustive search method is not amenable to distributed implementation. In the following, Section 4.3 and Section 4.4 provide two suboptimal solutions to (4.4) that are more amenable to distributed implementation.

4.3 Approximately Ideal Path Routing (AIPR)

The motivation for our first scheme is to approximate the ideal regular path; we thus refer to this routing scheme as the approximately ideal path routing (AIPR) scheme. AIPR directly utilizes the Euclidean distance to select relays, and thus differs from the n^{th} -nearest-neighbor routing schemes [33].

For a given source and destination, [62] suggests that in a regular linear network there is an optimum number of hops n_{opt} . More specifically, the number of links in an optimal *regular* linear network satisfies [62]

$$n_{opt}R \approx \frac{\alpha + \mathcal{W}(-\alpha e^{-\alpha})}{\ln 2}, \quad (4.5)$$

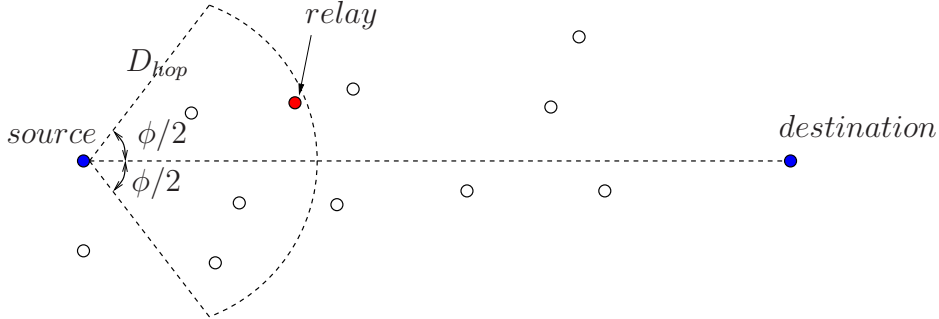


Figure 4.1. Illustration of the first step in AIPR.

where R is the path spectral efficiency, and $\mathcal{W}(\cdot)$ is the principal branch of the Lambert W function [17]. Combining (2.3), (4.3) and (4.5), we obtain the number of hops in an optimal regular linear network given the network SNR ρ :

$$n_{opt} \approx \left[\left(\frac{2^{[\alpha + \mathcal{W}(-\alpha e^{-\alpha})]/\ln 2} - 1}{\rho} \right)^{1/\alpha} \right]_+, \quad (4.6)$$

where $[\cdot]_+$ rounds the operand to the nearest positive integer. Assuming that the distance between the source and destination $D_{s,d}$ and the network SNR are known, the optimum inter-relay distance D_{hop} is

$$D_{hop} = D_{s,d}/n_{opt}. \quad (4.7)$$

Thus, an optimum regular linear network connecting node s and d consists of n_{opt} hops with a per-hop distance D_{hop} .

The above discussion applies to networks in which both the number and locations of relays can be designed. However, such a regular linear path with an ideal inter-relay distance most likely will not exist in more general network scenarios. As an alternative, we propose the procedure as shown in Algorithm 1 to find a path approximating this ideal path given that location information is available. The basic idea is demonstrated in Fig. 4.1. Starting from the source, we look for a relay node that lies within a distance D_{hop} and in the right direction to the destination. A

transmit node is assumed to know the destination location in order to proceed in the right direction and to know location information for at least neighboring nodes. To prevent the path from going in the wrong direction in the two-dimensional plane [33], the search for the relay is limited inside a sector originated from the transmitter with a radius D_{hop} and with an angle $\phi/2, 0 \leq \phi \leq \pi$ of the axis to the destination as suggested in [33]. After a transmit node choose its relay node, the relay node obtains the value of D_{hop} from the transmitter and repeats the process of finding its next relay node until the destination is reached. If at any stage no such node is available, Algorithm 1 increases the per-hop distance D_{hop} by δ . Only the source needs to know $D_{s,d}$ to compute D_{hop} .

Algorithm 1 AIPR

- 1: {Node s and d denotes the source and destination respectively; $Path[0...N]$ is a sufficiently large array with initial value $NULL$ and will contain the path; $sect(a, b, r, \phi/2)$ denotes the sector originated from node a , with a radius r and an angle $\phi/2$ of the axis from the node a to b ; δ is the given step in increasing the inter-relay distance D .}
 - 2: calculate D_{hop} based on (4.7);
 - 3: $\hat{s} = s$; $\hat{d} = NULL$; $Path[0] = s$; $n = 0$; $D = D_{hop}$;
 - 4: **while** $\hat{d} \neq d$ **do**
 - 5: **if** No relay node lies in $sect(\hat{s}, d, D, \phi/2)$ **then**
 - 6: $D = D + \delta$;
 - 7: **else if** node d lies in $sect(\hat{s}, d, D, \phi/2)$ **then**
 - 8: $\hat{d} = d$; $Path[n++] = \hat{d}$;
 - 9: **else**
 - 10: choose the relay node in $sect(\hat{s}, d, D, \phi/2)$ with the shortest distance to the destination, denote it as node \hat{d} and $Path[n++] = \hat{d}$;
 - 11: **end if**
 - 12: **end while**
-

4.4 Distributed Spectrum-Efficient Routing (DSER)

The motivation for our second routing scheme, namely, the *distributed spectrum-efficient routing* (DSER) scheme, is to balance the trade-off between the power efficiency and bandwidth efficiency, thus improving the spectral efficiency. More

specifically, the discussion in Section 4.2 suggests that longer per-hop distances might result in power inefficiency, and shorter per-hop distances might result in bandwidth inefficiency. Thus, there is both a penalty and a reward, in terms of spectral efficiency, with the addition of intermediate relay links. This motivates us to solve the following problem for a spectrum-efficient route:

$$\min_{L:t(L)=s,r(L)=d} \sum_{l \in L} 1 + \frac{\beta}{\rho_l}, \quad (4.8)$$

where, as before, nodes s and d form the desired source-destination pair, and $\beta \geq 0$, referred to as the *routing coefficient*, is a parameter that can be designed. Intuitively, the additive constant 1 represents the penalty on bandwidth efficiency for additional hops; the factor $1/\rho_l$ characterizes SNR gains by using links with short distances; and the parameter β weights the impact of power and bandwidth. A routing algorithm can use $1 + \beta/\rho_l$ as the link metric and use distributed Bellman-Ford or Dijkstra's algorithms to solve (4.8). As we will see, this routing scheme can offer significant gains in spectral efficiency compared to nearest-neighbor routing or single-hop. The DSER scheme does not depend on the particular path-loss model in (2.3). In practice, the link SNR can be directly measured by received signal strength indicators (RSSI) available on most devices and fed back to the transmitters.

4.4.1 Values of the Routing Coefficient

To determine the routing coefficient β , we note that (4.6) provides the optimum number of hops n_{opt} for the design of a regular linear network. Now, if we assume that DSER is used to design a regular linear network connecting a particular source-destination pair with a unit distance and SNR ρ , the objective function to be minimized becomes

$$f(|L|) = |L| \left[1 + \frac{\beta |L|^{-\alpha}}{\rho} \right]. \quad (4.9)$$

We relax $|L|$ as a real number, differentiate (4.9) with respect to $|L|$ and set $df(|L|)/d|L| = 0$ to obtain an expression for the optimum number of links $|L|_{opt}$. By setting $|L|_{opt} = n_{opt}$, we obtain

$$\beta = \frac{e^{\alpha + \mathcal{W}(-\alpha e^{-\alpha})} - 1}{\alpha - 1}. \quad (4.10)$$

The routing coefficient determined by (4.10) is independent of the network SNR and can be determined by the channel model. Furthermore, in the range $1 < \alpha \leq 5$, (4.10) can be very accurately approximated as

$$\beta \approx 2^\alpha. \quad (4.11)$$

In Section 4.6 we present simulation results to show that DSER performs quite well using these approximations. It can be observed from (4.11) that the value of routing coefficient increases drastically as the path loss exponent increases. This suggests that the SNR gain of shorter hops is assigned a higher weight as the path loss exponent increases. As a result, DSER favors a route with a shorter per-hop distance to combat the path loss when the path loss exponent is large.

We note that (4.10) is developed essentially assuming there are an infinite number of nodes and a continuum of locations from which to choose. Moreover, our derivation has not fully taken into account the effect of modulation, coding, queuing, and so forth. Therefore, for an arbitrary network with a finite number of nodes and practical communication schemes, the value of β can be further tuned, *e.g.*, for a specific route geometry, network SNR, modulation format, and so forth, to improve the spectral efficiency of the DSER scheme.

4.4.2 Properties

From (4.8), it is straightforward to see that for a given network, the route generated by DSER depends on the link SNRs. In the high SNR regime, $\beta/\rho_l \ll 1$, *i.e.*,

the cost of sharing bandwidth among many links outweighs the SNR gains of shorter inter-relay distances. Thus, the DSER route will approach single-hop between the source and destination in this regime. In the low SNR regime or the high path loss exponent regime, $\beta/\rho_l \gg 1$, *i.e.*, the SNR gains of shorter links outweigh the cost of sharing bandwidth. In such scenarios, the performance of DSER will approach that of nearest-neighbor routing. The discussion here agrees with simulation results we will present in Section 4.6.

For the DSER scheme, the weight of a path L is $W(L) = \sum_{l \in L} 1 + \beta/\rho_l$. It can be easily verified that the DSER metric is strictly isotonic and monotonic. As Chapter 2 summarizes, for link-state routing protocols, isotonicity of the path weight function is a necessary and sufficient condition for a generalized Dijkstra's algorithm to yield optimal paths. If the path weight function satisfies strict isotonicity, forwarding decisions can be based only on independent local computation, and the resulting path is loop free. For distance vector routing protocols, monotonicity of the path weight function implies protocol convergence in every network, and isotonicity assures convergence of algorithms to optimal paths [64]. Therefore, the DSER scheme can be implemented in existing networks with link-state or distance vector routing protocols. Also, the path metric of the DSER scheme is additive, meeting a standard assumption of most existing implementations of Bellman-Ford or Dijkstra's algorithms [18]. Thus, DSER can be implemented on top of existing wireless network routing protocols such as DSR and AODV [37, 20]. By contrast, AIPR is not as easy to incorporate into existing routing protocols. However, as we will see, AIPR offers certain advantages in low SNR regimes, thus is a good alternative for routing in wireless sensor networks, where location information can be available to sensor nodes.

4.5 Extensions

4.5.1 Spatial Reuse

AIPR and DSER have so far been developed without taking into account the effect of spatial reuse of bandwidth, *i.e.*, without considering interference. However, it is worth noting that the condition (4.5), which guides our design of AIPR and DSER, turns out to be equivalent to, up to a factor of 2, the condition for maximizing the intensity of transmission in an interference-limited network [38]. In the context of [38], n_{opt} can be viewed as the number of orthogonal sub-bands and R as the required spectral efficiency on each link.

As indicated in Section 4.1, joint design of routing and scheduling can be difficult. For the purpose of illustrating that our routing schemes can benefit from spatial reuse, it suffices to consider a separate design approach: apply the routing scheme assuming no interference to obtain a route, and then apply a scheduling algorithm to the selected route. In particular, we consider modulo- K scheduling [62], also called K -phase TDMA [76]: two links $l_i, l_j \in L$ can use the same time slot if $(i - j) \bmod K = 0$ where \bmod is the modulo operation. Note that we assume the transmission is scheduled in the right order. The idea of modulo- K scheduling is to limit the co-channel interference while reusing wireless resources spatially. For each route, we choose an optimum K that maximizes the path spectral efficiency. Even though allowing nodes to transmit with different levels of power might improve the efficiency of networks via power control, we only consider the constant transmit power assumption as argued in Section 2.1. Note that both modulo- K scheduling and constant transmit power assumption are not optimal in general, but they suffice to show that spatial reuse can improve the spectral efficiency of the DSER scheme. In Section 4.6, we present simulation results showing that at low SNR the spectral efficiency of DSER with modulo- K scheduling is larger than without spatial reuse.

Note that once a path L is determined, the spectral efficiency of the path with a modulo- K scheduling is given by

$$R_L = \min_{l \in L} \frac{1}{K} \log(1 + \gamma_l), \quad (4.12)$$

where K is the number of time slots needed for scheduling and γ_l is the signal-to-interference-and-noise ratio (SINR) of link $l \in L$, *i.e.*,

$$\gamma_l = \frac{\rho_l}{1 + \sum_{\{l_i: l_i \in L, l_i \neq l, \tau(l_i) = \tau(l)\}} \rho G_{t(l_i), r(l)}}, \quad (4.13)$$

with $\tau(l)$ denoting the time slot used by link l .

4.5.2 Relation of DSER to Other Protocols

It turns out that DSER is related to several widely known routing metrics. As we will show in the sequel, by adjusting the routing coefficient β , the DSER metric specializes to the minimal hop-count or to the expected transmission count (ETX) routing metric. These connections demonstrate the robustness of DSER and also provide a different interpretation of DSER.

When $\beta = 0$, DSER falls back to minimal hop-count routing. As demonstrated in [24], minimal hop-count routing is very robust and provides good performance when network devices are highly mobile. Thus, even though DSER is developed assuming a relatively static network, it can still apply to a highly mobile network by choosing $\beta = 0$.

With proper choice of β , the DSER metric can also approximate the ETX routing metric [20], a well-known metric for improving the throughput of wireless networks. To illustrate, we consider a network with an independently identical distributed (i.i.d) block Rayleigh fading model for each channel. Signals also suffer path loss as described in Section 2.1. The fading coefficients are complex Gaussian random variables with zero mean and unit variance. Each link l has a desired link data rate

R and uses automatic repeat-request (ARQ) until the message is correctly received. Denoting the packet error rate for link l as P_l^e , the average number of transmissions for a packet on a link is $1/(1 - P_l^e)$. To minimize the end-to-end delay, the ETX scheme proposed in [20] aims to minimize the expected total number of packet transmissions, *i.e.*,

$$\min_{L:t(L)=s,r(L)=d} \sum_{l \in L} \frac{1}{1 - P_l^e}. \quad (4.14)$$

In terms of diversity-multiplexing trade-off, [81] shows that the main error event causing packet losses is outage. Thus, we approximate the packet loss rate by the outage probability,

$$P_l^e = 1 - \exp \left[-\frac{2^R - 1}{\rho_l} \right]. \quad (4.15)$$

Substituting (4.15) into (4.14), and making use the approximation $e^x \approx 1 + x$ for small x , *i.e.*, small R or large ρ_l , (4.14) becomes

$$\min_{L:t(L)=s,r(L)=d} \sum_{l \in L} 1 + \frac{2^R - 1}{\rho_l}, \quad (4.16)$$

which is the same as (4.8) with

$$\beta = 2^R - 1. \quad (4.17)$$

Thus, with a proper choice of β , the DSER metric approximates the ETX routing metric for fading channels at high SNR. Comparing (4.17) with (4.10), we note that for ETX in the fading channel, the per-link data rate R assumes the role that α held in (4.10). As the per-link data rate requirement R increases, β increases, suggesting that the SNR gain provided by a shorter per-hop distance becomes more important in guaranteeing reliability.

The approximation of ETX by DSER offers another interpretation of (4.8) as a routing scheme to find a route that maximizes the reliability. When a path contains a large number of hops, the SNR gain of shorter inter-relay distance results in a smaller

error rate for each link along the path. However, as the number of links increases, the probability that all links are decoding correctly might actually decrease. From this perspective, the constant term 1 captures the requirement that all links are functioning, and the factor $1/\rho_l$ characterizes the gain in smaller individual link error probability by using shorter hops.

4.5.3 Finite Input Alphabet

So far, we have assumed Gaussian inputs, corresponding to an alphabet with infinite size. In practice, the input alphabet size is usually finite. The size of alphabet limits the maximum spectral efficiency as we will see in the sequel. Hence, it is of practical interest to see how our proposed routing schemes perform with a finite input alphabet.

We assume the input alphabet of $\mathcal{X} = \{a_1, a_2, \dots, a_M\}$, where M is the size of alphabet. We assume the input is of equal amplitude; hence the SNR for an AWGN channel with finite input is defined as $\rho = |a_1|^2/N_0$. As the input is of equal amplitude, we can further assume equiprobable inputs. A more general finite alphabet and a more general distribution of input letters might further increase the capacity. However, our assumptions are sufficient to demonstrate the effect of finite input alphabets. It can be shown [28, 69] that the channel capacity for a AWGN channel under the above constraints is

$$C_N = \log M - \frac{1}{M} \sum_{i=1}^M E \left[\log \sum_{j=1}^M \exp\left\{-\frac{|z + a_i - a_j|^2 - |z|^2}{N_0}\right\} \right], \quad (4.18)$$

where the expectation $E[\]$ is with respect to the Gaussian random variable z of zero mean and N_0 variance. Unfortunately, the evaluation of (4.18) turns out to be difficult for a general input size M . Hence, we will rely on Monte-Carlo simulation to evaluate (4.18). Fig. 4.2 demonstrates the impact of finite input size on a single input single output (SISO) AWGN channel. From Fig. 4.2, it is straightforward to see that

the constraint of finite input alphabet size adds another trade-off to the problem, *i. e.*, instead of power and bandwidth, the size of alphabet could also become the main constraint in limiting spectral efficiency. The impact of the constraint of alphabet size in multihop networks can be further seen in Fig. 4.3 and Fig. 4.4, which show the spectral efficiency as a function of SNR for linear regular multihop networks with Gaussian and BPSK input, respectively. From Fig. 4.4, we can observe that the finite alphabet size often becomes the dominant constraint in limiting the rates of multihop networks. The impact of the constraint of alphabet size in multihop networks is even more obvious if we consider a curve that represents the supremum of all curves with different number of hops. Comparing Fig. 4.3 and Fig. 4.4, we can see that the supremum curve for a small size alphabet has much more drastic changes in slope compared to that of an infinite size alphabet. As a result, we can expect the performance curves of routing schemes for a finite size alphabet to become less smooth compared to that with Gaussian inputs. Also, the supremum curve of finite size alphabets is upper bounded by $\log M$.

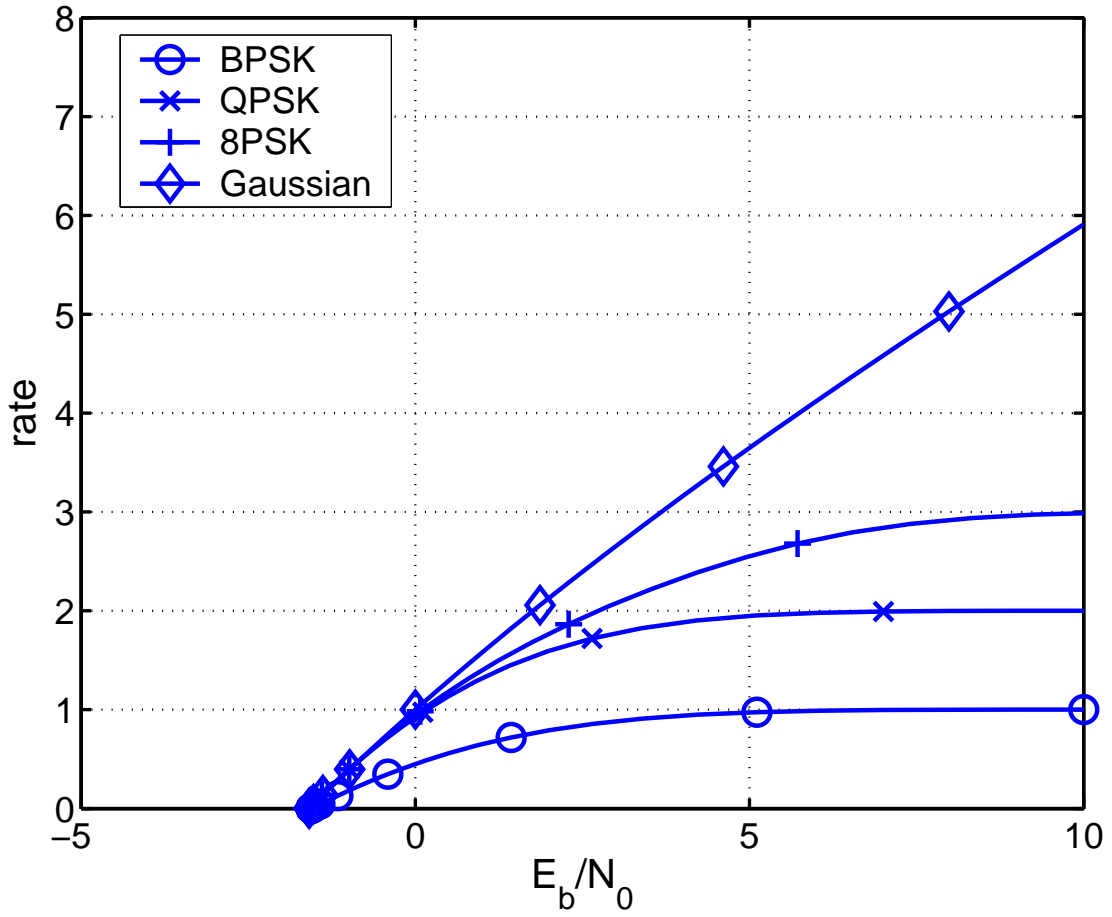


Figure 4.2. Spectral efficiency of a single-input single-output AWGN channel with finite size alphabets

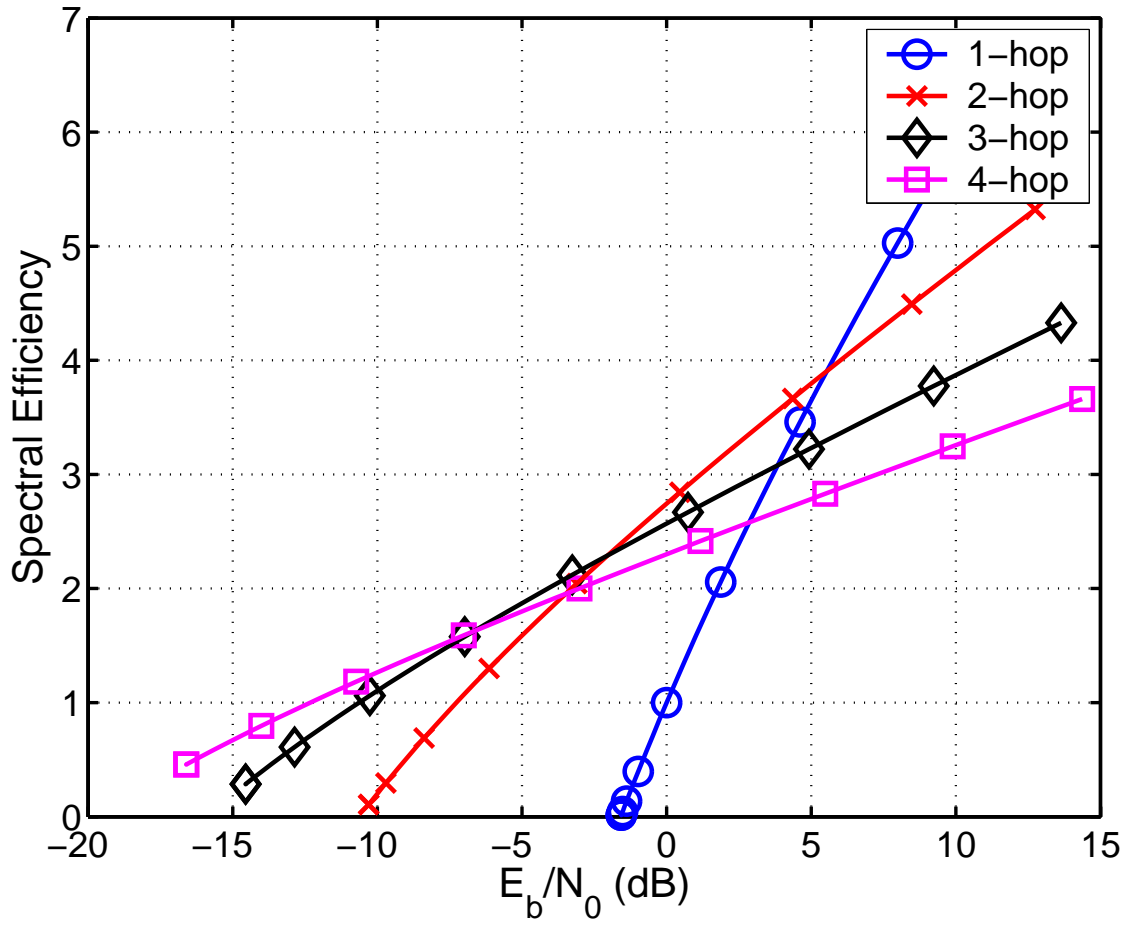


Figure 4.3. Spectral efficiency of linear regular multihop networks with Gaussian input

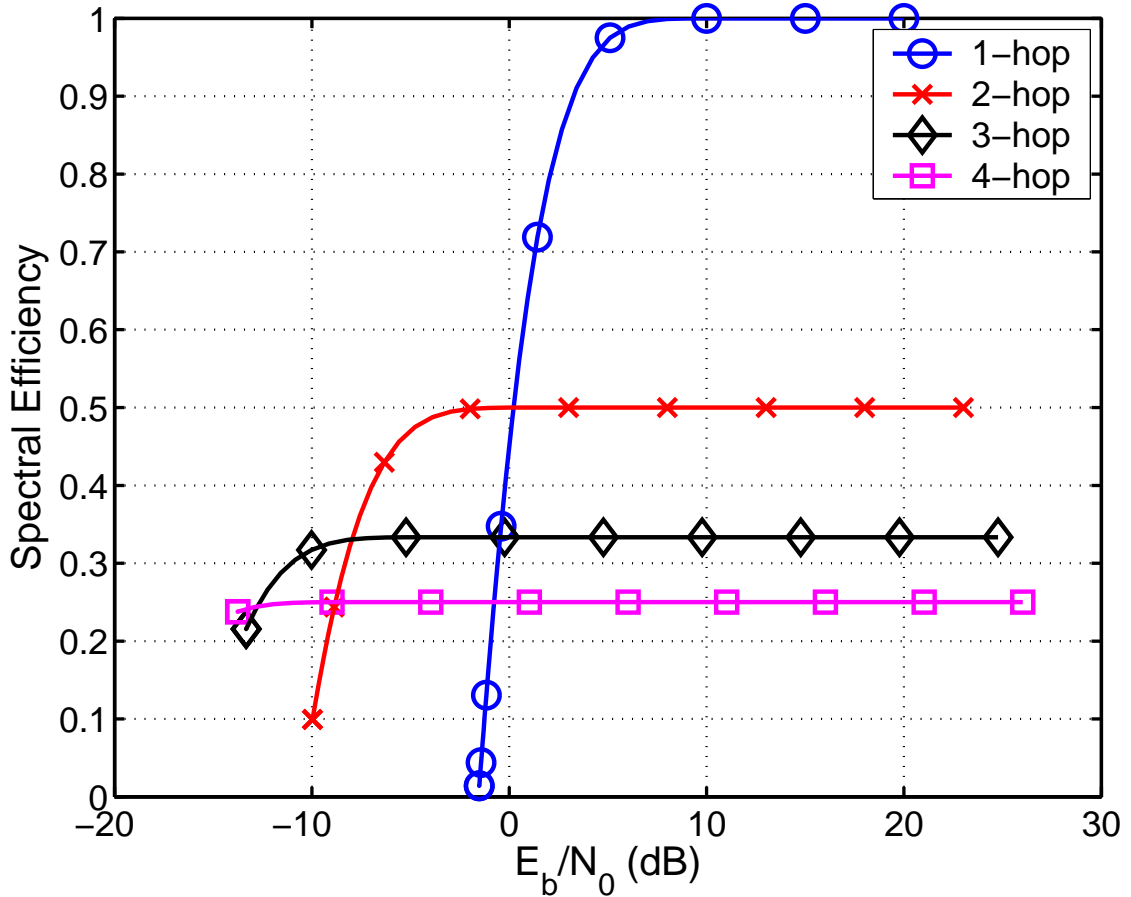


Figure 4.4. Spectral efficiency of linear regular multihop networks with BPSK input

4.6 Simulation Results

This section presents simulation results to compare spectral efficiencies of different routing schemes averaged over random network realizations. Our simulations focus on uniformly random networks. For a one-dimensional linear network, we assume the source and destination are located at coordinates $(0, 0)$ and $(1, 0)$, respectively, and the horizontal coordinates of intermediate relay nodes are independent random variables uniformly distributed between 0 and 1. For a two-dimensional network, we assume the source and destination are located at $(0, 0)$ and $(1, 1)$, re-

spectively. The horizontal and vertical coordinates of the potential relay nodes are independent random variables uniformly distributed between 0 and 1. To estimate average spectral efficiency over the ensemble of random networks, we take the mean over 10^4 network realizations. In our simulations, the boundaries of the 95% confidence interval are within $\pm 2\%$ of the average value, assuming the spectral efficiency of a routing scheme is Gaussian distributed. Thus, the confidence interval is sufficiently-small, allowing us to compare routing schemes using these simulation statistics.

We also assume a path-loss model described in (2.2), taking the path loss exponent $\alpha = 4$ and the far-field distance $D_f = 10^{-3}$ unless specified otherwise. Motivated by the approximation $\beta \approx 2^\alpha$ in Section 4.4, the routing coefficient is taken to be $\beta = 16$.

4.6.1 Gaussian Inputs

This section presents simulation results with Gaussian inputs.

As two examples, Fig. 4.5 and 4.6 show the average spectral efficiency of different routing schemes including nearest-neighbor routing, single-hop, AIPR, and DSER for uniformly random linear networks with 5 and 9 nodes, respectively. To see a wider dynamical range around the low SNR regimes, the horizontal coordinate is taken as E_b/N_0 , *i.e.*, the ratio between the SNR and the average spectral efficiency. In Fig. 4.5 and 4.6, the optimal spectral efficiency is obtained by an exhaustive search method and is provided as a reference. It is clear that the performance of single-hop only approaches the optimum performance in the high SNR regime and suffers from a significant loss in spectral efficiency in the low SNR regime. The performance of nearest-neighbor routing approaches the optimal performance in the low SNR regime, but degrades in the high SNR regime due to its inefficient use

of bandwidth. By contrast, one can observe that the curves of the DSER scheme track the optimal curves throughout the whole SNR regime. In particular, in the moderate SNR regime, DSER offers significant gains in spectral efficiency relative to AIPR, nearest-neighbor routing, and single-hop. In particular, when E_b/N_0 is around 5 dB or the network SNR is around 0 dB, the spectral efficiency of the DSER scheme is twice as large as those of nearest-neighbor routing and single-hop. Thus, networks can benefit significantly in spectral efficiency from the use of DSER.

In random networks, AIPR suffers from a significant performance loss in the moderate SNR regime because it is difficult to find a regular linear path. However, AIPR performs reasonably well in either the low SNR or the high SNR regimes in our simulation. This is because at low SNR, AIPR degenerates into nearest-neighbor routing. Hence, the impact of path irregularity at low SNR is not as serious as at moderate SNR. In the high SNR regime, AIPR degenerates to choosing the direct link from the source to destination, which is the optimum route. We stress that our simulation does not fully consider the impact of fading, which might cause significant degradation in performance for AIPR. We note the curves of AIPR bend backwards around 5 dB E_b/N_0 . This is because the horizontal coordinate is E_b/N_0 , not ρ . More specifically, we simulate the average rate as a function of SNR, i.e., $R = f(\rho)$. However, what we draw is $f(\rho)$ vs $\rho/f(\rho)$ in order to see a larger dynamical range. The regime that the AIPR curve bend backwards is the regime in which the rate of AIPR grows extremely slow as SNR increases. Consider the example of linear network with 9 nodes, at $\rho = 9$ dB, the rate R is 2.5. The corresponding E_b/N_0 is 5.0dB. However, at $\rho = 10$ dB, the rate R is 3.5, which corresponds to $E_b/N_0 = 4.6$ dB. The reason that the rate of AIPR as a function of SNR grows so slow is as follows. Around this regime, the optimum number of hops is approximately 2. So, the source node always look for a relay node between it and the destination node. Now, the

relay node can be quite far away from the ideal locations, which causes degradation. In particular, around this regime, even the spectral efficiency gain of an ideal two-hop linear regular path is small relative to the single-hop scheme. Therefore, the degradation caused by the non-optimum location of the relay is particularly serious in this regime and causes the “bending backwards” behavior. This is a price to pay for doing routing distributively.

Fig. 4.7 shows the average spectral efficiencies of different routing schemes in a two-dimensional random network with 9 nodes. Note that the nearest-neighbor routing in Fig. 4.7 selects its nearest neighbor that lies within an angle $\phi/2$ of the line from the source to destination, *i.e.*, Strategy A in [33]. We choose $\phi = \pi/2$. Compared to the case of one-dimensional random networks, the performance of DSER in two-dimensional random networks degrades in the low SNR regime. This could be explained by the fact that DSER does not require its relay node to lie within angle $\phi/2$ of the line from source to destination, *i.e.*, DSER does not require location information even in two-dimensional random networks. In contrast, both nearest-neighbor routing and AIPR require location information in two-dimensional networks. Other than this difference, most other observations from one-dimensional networks carry over to two-dimensional networks. Thus, in the remainder, we will only focus on the results from the one-dimensional case with the understanding that these observations carry over to the two-dimensional networks.

Fig. 4.8 shows how DSER and AIPR adapt to different network SNRs, choosing different paths in a sample linear random network with 8 nodes. Note that based on (4.6), the optimum hop number of an optimum regular linear path for the network SNR of -20, 0 and 20dB is 8, 3 and 1, respectively. As shown in Fig. 4.8, DSER and AIPR choose paths with shorter hops when SNR is low. As the SNR increases, they tend to choose paths with longer inter-relay distance. Paths selected by AIPR

and DSER are not necessarily the same. In particular, Fig. 4.8 shows that, relative to DSER, AIPR can choose a more balanced route at low SNR due to its utilization of location information. This observation is in line with our previous observation that AIPR can provide better performance at low SNR. Together with Fig. 4.6 – 4.7, Fig. 4.8 demonstrates that DSER and AIPR adapt to changes of the network SNR as we expected.

Fig. 4.9 compares the performance of DSER with that of optimal routing with bandwidth optimization (ORBO). The spectral efficiency improves for ORBO mainly in the low SNR regime. However, as the network SNR increases, the benefit of bandwidth optimization decreases and eventually vanishes. This is because at high SNR, ORBO corresponds to single-hop, which is also the case for DSER.

Fig. 4.10 show the average spectral efficiency as a function of the path-loss exponent for uniformly random linear networks with 9 nodes and network SNR of -40 and 20 dB. It might seem counter-intuitive that the average spectral efficiency grows as the path-loss exponent increases. However, this is because our network SNR is end-to-end normalized SNR. Thus, as the path-loss exponent increases, the effective link SNRs on intermediate links increases as well. From Fig. 4.10, when the network SNR is high, the impact of different path loss exponent on routing schemes decreases. When the network SNR is small, relative to single-hop, routing schemes with multi-hop relaying benefit significantly from the high path loss exponent.

Fig. 4.11 shows that even though DSER and AIPR are proposed assuming TDMA without spatial reuse, a modulo- K scheduling can further improve their performance at low SNR. Moreover, we observe that in the high SNR regime, there is not much to gain from spatial reuse, as single-hop between the source and destination is optimal.

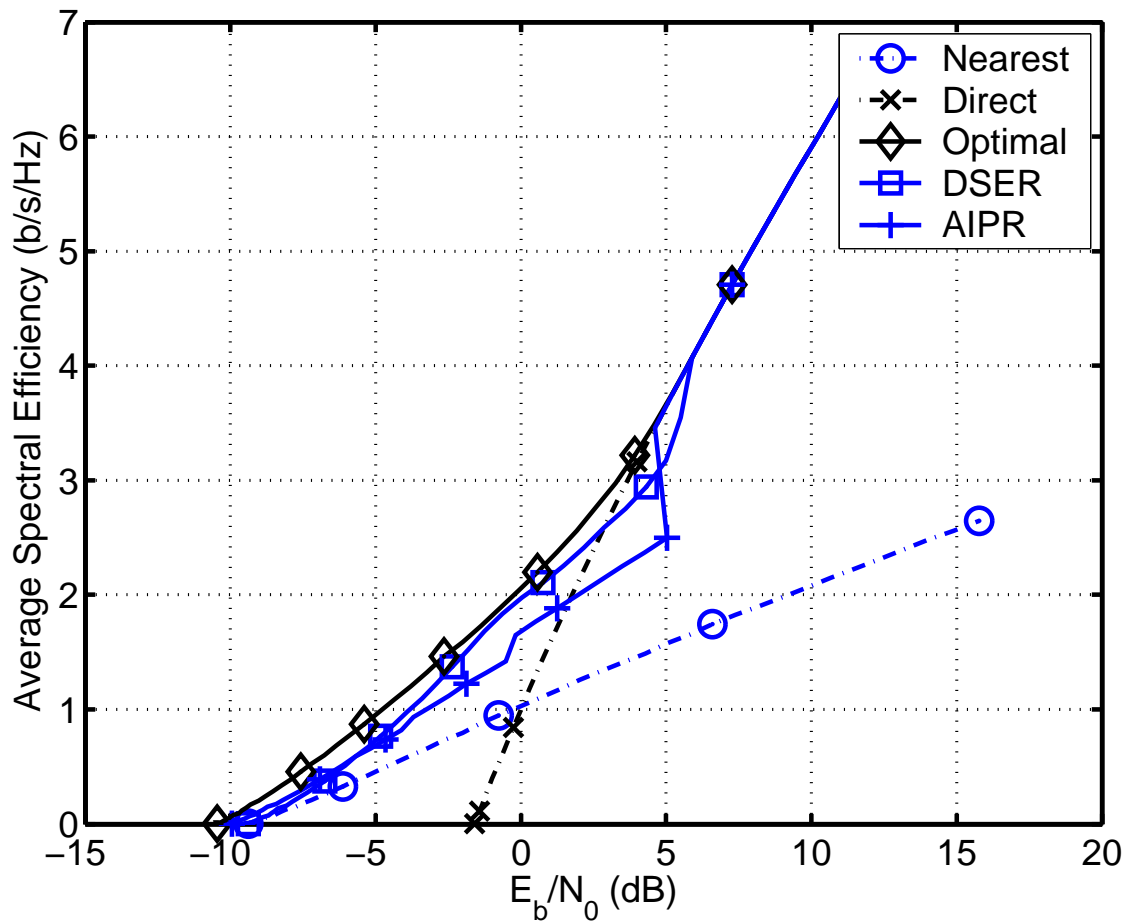


Figure 4.5. Average spectral efficiencies of different routing schemes for uniformly random linear networks with 5 nodes and $\alpha = 4$.

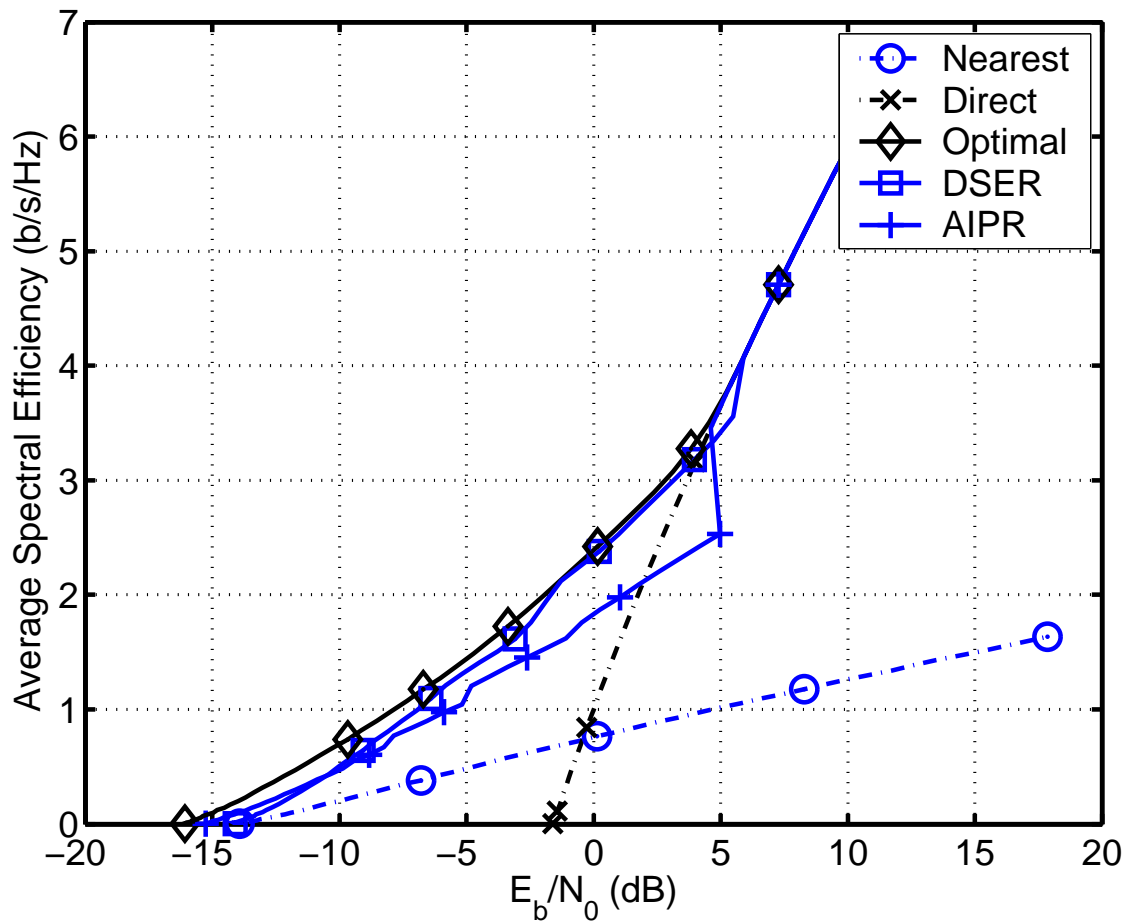


Figure 4.6. Average spectral efficiencies of different routing schemes for uniformly random linear networks with 9 nodes and $\alpha = 4$.

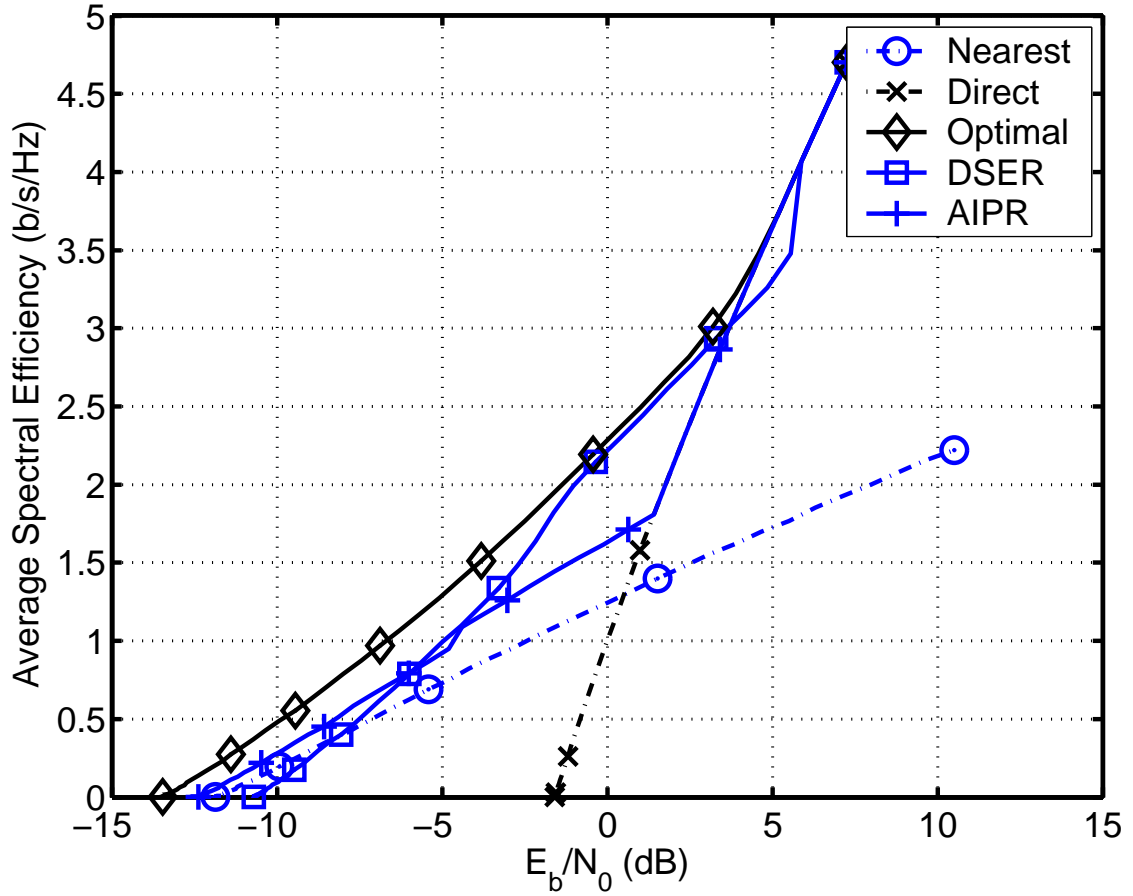


Figure 4.7. Average spectral efficiencies of different routing schemes for 2-D random networks with 9 nodes, $\alpha = 4$ and $\phi = \pi/2$.

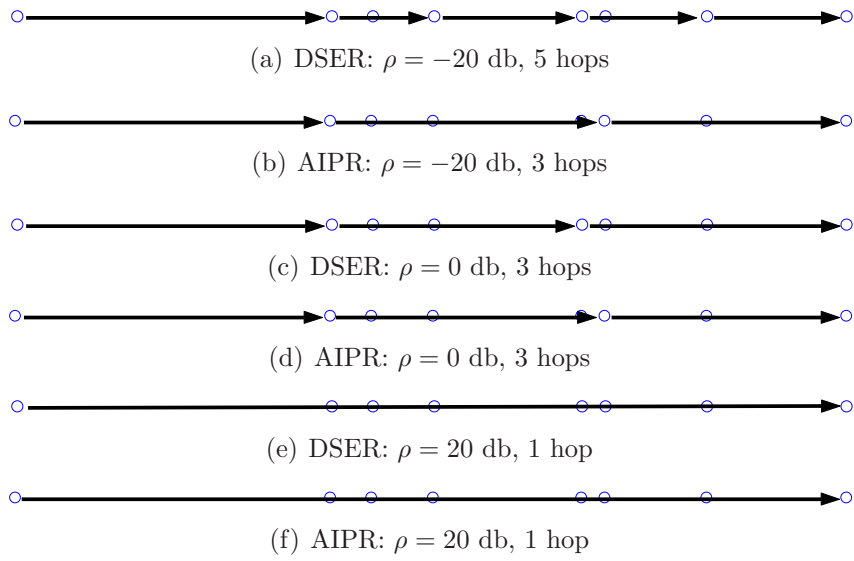


Figure 4.8. Sample DSER paths in a linear network.

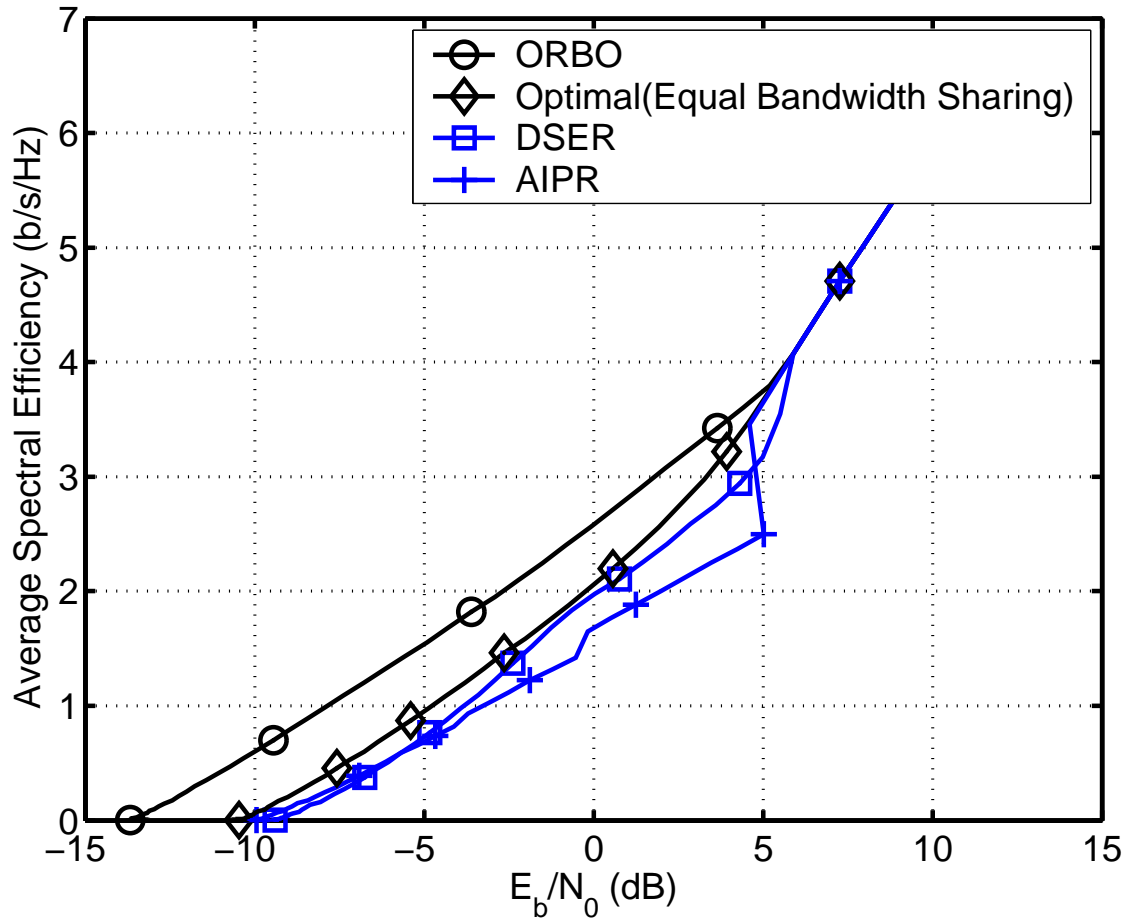


Figure 4.9. Average spectral efficiencies of the optimal routing with bandwidth optimization (ORBO) and DSER for uniformly random linear networks with 5 nodes and $\alpha = 4$.

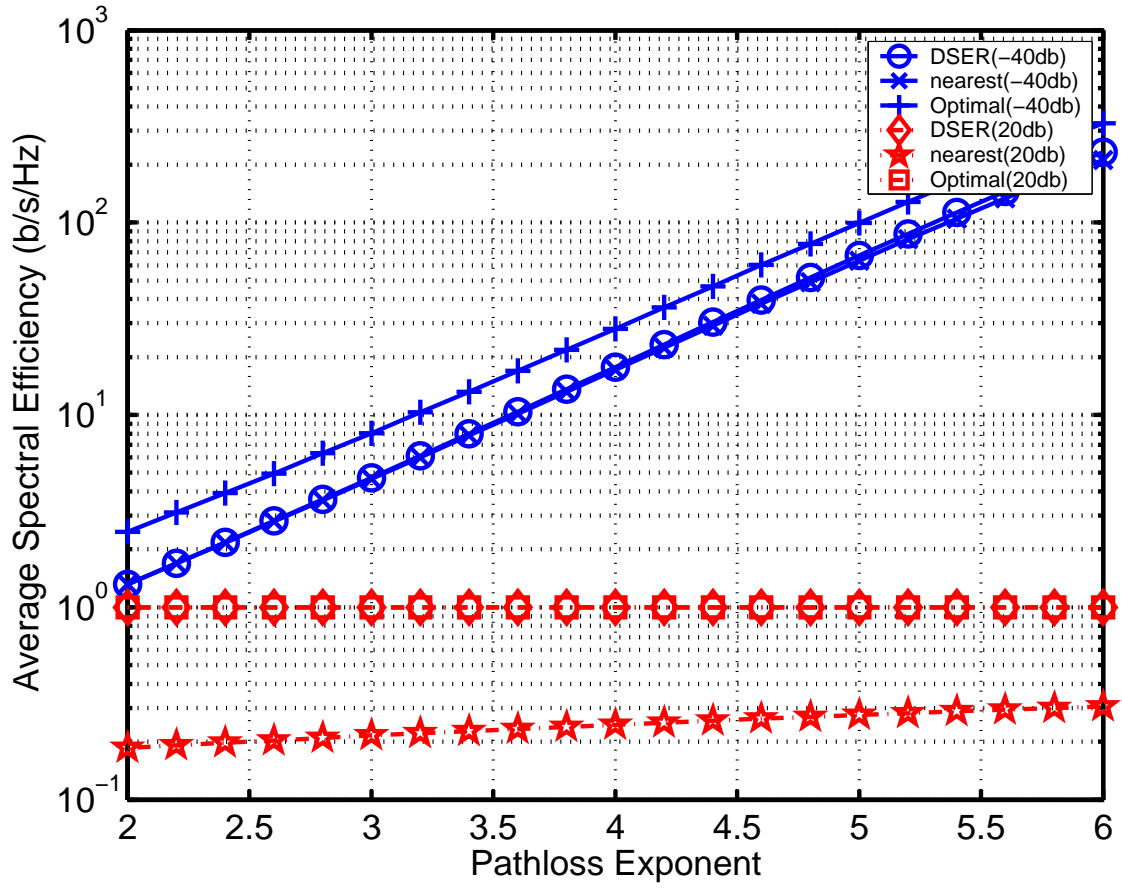


Figure 4.10. Average spectral efficiencies as a function of path-loss exponent for uniformly random linear networks with 9 nodes.

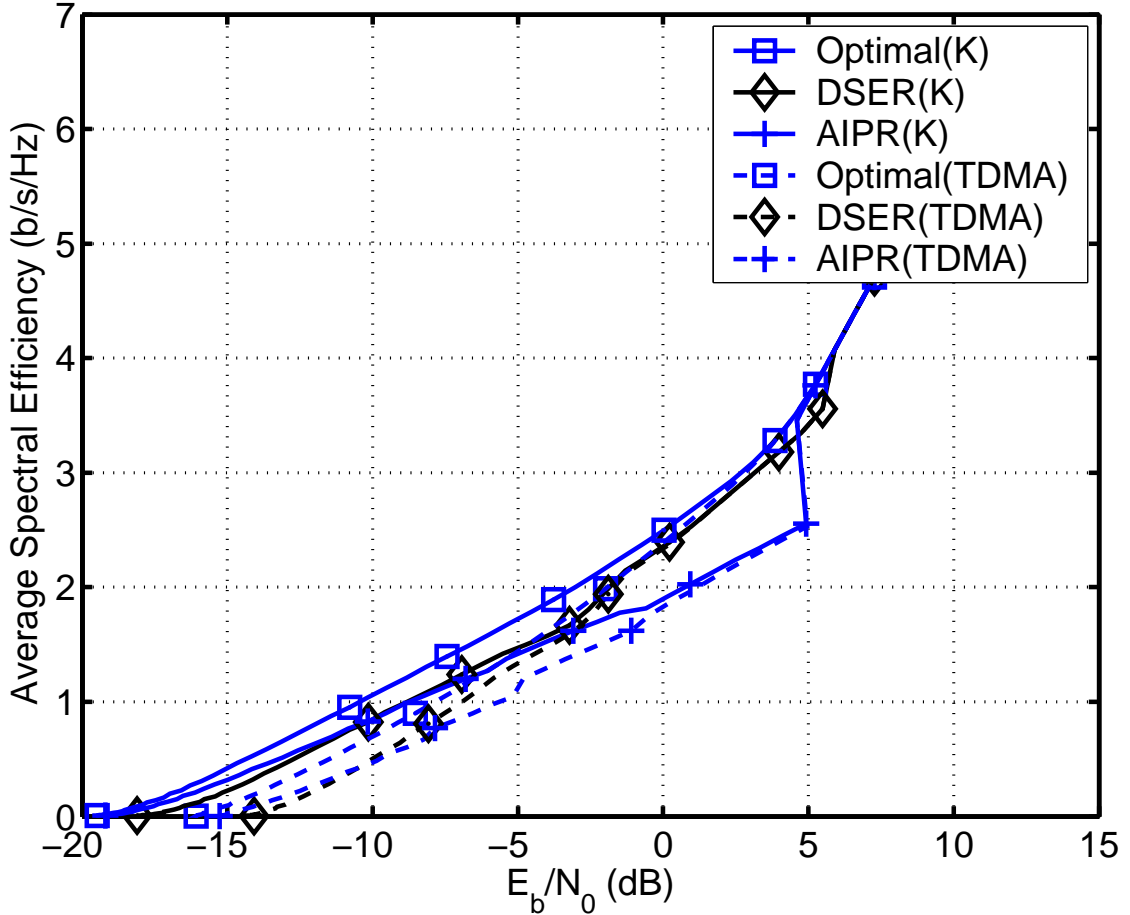


Figure 4.11. Average spectral efficiencies versus network SNR for uniformly random linear networks with 9 nodes and $\alpha = 4$. The dashed lines correspond to TDMA without spatial reuse and the solid lines correspond to modulo- K scheduling.

4.6.2 Finite Input Alphabet

We consider the impact of finite size alphabet on our selected AIPR and DSER routes by replacing the expression $\log(1 + \rho_l)$ in (4.3) with (4.18). We assume that all nodes are constrained with the same finite alphabet. This assumption models wireless transceivers, *e.g.*, Berkeley notes, that do not allow adaptive modulation. However, we note that the finite alphabet constraint often penalizes a multihop route by preventing it from fully taking advantage of link SNR improvement. A

better optimization of routing coefficient in DSER and a further modification of AIPR that takes the finite alphabet size into account might further improve the performance of our routing schemes. This direction is not further pursued in this dissertation.

Fig. 4.12 and Fig. 4.13 demonstrate the impact of BPSK on the performance of routing in uniformly random linear networks with 5 and 9 nodes, respectively. We assume that all nodes are using the same alphabet. The rate for all routes are computed based on (4.18). As we predicted before, the performance curves of all routing schemes often have drastic slope changes due to the impact of finite size alphabet on the spectral efficiency. Similar with that of Gaussian input, the performance curves of DSER and AIPR bend backwards due to representing the horizontal coordinate in E_b/N_0 . It is also straightforward to see that the minimum E_b/N_0 value at which the direct communication becomes optimal for finite input becomes smaller compared with that of AWGN input. This suggests that, due to the finite size of input alphabet, the cost of more hops becomes even more significant for low or moderate SNR. We can explain this by observing that (4.18) is upper bounded by the logarithm of the input alphabet size. Hence, even if an intermediate hop enjoys significant SNR gains due to multihop, the SNR gains along intermediate links could have very little impact on the end-to-end spectral efficiency. This also explains why the performance curves of both DSER and AIPR do not converge to the optimum curve as fast as in the AWGN input cases. DSER and AIPR assume AWGN input and choose route accordingly. As a result, DSER and AIPR still favors multiple hop routes when in fact, due to the finite alphabet size, the direct communication is already optimum. However, in contrast of the nearest-neighbor routing, DSER and AIPR eventually adapt themselves and choose the direct link as the route. A comparison of Fig. 4.12 and Fig. 4.13 also indicates that in the

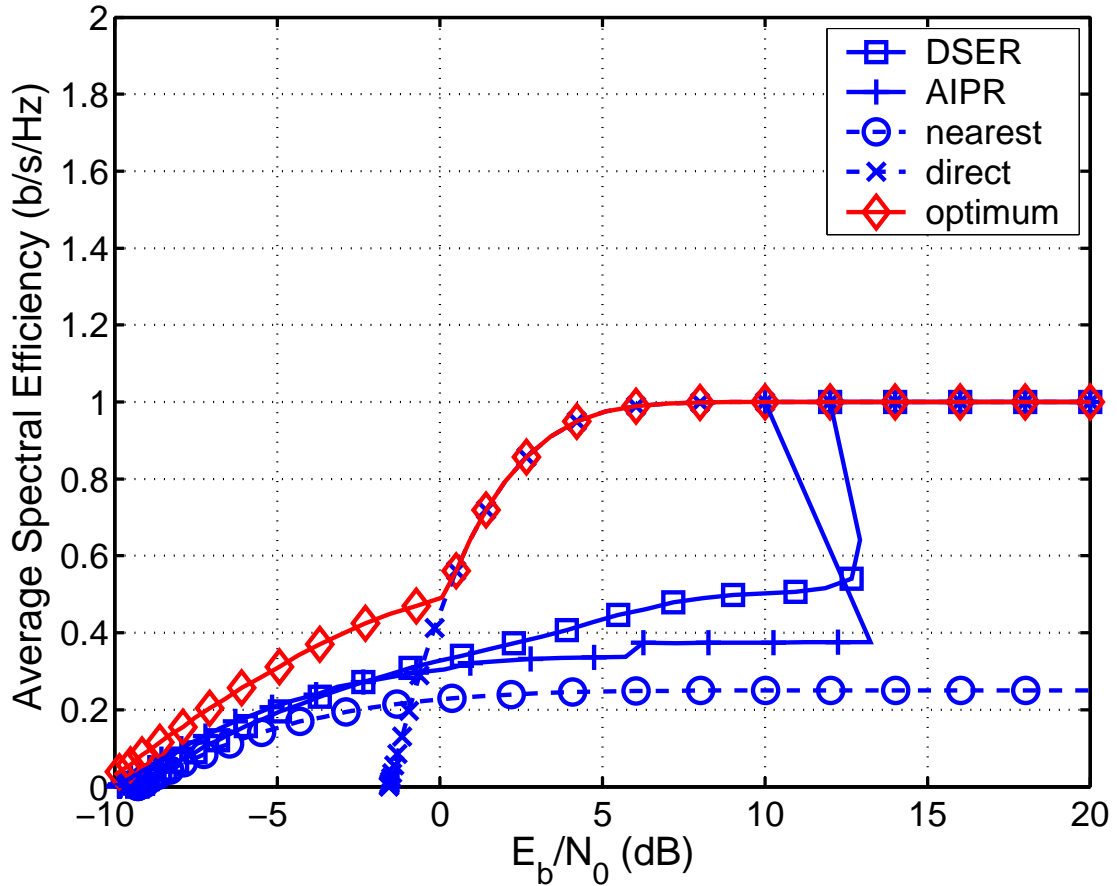


Figure 4.12. Average spectral efficiencies of different routing schemes for uniformly random linear networks with BPSK, 5 nodes and $\alpha = 4$.

low SNR regime, more hops could increase the end-to-end spectral efficiency. The minimum E_b/N_0 at which we have non-zero end-to-end spectral efficiency decreases as the size of networks grows.

Fig. 4.14 and Fig. 4.15 demonstrate the impact of 8-PSK on the performance of routing in uniformly random linear networks with 5 and 9 nodes, respectively. Not surprisingly, as the constellation size increases, the gap between the performance of our proposed routing schemes and that of optimum performance becomes smaller. Hence, a high dimensional constellation could be key to gain the benefit of adaptive routing in practical implementations.

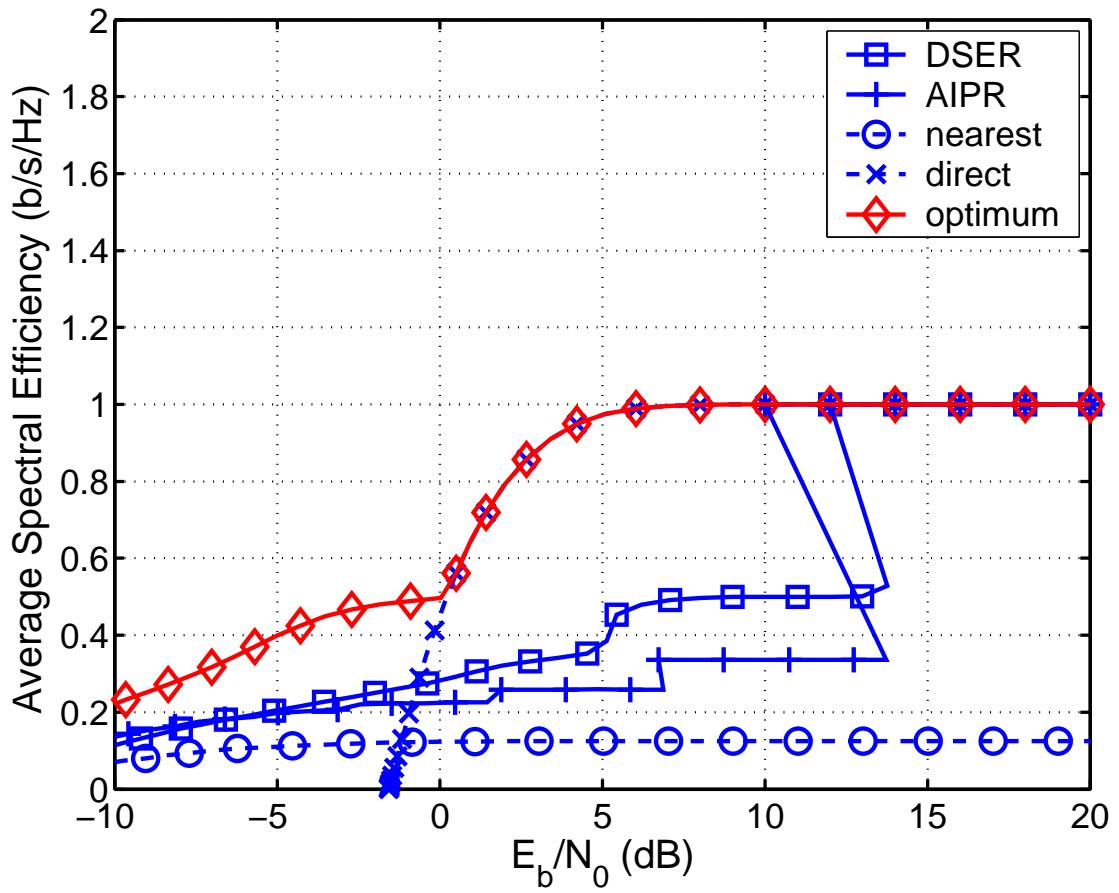


Figure 4.13. Average spectral efficiencies of different routing schemes for uniformly random linear networks with BPSK, 9 nodes and $\alpha = 4$.

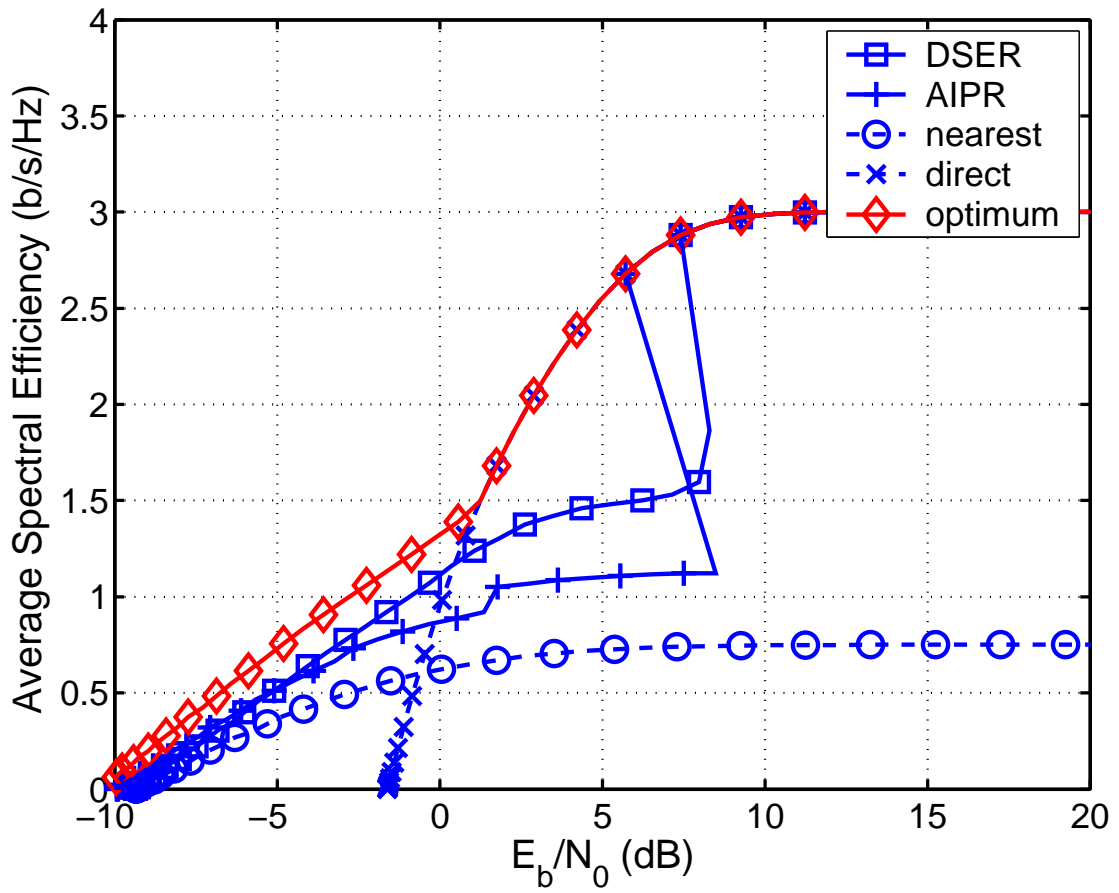


Figure 4.14. Average spectral efficiencies of different routing schemes for uniformly random linear networks with 8-PSK, 5 nodes and $\alpha = 4$.

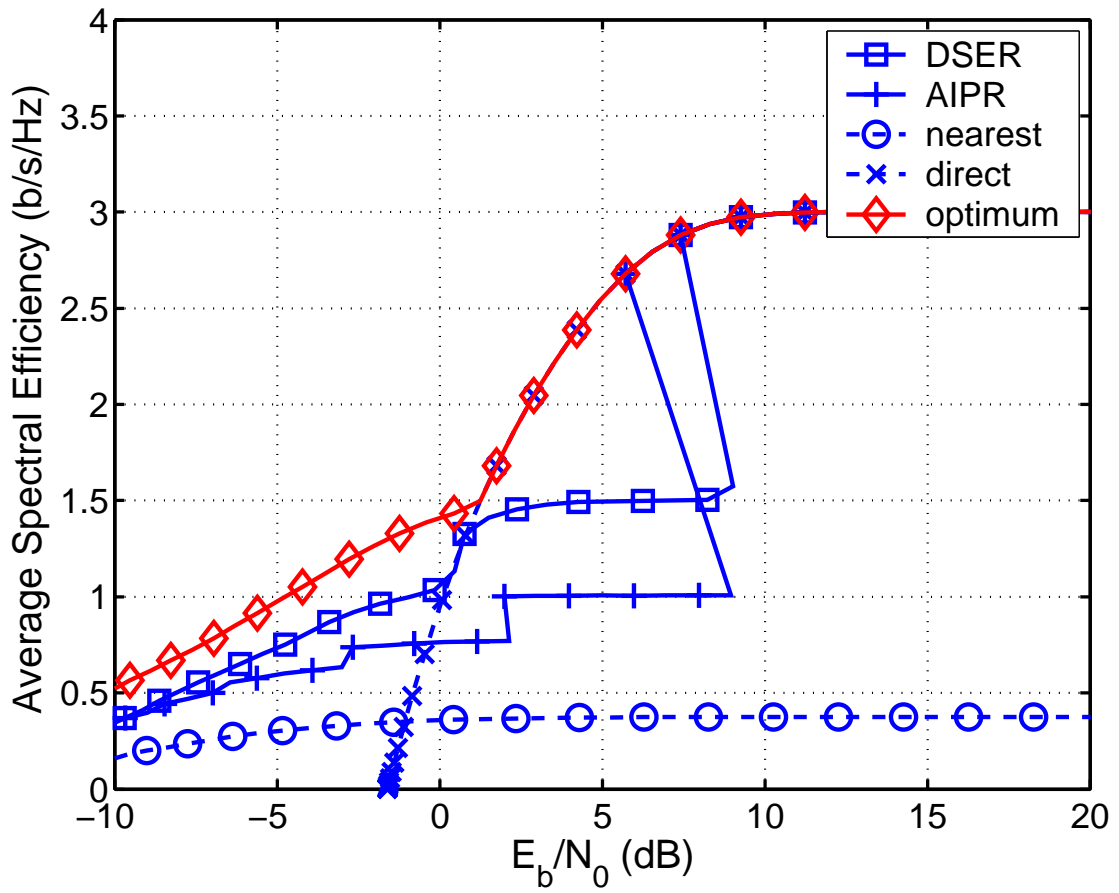


Figure 4.15. Average spectral efficiencies of different routing schemes for uniformly random linear networks with 8-PSK, 9 nodes and $\alpha = 4$.

4.7 Summary

This chapter studies end-to-end spectral efficiencies of different wireless routing schemes. This chapter's main contribution is to introduce two suboptimal solutions, namely, approximately ideal path routing (AIPR) and distributed spectrum-efficient routing (DSER), to the problem of finding routes with high spectral efficiency. AIPR is a location-assisted routing scheme. DSER can be based upon local link quality estimates, can be implemented using standard Bellman-Ford or Dijkstra's algorithms, and can be integrated into existing network protocols. Furthermore, the performance of DSER and AIPR is close to that of nearest-neighbor routing and that of minimum hop-count routing in the low and high SNR regimes, respectively. In the moderate SNR regime, DSER provides significant gains in spectral efficiency compared with both nearest-neighbor routing and minimum hop-count routing.

CHAPTER 5

RESOURCE ALLOCATION FOR BROADBAND MULTIHOP NETWORKS

In Chapter 4, we have observed that there exists a trade-off between the number of hops and the end-to-end spectral efficiency for multihop networks with AWGN channels. However, for a broadband system with frequency selective fading, when fading is slow and fading coefficient is available at the transmitter, it is well known that power allocation such as water-filling increases spectral efficiency [22]. Hence, it is not immediately clear whether the observation for narrowband system carriers over to broadband multihop networks. This chapter discusses the trade-off between the number of hops and the end-to-end spectral efficiency for broadband multihop networks by studying a resource allocation problem to maximize the end-to-end rate in a broadband multihop wireless network.

Assuming an orthogonal frequency division multiplexing (OFDM) system, we consider allocating power and subcarriers jointly according to channel state information. We focus on proposing low-complexity efficient algorithms for the multihop network with one destination, one source, and one or more relays. Simulation results suggest that the performance of our proposed algorithms closely follows the optimum performance. More importantly, our results yield that more hops help in the power limited regime, but do not help in the bandwidth limited regime, consistent with previous discussions for a narrowband system in Chapter 4.

5.1 Background

In [73], an algorithm for joint multiuser OFDM power, subcarrier and rate allocation is developed to minimize the total power consumption for a broadcast channel. In [78], the problem of assigning discrete frequency bins in a Gaussian multiple-access channel with intersymbol interference (ISI) to maximize a weighed sum rate is studied, and a practical low-complexity algorithm is proposed. Relative to [73] and [78], this chapter considers a multihop network, which adds flow conservation constraints to the problem formulation. As we will see, the addition of flow conservation constraints makes the optimization problem much more involved.

Even though [48] considers a broadband multihop wireless network, a frequency subcarrier is *shared* in a time division fashion by different hops along a path. By contrast, this chapter considers allocating subcarriers *exclusively* to different hops along the path, *i.e.*, subcarriers assigned to different hops do not overlap. Furthermore, this chapter also considers power allocation jointly with subcarrier allocation.

The remainder of the chapter is organized as follows. Section 5.2 describes the network and channel models. Section 5.3 formulates the joint power and subcarrier allocation problem for wireless networks with one destination and multiple sources and relays. Not surprisingly, it turns out that this problem is a combinatorial optimization problem. To gain some insight on the optimum solutions, we relax the problem to a convex optimization problem by adjusting the requirement on the subcarrier allocation. The relaxed problem provides an upper bound to the solution of the original problem. Moreover, Section 5.4 discusses Karush-Kuhn-Tucker (KKT) conditions for optimal solutions to the relaxed problem and provides important insights that motivate low-complexity algorithms for solving the original problem in Section 5.5. Specifically, Section 5.5.1 proposes a high-SNR approximation algorithm based on the observation that the optimal frequency allocation in a two-hop

network has a two-band structure assuming high SNR. For a network with more than two hops, a low complexity greedy algorithm is proposed in Section 5.5.2. Section 5.6 provides numerical simulation results and Section 5.7 concludes the chapter.

5.2 System Model

In this chapter, we formulate a power and subcarrier allocation problem for a wireless network with one base station (BS) and multiple sources and relays. A special case of this type of network is the multihop network of one source, one destination, and one or multiple relays. We assume relay stations do not have their own information to send and the BS is the only information sink for all sources. Following Chapter 4, we represent the network as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the set of nodes \mathcal{V} represents stations in the network and the set of links \mathcal{E} represents potential communication between two stations in the network. For each link $e \in \mathcal{E}$, we use $t(e)$ to represent the transmit end of the link and $r(e)$ to represent the receive end. We denote by \mathcal{S} the set of nodes with information sources and by \mathcal{R} the set of relay stations. We assume that sources do not relay information, *i.e.*, $\mathcal{S} \cap \mathcal{R} = \emptyset$. The sink is denoted d . Hence, $t(e) \neq d, \forall e \in \mathcal{E}$ and $d \notin \mathcal{R} \cup \mathcal{S}$.

We consider an OFDM-type system in which the whole frequency band is split into multiple subcarriers to share among different transmissions. We neglect the impact of cyclic prefix on rates. We denote by \mathcal{K} the set of available frequency subcarriers. The bandwidth of a subcarrier $k \in \mathcal{K}$ is w_k Hz. Note that in practice, the bandwidth of all subcarriers is often identical. We denote by $h_{l,k}, l \in \mathcal{E}, k \in \mathcal{K}$, the fading coefficient of the k th subcarrier for link l . We assume channel gains of all links are available to a central scheduler to allow for subcarriers allocation. We also assume channel gains are available for transmitting nodes to allocate power.

Following [44], we assume that the nodes cannot transmit and receive at the

same time. Furthermore, we consider no frequency reuse inside the cell, *i.e.*, in the network we study, any part of the bandwidth cannot be assigned to more than one link.

Each transmit node $i \in \mathcal{V}$ has a total power constraint of ρ_i , *i.e.*,

$$\sum_{\{l:t(l)=i\}} \sum_{k \in \mathcal{K}} \rho_{l,k} \leq \rho_i, \quad (5.1)$$

where $\rho_{l,k}$ denotes the allocated power on link $l \in \mathcal{E}$ at subcarrier $k \in \mathcal{K}$.

For simplicity of presentation, we assume the one-sided noise power spectral density level N_0 is equal to unity for all subcarriers and all nodes.

5.3 Problem Formulation

The goal of the optimization is to maximize the total information received by the sink from the sources subjected to transmit power and subcarrier constraints. Furthermore, we impose a flow conservation constraint to reflect the fact that the rate of a multihop route is limited by the rate of the bottleneck link, *i.e.*, the link that has the minimal rate among all intermediate hops [62, 49]. As we will see in the sequel, the problem of allocating power and subcarriers optimally becomes much more involved due to this constraint.

To find out optimal allocation of power and subcarriers, we formulate the following optimization problem:

$$C^* = \max \sum_{k \in \mathcal{K}} \sum_{\{l:t(l) \in \mathcal{S}\}} C_{l,k}, \quad (5.2)$$

subjected to the following constraints:

$$C_{l,k} \leq w_{l,k} \log(1 + |h_{l,k}|^2 \rho_{l,k} / w_{l,k}), \forall l \in \mathcal{E}, \forall k \in \mathcal{K}, \quad (5.3a)$$

$$0 \leq C_{l,k}, \forall l \in \mathcal{E}, \forall k \in \mathcal{K}, \quad (5.3b)$$

$$0 \leq w_{l,k} \leq w_k, \forall k \in \mathcal{K}, \quad (5.3c)$$

$$w_{l,k}w_{j,k} = 0, l \neq j, \forall k \in \mathcal{K}, \quad (5.3d)$$

$$\sum_{\{l:t(l)=i\}} \sum_{k \in \mathcal{K}} \rho_{l,k} \leq \rho_i, \forall i \in \mathcal{V}, \quad (5.3e)$$

$$0 \leq \rho_{l,k}, \forall l \in \mathcal{E}, \forall k \in \mathcal{K}, \quad (5.3f)$$

$$\sum_{k \in \mathcal{K}} \sum_{\{l:t(l)=i\}} C_{l,k} \geq \sum_{k \in \mathcal{K}} \sum_{\{l:r(l)=i\}} C_{l,k}, \forall i \in \mathcal{R}, \quad (5.3g)$$

where: $C_{l,k}$ is the information rate transmitted on link l at subcarrier k ; $w_{l,k}$ denotes whether the subcarrier k is used on link l ; and $\rho_{l,k}$ denotes the transmit power of link l at subcarrier k . The constraint (5.3g) captures flow conservation at the relay stations, *i.e.*, the rate of flow arriving at the node must be no larger than the rate of flow coming out of the node. Under this constraint, the rate of a multihop route is limited by the minimum of the rates of intermediate hops, consistent with the conclusions from information theoretic study in [62, 49]. Constraint (5.3e) represents the transmit power constraint at the node, and constraint (5.3c) reflects the bandwidth constraint of the carrier k . Finally, constraint (5.3d) reflects the practical constraint that one subcarrier is allocated to one link only. Constraints (5.3c) and (5.3d) suggests that $w_{l,k}$ is either 0 or w_k .

Unfortunately, the combination of (5.3c) and (5.3d) makes the problem intractable due to its combinatorial nature. In order to utilize convex optimization techniques to gain some insight on the problem, we relax the frequency constraint and allow a subcarrier to be shared by multiple links, *i.e.*, $w_{l,k}$ can take continuous value between 0 and w_k . This relaxation corresponds to further splitting the subcarrier's bandwidth. The relaxed problem is

$$C^* = \max \sum_{k \in \mathcal{K}} \sum_{\{l:t(l) \in \mathcal{S}\}} C_{l,k}, \quad (5.4)$$

subjected to the following constraints:

$$C_{l,k} \leq w_{l,k} \log(1 + |h_{l,k}|^2 \rho_{l,k}/w_{l,k}), \forall l \in \mathcal{E}, \forall k \in \mathcal{K}, \quad (5.5a)$$

$$0 \leq C_{l,k}, \forall l \in \mathcal{E}, \forall k \in \mathcal{K}, \quad (5.5b)$$

$$0 \leq w_{l,k} \leq w_k, \forall k \in \mathcal{K}, \quad (5.5c)$$

$$\sum_{l \in \mathcal{E}} w_{l,k} \leq w_k, \forall k \in \mathcal{K}, \quad (5.5d)$$

$$\sum_{\{l:t(l)=i\}} \sum_{k \in \mathcal{K}} \rho_{l,k} \leq \rho_i, \forall i \in \mathcal{V}, \quad (5.5e)$$

$$0 \leq \rho_{l,k}, \forall l \in \mathcal{E}, \forall k \in \mathcal{K}, \quad (5.5f)$$

$$\sum_{k \in \mathcal{K}} \sum_{\{l:t(l)=i\}} C_{l,k} \geq \sum_{k \in \mathcal{K}} \sum_{\{l:r(l)=i\}} C_{l,k}, \forall i \in \mathcal{R}. \quad (5.5g)$$

We note that (5.5d) ensures that frequency is not reused, *i.e.*, there is no interference.

Thus, (5.5a) holds.

It can be shown that the new problem is a convex optimization problem. Hence, standard optimization tools such as the KKT conditions can be employed, leading to insights on optimum solutions that can help design suboptimum, yet efficient, numerical algorithms. We note that the constraint set of the new problem, *e.g.*, frequency constraints, is no smaller than that of the original problem; thus, the solution to the new problem provides an upper bound on the solution of the original problem (5.2).

5.4 KKT Conditions After the Relaxation

The Lagrangian for the relaxed optimization problem (5.4) is

$$\begin{aligned}
L = & - \sum_{k \in \mathcal{K}} \sum_{\{l: t(l) \in \mathcal{S}\}} C_{l,k} \\
& + \sum_{k \in \mathcal{K}} \sum_{l \in \mathcal{E}} \lambda_{l,k} [C_{l,k} - w_{l,k} \log(1 + |h_{l,k}|^2 \rho_{l,k} / w_{l,k})] \\
& + \sum_{k \in \mathcal{K}} \nu_k [-w_k + \sum_{l \in \mathcal{E}} w_{l,k}] \\
& + \sum_{i \in \mathcal{V}} \tau_i [-\rho_i + \sum_{k \in \mathcal{K}} \sum_{\{l: t(l)=i\}} P_{l,k}] \\
& + \sum_{i \in \mathcal{R}} \beta_i [-\sum_{k \in \mathcal{K}} \sum_{\{l: t(l)=i\}} C_{l,k} + \sum_{k \in \mathcal{K}} \sum_{\{l: r(l)=i\}} C_{l,k}],
\end{aligned}$$

where $\lambda_{l,k}$, ν_k and β_i are Lagrange multipliers.

The partial derivatives of the Lagrangian with respect to the optimization parameters, *i.e.*, rate, power and bandwidth, are as follows:

$$\frac{\partial L}{\partial C_{l,k}} = \begin{cases} -\beta_{t(l)} + \beta_{r(l)} + \lambda_{l,k}, & t(l) \notin \mathcal{S}, r(l) \neq d, \\ -\beta_{t(l)} + \lambda_{l,k} & t(l) \notin \mathcal{S}, r(l) = d, \\ -1 + \beta_{r(l)} + \lambda_{l,k} & t(l) \in \mathcal{S}, r(l) \neq d, \\ -1 + \lambda_{l,k} & t(l) \in \mathcal{S}, r(l) = d, \end{cases} \quad (5.6)$$

$$\frac{\partial L}{\partial \rho_{l,k}} = -\log e \frac{\lambda_{l,k} |h_{l,k}|^2 w_{l,k}}{w_{l,k} + \rho_{l,k} |h_{l,k}|^2} + \tau_{t(l)}, \quad (5.7)$$

$$\begin{aligned}
\frac{\partial L}{\partial w_{l,k}} = & -\lambda_{l,k} \log\left(1 + \frac{\rho_{l,k} |h_{l,k}|^2}{w_{l,k}}\right) + \log e \frac{\lambda_{l,k} |h_{l,k}|^2 \rho_{l,k}}{w_{l,k} + |h_{l,k}|^2 \rho_{l,k}} \\
& + \nu_k, \quad (5.8)
\end{aligned}$$

A necessary condition for the optimal solutions $(C_{l,k}^*, \rho_{l,k}^*, w_{l,k}^*)$ are that

$$\frac{\partial L}{\partial C_{l,k}} \Big|_{(C_{l,k}^*, \rho_{l,k}^*, w_{l,k}^*)} \begin{cases} = 0 & \text{for } C_{l,k}^* > 0, \\ \geq 0 & \text{for } C_{l,k}^* = 0 \end{cases} \quad (5.9)$$

$$\frac{\partial L}{\partial \rho_{l,k}} \Big|_{(C_{l,k}^*, \rho_{l,k}^*, w_{l,k}^*)} \begin{cases} = 0 & \text{for } \rho_{l,k}^* > 0, \\ \geq 0 & \text{for } \rho_{l,k}^* = 0, \end{cases} \quad (5.10)$$

$$\frac{\partial L}{\partial w_{l,k}} \Big|_{(C_{l,k}^*, \rho_{l,k}^*, w_{l,k}^*)} \begin{cases} = 0 & \text{for } w_{l,k}^* > 0, \\ \geq 0 & \text{for } w_{l,k}^* = 0, \\ \leq 0 & \text{for } w_{l,k}^* = w_k, \end{cases} \quad (5.11)$$

and

$$\lambda_{l,k} \geq 0,$$

$$\beta_i \geq 0,$$

$$\nu_i \geq 0,$$

$$\tau_i \geq 0.$$

From (5.11), if a subcarrier k is assigned exclusively to a link l , *i.e.*, $w_{l,k}^* = w_k$, the necessary condition is

$$\nu_k \leq \lambda_{l,k} \log\left(1 + \frac{\rho_{l,k}^* |h_{l,k}|^2}{w_{l,k}^*}\right) - \log e \frac{\lambda_{l,k} |h_{l,k}|^2 \rho_{l,k}^*}{w_{l,k}^* + |h_{l,k}|^2 \rho_{l,k}^*}. \quad (5.12)$$

An important observation from (5.12) is that its left hand side only depends on the subcarrier k ; hence, only the link with the largest right hand side can use the subcarrier exclusively. Thus, the subcarrier assignment problem can be viewed as a bidding process. Consider the right hand side of (5.12) as the bidding price offered by link l towards subcarrier k for utilizing its bandwidth; the highest bidder win the exclusive right to use the bandwidth in subcarrier k . When there are multiple highest bidders, the bandwidth of subcarrier k is shared.

Following (5.10), the optimal power allocation for a node given a subcarrier assignment and Lagrange multipliers is,

$$\rho_{l,k}^* = w_{l,k}^* \left[\log e \frac{\lambda_{l,k}}{\tau_{t(l)}} - \frac{1}{|h_{l,k}|^2} \right]^+. \quad (5.13)$$

The power allocation (5.13) is similar to the standard water-filling solution [22].

From (5.11) and (5.6), the following necessary conditions for $C_{l,k}^* \neq 0$ can be inferred,

$$\lambda_{l,k} = \beta_{t(l)} - \beta_{r(l)}, \quad t(l) \notin \mathcal{S}, r(l) \neq d, \quad (5.14)$$

$$\lambda_{l,k} = \beta_{t(l)}, \quad t(l) \notin \mathcal{S}, r(l) = d, \quad (5.15)$$

$$\lambda_{l,k} = 1, \quad t(l) \in \mathcal{S}, r(l) = d, \quad (5.16)$$

$$\lambda_{l,k} = 1 - \beta_{r(l)}, \quad t(l) \in \mathcal{S}, r(l) \neq d. \quad (5.17)$$

We can further unify (5.14) – (5.17) into a single expression,

$$\lambda_{l,k} = \beta_{t(l)} - \beta_{r(l)}, \quad (5.18)$$

by the following extension of β_i , *i.e.*,

$$\beta_d = 0 \quad , \quad \beta_i = 1, \quad \forall i \in \mathcal{S}. \quad (5.19)$$

Similarly, the necessary condition for $C_{l,k}^* = 0$ is

$$\lambda_{l,k} \geq \beta_{t(l)} - \beta_{r(l)}. \quad (5.20)$$

An analogy between information flow inside the wireless network and water flow inside a pipeline network can be drawn by considering β_i as the pressure at node $i \in \mathcal{V}$ and $\lambda_{l,k}$ as the resistance of link l at subcarrier k . By (5.20), information will flow through link l via subcarrier k only if the difference in pressure between the transmit node and the received node, *i.e.*, $\beta_{t(l)} - \beta_{r(l)}$ is no smaller than the resistance

$\lambda_{l,k}$. Expression (5.18) further suggests that the pressure inside the network reaches an equilibrium, *i.e.*, the difference in pressure is equal to the resistance, at the optimum solution. Again, it is interesting to observe that the pressure difference between two nodes does not depend on the subcarrier index.

Even though the relaxed problem (5.4) is convex and can be solved efficiently [11], the computational complexity can still be large for a large number of subcarriers. Furthermore, solutions to the relaxed problem (5.4) do not always provide a feasible solution to the original problem (5.2). Hence, the following section further exploits characteristics of multihop networks and insights from the relaxed problem to develop practical, low-complexity algorithms for the original problem (5.2).

5.5 Low-Complexity Algorithms

This section focuses on developing practical, low-complexity algorithms for multihop networks with one destination, one source, and one or multiple relay stations. In Section 5.5.1, we develop a practical algorithm for a two-hop network motivated by the observation that a two-band partition of bandwidth is an optimum solution to the relaxed problem under certain assumptions. For the network with more than two hops, a greedy approach is proposed in Section 5.5.2.

5.5.1 High-SNR Approximation Algorithm for Two-Hop Network

For a two-hop network, we denote the source as node 1 and the relay as node 2, the source-relay link as link l_1 and the relay-destination link as link l_2 . We also define the per-subcarrier effective SNR as

$$\hat{\rho}_{l_i,k} := \log e \frac{\lambda_{l_i,k} |h_{l_i,k}|^2}{\tau_i} - 1, \quad i = 1, 2. \quad (5.21)$$

It can be verified from (5.21) and (5.13) that

$$\hat{\rho}_{l_i,k} = \frac{\rho_{l_i,k}}{w_{l_i,k}} |h_{l_i,k}|^2 \quad \text{for } \rho_{l_i,k} > 0. \quad (5.22)$$

The following Proposition indicates the optimum frequency assignment for the relaxed problem is a two-band partition of the total available bandwidth given a set of Lagrange multipliers, hence suggesting a way to significantly reduce the computation complexity.

Proposition 1 *Assuming*

1. Lagrange multipliers $\lambda_{l_1,k}, \lambda_{l_2,k}, \forall k$ are known;
2. $\hat{\rho}_{l_i,k} \gg 1, \forall k, i = 1, 2$;
3. $|h_{l_1,k}|^{2\lambda_{l_1,k}} / |h_{l_2,k}|^{2\lambda_{l_2,k}}$ decreases as k increases;

the optimum subcarrier allocation to (5.4) is a two-band partition of bandwidth, i.e., there exists $1 \leq L_1 \leq |\mathcal{K}|$, such that

$$\begin{aligned} w_{l_1,k} &= w_k, & w_{l_2,k} &= 0 & \text{for } k < L_1, \\ w_{l_1,k} &= 0, & w_{l_2,k} &= w_k & \text{for } L_1 < k \leq L_2. \end{aligned}$$

Proof: The proof of Proposition 1 follows similar lines as Theorem 2 in [78]. See Appendix C for details. ■

We note that Assumption 3 can be easily met by reordering the subcarrier indices. Recall (5.22), Assumption 2 can be met if SNR is high. Also, by the KKT conditions, it can be shown that $\lambda_{l_1,k} = \lambda_1 \geq 0, \lambda_{l_2,k} = \lambda_2 \geq 0, \forall k$, and $\lambda_1 + \lambda_2 = 1$ for the two-hop network. Combining this observation with Proposition 1, we propose Algorithm 5.5.1 that returns the optimum rate given λ_1 .

A line search method [8] can then be employed to find λ_1 that maximizes the rate, i.e.,

$$\max_{\lambda_1 \in (0,1)} g_\lambda(\lambda_1), \tag{5.23}$$

where $g_\lambda(\lambda_1)$ is the function that returns the optimum rate of the two-hop network given λ_1 , as defined in Algorithm 5.5.1. We refer to this method as the high-SNR approximation algorithm.

- 1: Function $g_\lambda(\lambda_1)$
- 2: {return the optimum rate of the two-hop network given λ_1 }
- 3: initialize an array $c[1, \dots, |\mathcal{K}|]$;
- 4: relabel subcarrier indices such that $|h_{l_1,k}|^{2\lambda_1}/|h_{l_2,k}|^{2(1-\lambda_1)}$ decreases as k increases;
- 5: **for** $k = 1$ to $|\mathcal{K}| - 1$ **do**
- 6: Perform water-filling for node 1 using subcarrier 1 to k ;
- 7: Perform water-filling for node 2 using subcarrier $k + 1$ to $|\mathcal{K}|$;
- 8: Calculate the end-to-end rate based on the above subcarrier assignment and power allocation; store the rate into $c[k]$;
- 9: **end for**
- 10: return the maximum element inside the array c .

Although the high-SNR approximation algorithm is motivated assuming high SNR, Section 5.6 shows by simulation that it also works reasonably well for even moderate SNR. Unfortunately, the algorithm only applies to two-hop networks. For the more general multihop network, we suggest the following greedy algorithm as an alternative suboptimum solution.

5.5.2 Greedy Algorithm

The greedy algorithm proposed in this section is motivated from the KKT condition (5.12), which suggests nodes bid for exclusive rights to subcarriers. Obviously, the price each node would offer to a subcarrier depends on how much reward this subcarrier bring to the system. The better the channel quality of a subcarrier, the more reward it brings. This idea has been exploited by opportunistic communication in a multi-access channel (MAC) [67, 39]. Opportunistic communication [67, 39] in MAC achieves a higher throughput by assigning resources to the user with the best channel condition. However, straightforward extensions of opportunistic communications to a multihop network, *e.g.*, assigning a subcarrier to the link with the best channel, might result in a waste of bandwidths as the end-to-end system rate is limited by the minimum rate of all hops. Instead, we propose the greedy algorithm

to assign subcarriers to hops with the best reward in increasing the end-to-end rate for the whole network. Consider that, initially, we have a set of available subcarriers that are not assigned to any links. To increase the end-to-end rate of a route, we choose one available subcarrier that maximizes the rate of the bottleneck link and assign it to the bottleneck link. After this assignment, the bottleneck link that limits the route performance may change and we repeat the above process until all subcarriers are assigned. Note that at each step, the end-to-end rate is non-decreasing. A more precise description of the greedy algorithm is given as Algorithm 2.

Algorithm 2 Greedy Algorithm

- 1: Initialize the first link as the bottleneck link l_b ;
 - 2: Initialize the set of unassigned subcarrier \mathcal{K}' as \mathcal{K} ;
 - 3: **for** $k = 1$ to $|\mathcal{K}|$ **do**
 - 4: Assign from \mathcal{K}' the subcarrier j with the maximum $|h_{l_b,j}|^2$ to the bottleneck link l_b ; remove subcarrier j from \mathcal{K}' .
 - 5: Each transmit node performs water-filling based on the current subcarrier assignment.
 - 6: Calculate each link's rate; designate the link with the smallest rate as the new bottleneck link l_b .
 - 7: **end for**
 - 8: Each transmit node performs water-filling based on the current subcarrier assignment; calculate each link's rate and return the minimum rate.
-

In general, the greedy algorithm is only suboptimum.

5.6 Simulation Results

This section provides numerical simulation results for a linear uniform multihop network with independent multipath fading among links. Specifically, all nodes are located on a straight line, and the distance between the source and destination is normalized to be 1. Without loss of generality, the source is assumed to be located at $(0, 0)$, the i^{th} relay station located at $(i/N, 0)$ for $i = 1, \dots, N - 1$, and the destination located at $(1, 0)$, where N is the number of hops. We adopt a 4-path

Rayleigh fading model with unity power decay factor and exponential power delay profile for all links. The path-loss factor is assumed to be proportional to $d_{i,j}^{-v}$, where $d_{i,j}$ is the distance from node i to node j , and v is a constant value, chosen to be 4 in our setup. Each node is assumed to have a transmit power constraint ρ . All subcarriers have the same bandwidth of $1/|\mathcal{K}|$ such that the total system bandwidth is normalized to 1, *i.e.*, $\sum_{k \in \mathcal{K}} w_k = 1$.

We consider a fixed allocation of subcarriers as a baseline reference. In the fixed scheduling algorithm, each subcarrier is assigned to a hop regardless of channel conditions. Specifically, assuming $\mathcal{K} = \{1, \dots, K\}$, for n^{th} hop, if $n + (\lfloor \frac{K}{N} \rfloor)N > K$ ¹, we assign subcarriers $[n, n + N, \dots, n + (\lfloor \frac{K}{N} \rfloor - 1)N]$; and if $n + (\lfloor \frac{K}{N} \rfloor)N \leq K$, we assign subcarriers $[n, n + N, \dots, n + (\lfloor \frac{K}{N} \rfloor - 1)N, n + (\lfloor \frac{K}{N} \rfloor)N]$. After the subcarriers are assigned, we run water-filling power allocation for each hop, compute the rate for each hop and find the minimum of rates of all hops.

Fig. 5.1 and Fig. 5.2 compare the performance of different algorithms for two-hop and three-hop wireless networks, respectively. The optimum performances in Fig. 5.1 and Fig. 5.2 are obtained by exhaustively searching all possible subcarrier allocations. We note that the complexity of exhaustive search grows exponentially as the number of subcarriers increases. Fig. 5.1 and Fig. 5.2 demonstrate that the greedy algorithm offers significant gains, up to 5 dB, compared with the fix scheduling in the high SNR regime. This is because that a better allocation of subcarriers results in more significant improvement of rates in the bandwidth limited regime. The average performance of high-SNR approximation algorithm is slightly better than that of the greedy algorithms. However, the performances of both algorithms, *i.e.*, the greedy algorithm and the high-SNR approximation algorithm, are close to the optimum performance. Note that given the same number of subcarriers, the

¹ $\lfloor x \rfloor$ rounds x to the nearest integer towards minus infinity.

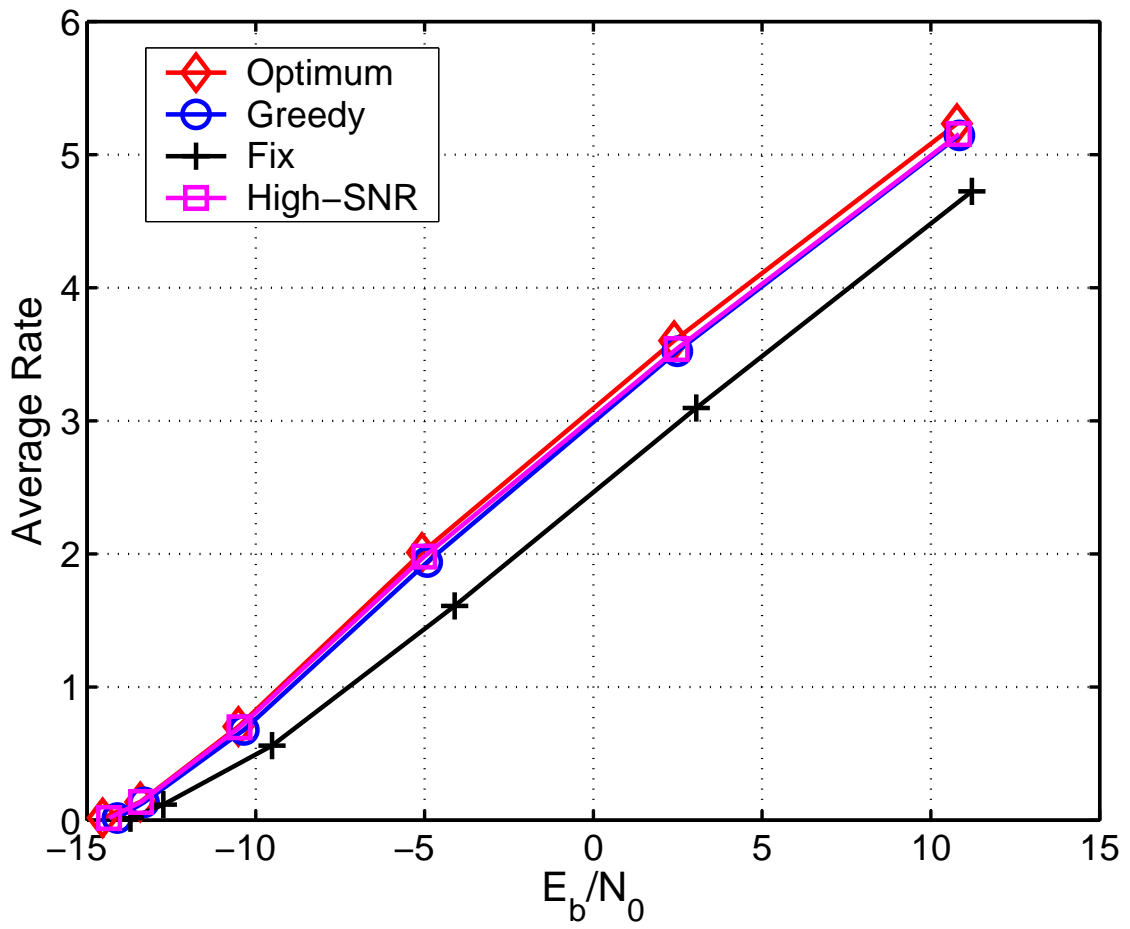


Figure 5.1. Average rate versus SNR for a two-hop wireless network with 8 subcarriers.

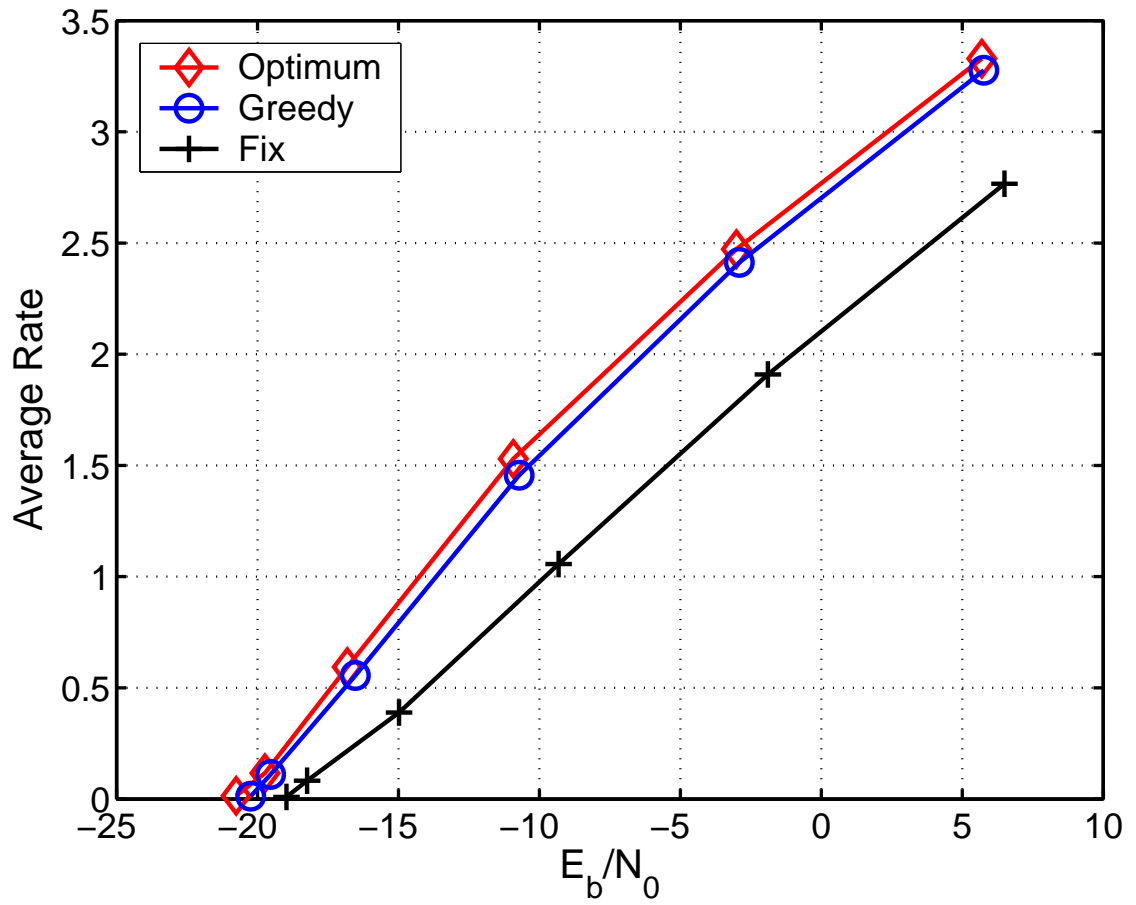


Figure 5.2. Average rate versus SNR for a three-hop wireless network with 8 sub-carriers.

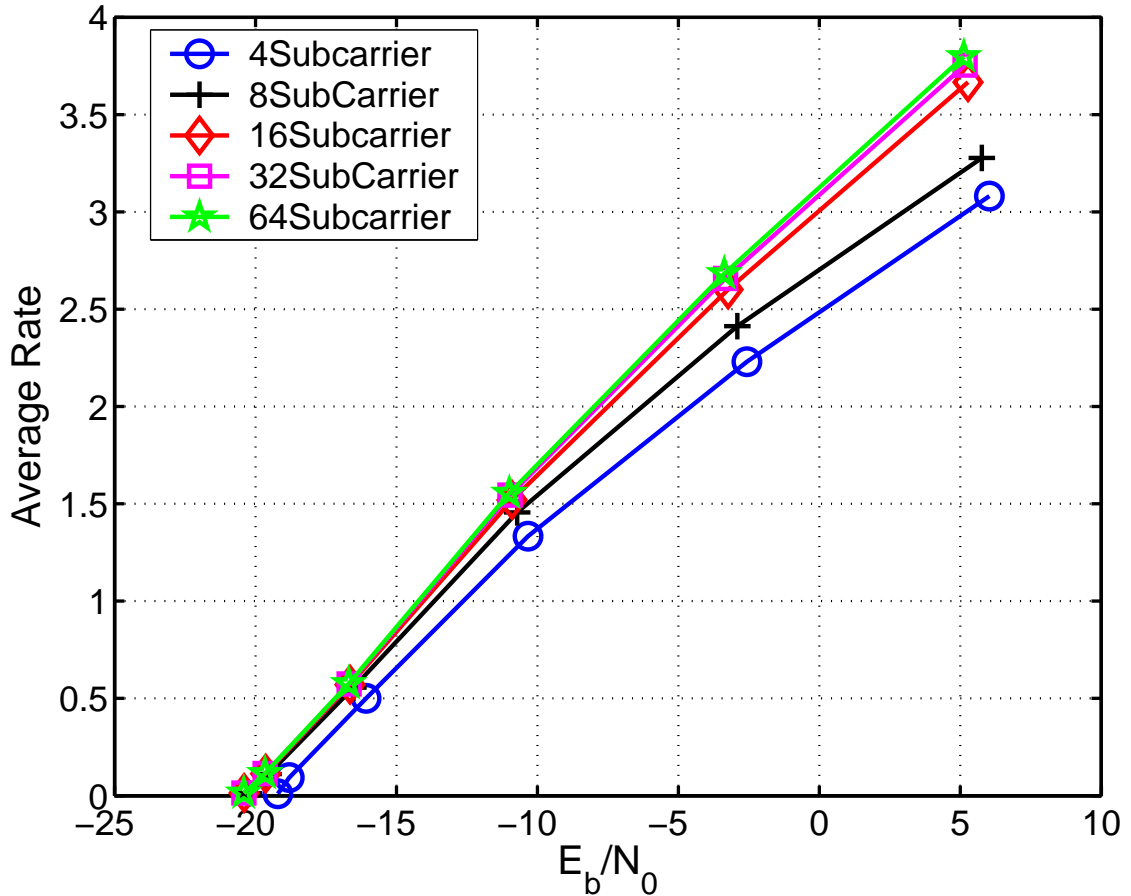


Figure 5.3. Average rate versus SNR for a three-hop wireless network with different number of subcarriers. The greedy algorithm is used.

performance gap between the greedy algorithm and exhaustive search increases as the number of hops increases.

Fig. 5.3 shows that as the number of subcarriers increases, the performance of the greedy algorithm generally improves due to a finer partition of bandwidth. The gain, however, diminishes when the number of subcarriers is large because of increasing correlations of channel states between adjacent subcarriers.

Fig. 5.4 compares the performance of networks with different number of hops given the same number of subcarriers. It is clear from Fig. 5.4 that as SNR increase, the optimum number of hops in terms of maximizing the spectral efficiency

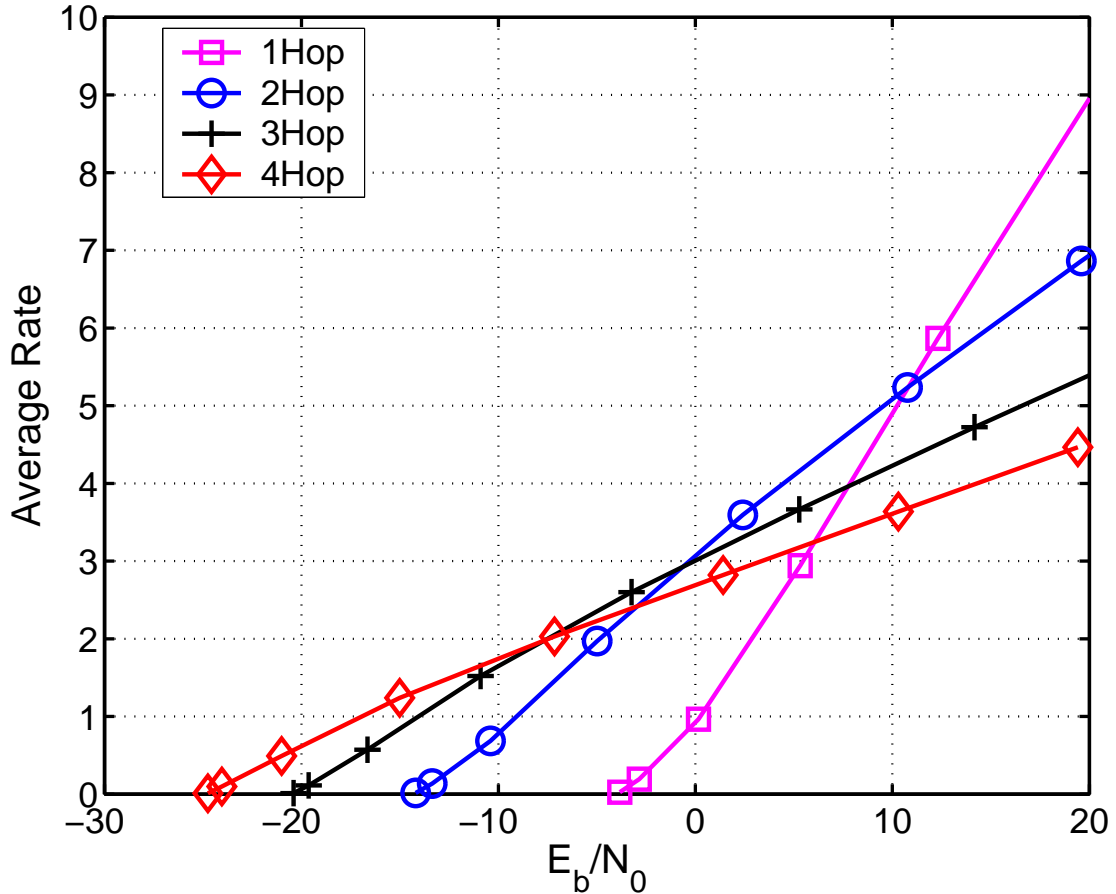


Figure 5.4. Average rate versus SNR of networks with different number of hops, 16 subcarriers. The greedy algorithm is used.

decreases. We emphasize that even though this observation is made for a broadband OFDM system assuming frequency selective fading channels and channel state information available to transmitters for power allocations, it is consistent with the observation from a narrowband AWGN system [62, 49].

5.7 Conclusion

This chapter formulates an optimization problem for the joint power allocation and subcarrier assignment for a multihop network with OFDM. More importantly, the chapter proposes two low-complexity algorithms, namely, the high-SNR approx-

imization algorithm and the greedy algorithm, to solve the optimization problem. Simulation results suggest both algorithms closely tracks the optimum performance. Furthermore, our simulations results indicate that in a broadband OFDM system, more hops help in the power limited regime, but do not help in the bandwidth limited regime, consistent with previous observations from a narrowband system.

CHAPTER 6

CONCLUSIONS

This chapter summarizes contributions of the dissertation and suggests future research directions.

6.1 Contributions

The overarching message of this dissertation is that there exist relaying or routing protocols that can take advantage of relay stations in improving performance of wireless networks, and yet are simple enough for practical implementation. Design of these protocols must consider different trade-offs in wireless relay networks, *e.g.*, complexity versus performance, power efficiency versus bandwidth efficiency, and diversity versus multiplexing. Various previous works have focused on proposing and studying complicated relaying protocols. Not surprisingly, these works demonstrate significant benefits of adding *complicated* relay stations. By contrast, we emphasize throughout this dissertation that wireless networks can benefit from addition of *simple* relay stations. We designed simple relaying (routing) protocols that led to improved robustness to fading or substantial power savings.

Specifically, in Chapter 3, we propose a simple relaying protocol, namely, multi-access amplify-forward (MAF), for the block fading multi-access channel with one relay station. We derive the diversity-multiplexing trade-off (DMT) of MAF for the two-user symmetric MARC. We demonstrate that MAF is optimal in terms of DMT

in the high multiplexing regime for the two-user symmetric MARC. Surprisingly, for a regime in which the more complicated relaying protocols, *i.e.*, dynamic decode-forward (DDF) and compress-forward (CF), are not DMT-optimal, the simple MAF protocol is proved to be DMT-optimal. For the MARC with more than two users employing MAF, we provide upper and lower bounds on the DMT, and show that the upper bound is tight in the low multiplexing regime. We further show that MAF outperforms DDF in the high multiplexing regime and outperforms CF in the low multiplexing regime.

For a wireless network with many relay stations, Chapter 4 develops distributed routing schemes that offer higher spectral efficiency compared to nearest neighbor routing and direct communication. By contrast with previous works, Chapter 4 intends to put a networking problem on a solid information-theoretic basis. Specifically, Chapter 4 proposes two routing schemes, approximately-ideal-path routing (AIPR) and distributed spectrum-efficient routing (DSER), that take into account the trade-off between power and bandwidth efficiency. AIPR finds a path to approximate an optimum regular path and requires location information. DSER is more amenable to distributed implementations based on Bellman-Ford or Dijkstra's algorithms. Our simulations show that the spectral efficiencies of AIPR and DSER for random networks approach that of nearest-neighbor routing in the low signal-to-noise ratio (SNR) regime and that of single-hop routing in the high SNR regime. In the moderate SNR regime, the spectral efficiency of DSER is up to twice that of nearest-neighbor or single-hop routing when path-loss exponent is 4.

Given that a route is selected, Chapter 5 further considers resource allocation, *e.g.*, power and bandwidth allocation, along multiple hops in the selected route. Based on intuition from KKT solutions to a convex relaxation of the problem, we propose a low-complexity efficient algorithm that is optimum at high-SNR for a

two-hop network. Extending the idea of opportunistic transmission, we propose a greedy approach for subcarrier allocation for a route with more than two hops. Furthermore, we also observe that more hops do not always improve spectral efficiency in a broadband system, consistent with previous results for a narrowband system.

6.2 Future Research

In this section, we collect a few directions for future research on wireless relay networks as follows:

- **Practical Coding and Decoding Algorithms:** Throughout the dissertation we have employed random coding arguments to evaluate performance of various relaying and routing schemes. Although recent work [36, 14] has appeared for designing coding schemes for the relay channel and MARC, further effort in designing coding and decoding algorithms and evaluating their performance is necessary for practical implementation of MAF protocols. In practice, practical coding schemes are often combined with ARQ schemes and can induce new trade-offs for wireless multihop networks. For example, a larger coding block length reduces the error probability per link, but might increase the end-to-end delay and increase the probability of queueing overflow [49].
- **Low SNR Analysis:** Most practical systems operate in the regime of moderate SNR. However, in general, an analysis for the regime of moderate SNR is difficult. Hence, in Chapter 3, we have focused on the high SNR DMT analysis to get some insight. It might be also possible to develop an analysis for the MARC in the regime of low SNR. And a combination of different perspectives from high and low SNR can provide a more comprehensive perspective on how wireless relay networks operate. Recent work on the outage probability at low

SNR [4] and work at the wideband spectral efficiency [72] provide promising frameworks that can be applied to the MARC and multihop networks.

- **Practical Radio Implementation:** We have focused on proposing simple protocols for wireless relay networks. Recent emergence of software defined radio (SDR), *e.g.*, GNU-Radio, has offered a powerful, flexible and inexpensive platform to implement the proposed protocols. A testbed implementation would indicate the degree to which the assumptions in the dissertation, *e.g.*, block and symbol synchronization among the radios and availability of channel state information, are reasonable. A testbed would also allow for study of the important issues of overhead required for channel estimation, multi-access control, and routing.
- **Routing with Cooperative Diversity:** In Chapter 4, we have focused on designing routing protocols that improve spectral efficiency of networks without exploiting cooperative diversity. Combining a judiciously chosen route with user cooperation schemes might further increase the rate and improve the reliability. However, interactions between routing and user cooperation have not been fully studied. To study the impact of user cooperation on designing routing protocols remains a challenging task. Recent work in [77] may provide a useful framework for studying the interaction between routing and user cooperation.
- **Queueing:** This dissertation has adopted the typical assumption of continuous traffic from the source(s), hence allowing us to make use of Shannon theory. This assumption corresponds to scenarios in which networks are heavily loaded with traffic, highlighting the importance of spectral efficiency. However, when networks are not heavily loaded with traffic, the bursty nature of traffic should

not be neglected and requires a potentially different perspective. Formulating the relaying or routing problem in the MARC or the multihop network under the assumption of bursty traffic can lead to related, but different, communication strategies and requires further combination of networking and information theoretic perspectives. The unification of networking and information theory has remained a challenging but promising topic [27]. Different sets of analysis tools, *e.g.*, dynamical programming, and different performance metrics, *e.g.*, queueing delay, might be needed for further pursuing this direction. An interesting work on this direction is in [56], where it is shown that, by exploiting the large number of idle users, the throughput of a cooperative wireless random-access network with Rayleigh fading approaches that of random-access network with AWGN.

- **Distributed Resource Allocation:** Resource allocation in wireless relay networks becomes difficult as more information and constraints need to be considered. Chapter 5 has developed a centralized approach towards subcarrier allocation, but allowed distributed power allocation in each transmit node. A promising approach in developing distributed resource allocation algorithms is the network utility optimization framework [51]. The challenge in applying network utility optimization to wireless relay networks is in taking into account, in a distributed way, the broadcast nature of wireless communication.
- **General networks:** This dissertation has focused on two types of wireless networks, *i.e.*, the MARC and multihop networks. Further study of more general networks would provide more guidelines for building wireless relay networks. One challenge in studying general networks is to come up with a basic network model that is more general than what we studied in this dissertation,

but not so complicated that the resulting analysis becomes intractable. The work in [12] provides an interesting framework that models how cooperating terminals can be connected to each other in wireless relay networks subjected to different system resource constraints. Much more work is needed to obtain better understanding of such models and issues involved.

APPENDIX A

PROOF OF THEOREM 1

Proof: The proof uses the machinery of Theorem 2 in [68] and Lemma 2 in [7] as well as some of the techniques of Theorem 3 in [5]. Section A.2 and Section A.3 provides the lower and upper bound, respectively. We provide detail development on bounding the error probability that only one user cannot be decoded in Section A.2.2. Section A.2.2 also intends to familiarize readers with the standard proof technique for DMT so that we don't have to provide detail development in other part of the disserataion. On other parts of the proof, *e.g.*, bounding the error probability that both users cannot be decoded and bounding the outage probabilities, we only provide a sketech on the standard part and focus on the novel part of our proof.

A.1 Preliminary

A.1.1 Notation

In this chapter, random variables and random vectors are denoted using the sans serif (*e.g.*, x) and bold sans serif (*e.g.*, \mathbf{X}) fonts, respectively. Calligraphic letters denote events or sets (*e.g.*, \mathcal{S}), and \mathcal{S}^c represents the complementary set. \mathcal{R}^N and \mathcal{C}^N represents the set of real and complex N -tuples, respectively, and \mathcal{R}^{N+} represents the set of non-negative real N -tuples. \mathcal{S}^{c+} represents the non-negative complementary set provided \mathcal{S} is a non-negative set, *i.e.*, $\mathcal{S}^{c+} := \mathcal{S}^c \cap \mathcal{R}^{N+}$ if $\mathcal{S} \in \mathcal{R}^N$. H^\dagger represents a Hermitian transpose of H and x^+ means $\max\{x, 0\}$.

A.1.2 Exponential Equality

Exponential equality is denoted by \doteq , *e.g.*, $f(\rho) \doteq \rho^v$ when

$$\lim_{\rho \rightarrow \infty} \frac{\log f(\rho)}{\log \rho} = v.$$

In this expression, v is called the exponential order of $f(\rho)$. $\dot{\leq}$ and $\dot{\geq}$ are defined similarly.

For a Gaussian random variable h with zero mean and unit variance, the exponential order ν of $1/|h|^2$ is

$$\nu = \lim_{\rho \rightarrow \infty} -\frac{\log |h|^2}{\log \rho}.$$

And the probability density function (PDF) of ν is shown [7] to be

$$P_\nu(v) = \lim_{\rho \rightarrow \infty} \log \rho \rho^{-v} \exp(-\rho^{-v}).$$

Using the definition of exponential equality, we can have

$$P_\nu(v) \doteq \begin{cases} \rho^{-v} & \text{for } v \geq 0, \\ 0 & \text{for } v < 0. \end{cases} \quad (\text{A.1})$$

Intuitively, (A.1) simplifies diversity-multiplexing analysis by allowing us to focus on the non-negative set of exponential orders of random variables. More precisely, for random variables $\{\nu_j\}_{j=1}^N$ that are independent and identically distributed according to $P_\nu(v)$, the probability $P_{\mathcal{S}}$ that (ν_1, \dots, ν_N) belongs to set \mathcal{S} can be characterized by

$$P_{\mathcal{S}} \doteq \rho^{-d_{\mathcal{S}}}, \quad (\text{A.2})$$

where

$$d_{\mathcal{S}} = \inf_{(\nu_1, \dots, \nu_N) \in \mathcal{S}^+} \sum_{j=1}^N \nu_j. \quad (\text{A.3})$$

In general, for the fading coefficient between node i and j , *i.e.*, $h_{i,j}$, we denote $\nu_{i,j}$ as the exponential order of $1/|h_{i,j}|^2$.

A.1.3 Codebooks

We consider the ensemble of i.i.d Gaussian random codes. Specifically, each user generates a codebook \mathcal{C}^i containing $\rho^{r_i \times l}$ codewords, denoted as $X_1^{(i)}, X_2^{(i)}, \dots, X_{\rho^{r_i \times l}}^{(i)}$ for $i = 1, 2$. Each codeword is a $2 \times l/2$ matrix with $\mathcal{C.N.}(0, 1)$ i.i.d. entries. Once picked, the codebooks are revealed to the receiver. In each block period, the transmitted signals of users are simply chosen from the corresponding codebook \mathcal{C}^i equiprobably according to the message to be transmitted. We focus on the symmetric case, *i.e.*, $r_i = r/2$ for $i = 1, 2$.

A.2 Lower Bound

Following the outline of [68] and [7], we split the joint error event \mathcal{E} into mutually exclusive error events \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 , *i.e.*,

$$\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3, \quad (\text{A.4})$$

where \mathcal{E}_1 (\mathcal{E}_2) represents the error event that only user 1 (2) is detected in error, and \mathcal{E}_3 represents the error event that both of the users are detected in error. Applying the union bound to (A.4) yields

$$P_{\mathcal{E}} \leq P_{\mathcal{E}_1} + P_{\mathcal{E}_2} + P_{\mathcal{E}_3}. \quad (\text{A.5})$$

$P_{\mathcal{E}_1}$, $P_{\mathcal{E}_2}$ and $P_{\mathcal{E}_3}$ are bounded by first eliminating the contribution of the correctly decoded user, if any, from the received signal, and then bounding the corresponding pairwise error probabilities for the remaining user(s) [5].

To proceed, we consider the average pairwise error probability for a linear Gaussian MIMO channel $\mathbf{Y} = H\mathbf{X} + \mathbf{Z}$ assuming Gaussian codewords. From [7, 81], the average pairwise error probability is approximated by

$$P_{p_{\mathcal{E}}} = \det\left(I + \frac{1}{2}H\Sigma_x H^H \Sigma_z^{-1}\right)^{-l}, \quad (\text{A.6})$$

where: I is a unity matrix; Σ_x and Σ_z denote the covariance matrices of signals \mathbf{X} and noise \mathbf{Z} , respectively and l is the codeword length.

Utilizing (A.6), we have the average pairwise error probabilities corresponding \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 , *i.e.*,

$$P_{\mathcal{P}\mathcal{E}_i|\mathbf{H}} \leq \left[1 + \frac{\rho}{2} |h_{i,d}|^2 + \frac{\rho}{2} \frac{|h_{i,d}|^2 + |\mathbf{b}h_{r,d}h_{i,r}|^2 + \frac{\rho}{2} |h_{i,d}|^4}{1 + |\mathbf{b}h_{r,d}|^2} \right]^{-l/2} \quad \text{for } i = 1, 2 \quad (\text{A.7})$$

$$P_{\mathcal{P}\mathcal{E}_3|\mathbf{H}} \leq \left[1 + \frac{\rho}{2} (|h_{1,d}|^2 + |h_{2,d}|^2) + \frac{\rho}{2} \frac{(|h_{1,d}|^2 + |h_{2,d}|^2) + |\mathbf{b}h_{r,d}|^2 (|h_{1,r}|^2 + |h_{2,r}|^2)}{1 + |\mathbf{b}h_{r,d}|^2} + \frac{\rho^2}{4} \frac{(|h_{1,d}|^2 + |h_{2,d}|^2)^2 + |\mathbf{b}h_{r,d}|^2 |h_{1,d}h_{2,r} - h_{2,d}h_{1,r}|^2}{1 + |\mathbf{b}h_{r,d}|^2} \right]^{-l/2}, \quad (\text{A.8})$$

where $\mathbf{H} = [h_{1,r}, h_{2,r}, h_{r,d}, h_{1,d}, h_{2,d}]$.

A.2.1 Bounding Single User Error Event

First, we consider bounding $P_{\mathcal{E}_1}$ from (A.7). Applying the union bound on $\rho^{rl/2}$ codewords and taking high SNR approximation, we have the following error probability conditioned on channel realizations:

$$P_{\mathcal{E}_1|\mathbf{H}} \leq \rho^{-\{\max[1-v_{1,d}, 2(1-v_{1,d}), 1-v_{r,d}-v_{1,r}]-r\} \frac{l}{2}}. \quad (\text{A.9})$$

We define the following set

$$\mathcal{G}_1 = \{(v_{1,d}, v_{r,d}, v_{1,r}) \geq 0 : \max[1 - v_{1,d}, 2(1 - v_{1,d}), 1 - v_{r,d} - v_{1,r}] \leq r\}. \quad (\text{A.10})$$

By Bayes' rule, we can bound $P_{\mathcal{E}_1}$ conditional on \mathcal{G}_1 , *i.e.*,

$$P_{\mathcal{E}_1} \leq P_{\mathcal{E}_1, \mathcal{G}_1^c} + P_{\mathcal{G}_1}, \quad (\text{A.11})$$

where: $P_{\mathcal{G}_1}$ is the probability that exponential order of channel realizations belongs to set \mathcal{G}_1 and $P_{\mathcal{E}_1, \mathcal{G}_1^c}$ is the probability that the event \mathcal{E}_1 happens and exponential order of channel realizations belongs to set \mathcal{G}_1^c .

Recall (A.2), we can have

$$P_{\mathcal{G}_1} \doteq \rho^{-d_{\mathcal{G}_1}}, \quad (\text{A.12})$$

where

$$d_{\mathcal{G}_1} = \inf_{(v_{1,d}, v_{r,d}, v_{1,r}) \in \mathcal{G}_1^+} v_{1,d} + v_{r,d} + v_{1,r} \quad (\text{A.13})$$

$$= (1 - r/2)^+ + (1 - r)^+. \quad (\text{A.14})$$

We then average (A.9) on channel realizations that belongs to set \mathcal{O}_1^c to bound $P_{\mathcal{E}_1, \mathcal{G}_1^c}$, *i.e.*,

$$P_{\mathcal{E}_1, \mathcal{G}_1^c} \leq \rho^{-d_{\mathcal{G}_1^c}}, \quad (\text{A.15})$$

where

$$d_{\mathcal{G}_1^c} = \inf_{(v_{1,d}, v_{r,d}, v_{1,r}) \in \mathcal{G}_1^c} v_{1,d} + v_{r,d} + v_{1,r} + \frac{l}{2} \{ \max [1 - v_{1,d}, 2(1 - v_{1,d}), 1 - v_{r,d} - v_{1,r}] - r \}. \quad (\text{A.16})$$

Since we assume l is large, the infimum in (A.16) is achieved when the equality in \mathcal{G}_1^c holds. After some algebra manipulation, we have

$$d_{\mathcal{G}_1^c} = (1 - r/2)^+ + (1 - r)^+. \quad (\text{A.17})$$

We combine (A.12), (A.15) and (A.11) to have

$$P_{\mathcal{E}_1} \leq \rho^{-d_{\mathcal{E}_1}}, \quad (\text{A.18})$$

where

$$d_{\mathcal{E}_1} = [(1 - \frac{r}{2})^+ + (1 - r)^+]. \quad (\text{A.19})$$

Similar results are obtained for $P_{\mathcal{E}_2}$.

A.2.2 Bounding Two User Error Event

Next, we bound $P_{\mathcal{E}_3}$. However, averaging (A.8) to bound $P_{\mathcal{E}_3}$ is not straightforward due to the term

$$|h_{1,d}h_{2,r} - h_{2,d}h_{1,r}|^2,$$

which involves subtraction. To circumvent this problem, define

$$\Theta := \frac{h_{2,r}h_{1,d} - h_{1,r}h_{2,d}}{\sqrt{|h_{1,r}|^2 + |h_{2,r}|^2}} \quad \text{and} \quad (\text{A.20})$$

$$\Omega := \frac{h_{1,r}^*h_{1,d} + h_{2,r}^*h_{2,d}}{\sqrt{|h_{1,r}|^2 + |h_{2,r}|^2}}. \quad (\text{A.21})$$

It is then straightforward to see that *conditioned* on $h_{1,r}$ and $h_{2,r}$, Θ and Ω are two complex Gaussian random variables with zero mean and unit variance. Furthermore, $E\{\Theta\Omega^*|h_{1,r}, h_{2,r}\} = 0$, meaning that Θ and Ω are conditionally uncorrelated and therefore independent. Realizing that,

$$|\Theta|^2 + |\Omega|^2 = |h_{1,d}|^2 + |h_{2,d}|^2,$$

(A.8) can be written as

$$P_{\mathcal{P}\mathcal{E}_3|\mathbf{H}} \leq \left[1 + \frac{\rho}{2}(|\Theta|^2 + |\Omega|^2) + \frac{\rho}{2} \frac{(|\Theta|^2 + |\Omega|^2) + |bh_{r,d}|^2 (|h_{1,r}|^2 + |h_{2,r}|^2)}{1 + |bh_{r,d}|^2} + \frac{\rho^2}{4} \frac{(|\Theta|^2 + |\Omega|^2)^2 + |bh_{r,d}|^2 (|h_{1,r}|^2 + |h_{2,r}|^2) |\Theta|^2}{1 + |bh_{r,d}|^2} \right]^{-l/2}. \quad (\text{A.22})$$

Note that since Θ , Ω , $h_{1,r}$ and $h_{2,r}$ are correlated, the techniques of [5] cannot directly be applied to average (A.22). However, by averaging in two steps, *i.e.*, fixing $h_{1,r}$ and $h_{2,r}$ and taking the conditional average with respect to Θ , Ω and $h_{r,d}$, and then taking the average with respect to $h_{1,r}$ and $h_{2,r}$, $P_{\mathcal{E}_3}$ can be bounded.

We denote the average error probability conditional on $|h_{1,r}|$ and $|h_{2,r}|$ as $P_{\mathcal{E}_3|h_{1,r}, h_{2,r}}$.

Utilizing the Bayes' rule, we have

$$P_{\mathcal{E}_3|h_{1,r}, h_{2,r}} \leq P_{\mathcal{E}_3|h_{1,r}, h_{2,r}, \mathcal{G}_{h_{1,r}, h_{2,r}}} + P_{\mathcal{G}_{h_{1,r}, h_{2,r}}},$$

where $\mathcal{G}_{h_{1,r}, h_{2,r}}$ is a set of $(v_\Theta, v_\Omega, v_{r,d})$.

By defining

$$\mathcal{G}_{h_{1,r}, h_{2,r}} = \{(v_\Theta, v_\Omega, v_{r,d}) \in \mathcal{R}^{3+} : \{\max[2(1 - \min(v_\Theta, v_\Omega)), 1 - v_{r,d} - \min(v_{1,r}, v_{2,r}), 2 - v_{r,d} - v_\Theta - \min(v_{1,r}, v_{2,r})]\}^+ \leq 2r\}, \quad (\text{A.23})$$

we have

$$P_{\mathcal{G}_{h_{1,r},h_{2,r}}} \doteq \rho^{-\inf_{(v_{\Theta},v_{\Omega},v_{r,d}) \in \mathcal{G}_{h_{1,r},h_{2,r}}} v_{\Theta} + v_{\Omega} + v_{r,d}}, \quad (\text{A.24})$$

and

$$P_{\mathcal{E}_3|h_{1,r},h_{2,r},\mathcal{G}_{h_{1,r},h_{2,r}}^c} \doteq \rho^{-d'_{\mathcal{E}_3|h_{1,r},h_{2,r}}}, \quad (\text{A.25})$$

where: v_{Θ} and v_{Ω} are the exponential order of $1/|\Theta|^2$ and $1/|\Omega|^2$, respectively; and

$$d'_{\mathcal{E}_3|h_{1,r},h_{2,r}} = \inf_{(v_{\Theta},v_{\Omega},v_{r,d}) \in \mathcal{G}_{h_{1,r},h_{2,r}}^c} v_{\Theta} + v_{\Omega} + v_{r,d} \quad (\text{A.26})$$

$$+ \frac{l}{2} \{ \max[2(1 - \min(v_{\Theta}, v_{\Omega})), 1 - v_{r,d} - \min(v_{1,r}, v_{2,r}), 2 - v_{r,d} - v_{\Theta} - \min(v_{1,r}, v_{2,r})] - 2r \}^+. \quad (\text{A.27})$$

Consider a large l , we can see that the infimum for both (A.26) and (A.24) is achieved at the boundary of $\mathcal{G}_{h_{1,r},h_{2,r}}$, leading to

$$P_{\mathcal{E}_3|h_{1,r},h_{2,r}} \dot{\leq} \rho^{-d_{\mathcal{E}_3|h_{1,r},h_{2,r}}} \quad (\text{A.28})$$

where

$$d_{\mathcal{E}_3|h_{1,r},h_{2,r}} = \inf_{(v_{\Theta},v_{\Omega},v_{r,d}) \in \mathcal{G}_{h_{1,r},h_{2,r}}} v_{\Theta} + v_{\Omega} + v_{r,d} \quad (\text{A.29})$$

$$= \begin{cases} 2(1-r)^+ & \text{for } \min\{v_{1,r}, v_{2,r}\} > (1-r)^+ \\ [3(1-r) - \min\{v_{1,r}, v_{2,r}\}]^+ & \text{for } 0 \leq \min\{v_{1,r}, v_{2,r}\} \leq (1-r)^+ \end{cases} \quad (\text{A.30})$$

To obtain (A.30) from (A.29), we first notice that when $r > 1$, $d_{\mathcal{E}_3|h_{1,r},h_{2,r}} = 0$. Therefore, in the following, we assume $0 \leq r \leq 1$. In Fig. A.1 and Fig. A.2.2, we then show the projection of $\mathcal{G}_{h_{1,r},h_{2,r}}$ on the plane $v_{\Omega} = C \geq 1-r$ as the shaded region for $0 \leq \min\{v_{1,r}, v_{2,r}\} \leq 1-r$ and $\min\{v_{1,r}, v_{2,r}\} \geq 1-r$, respectively. From Fig. A.1 and Fig. A.2.2, it is then clear that (A.30) is the desired infimum.

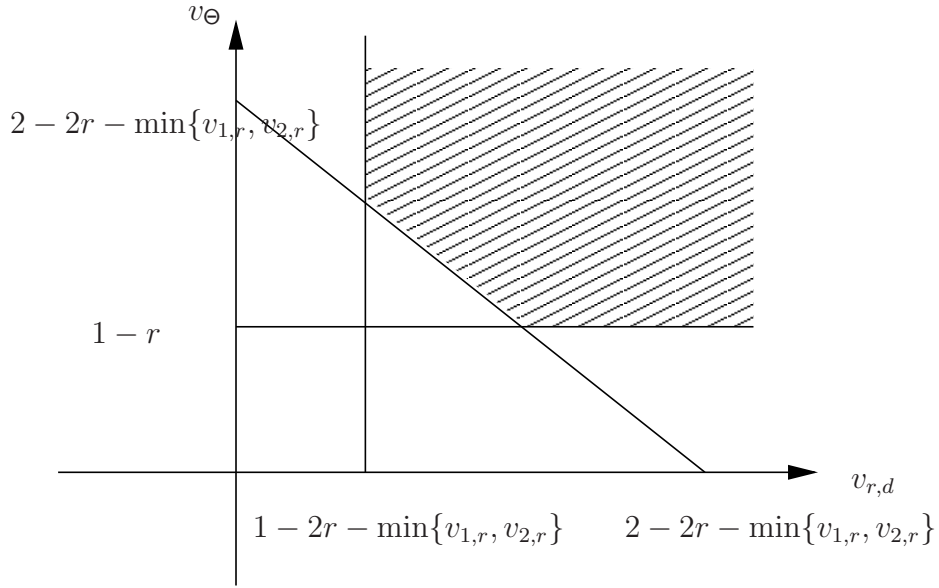


Figure A.1. The shadowed regions are feasible regions for $(v_{\Theta}, v_{r,d})$ on the plane $v_{\Omega} = C \geq 1 - r$ when $0 \leq \min\{v_{1,r}, v_{2,r}\} \leq 1 - r$ for $0 \leq r \leq 1$

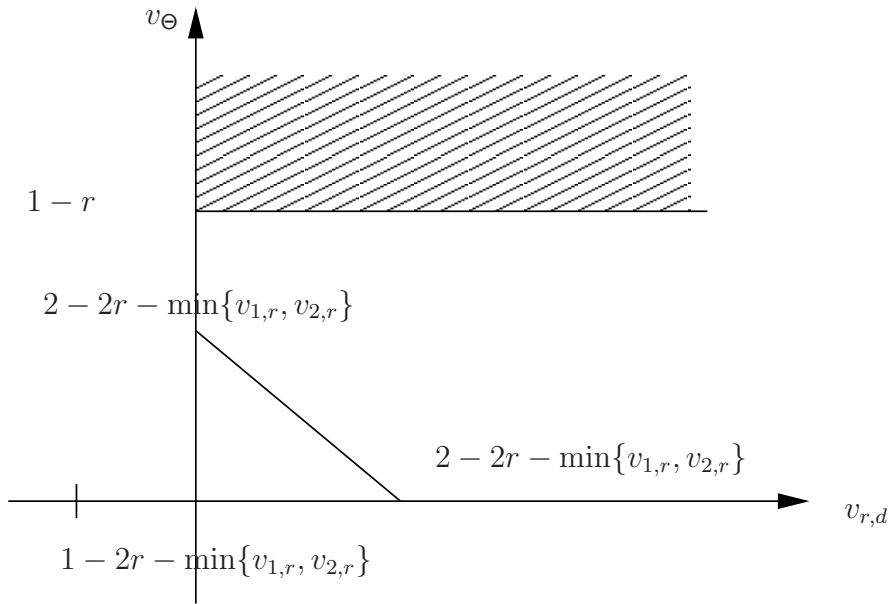


Figure A.2. The shadowed regions are feasible regions for $(v_{\Theta}, v_{r,d})$ on the plane $v_{\Omega} = C \geq 1 - r$ when $\min\{v_{1,r}, v_{2,r}\} \geq 1 - r$ for $0 \leq r \leq 1$

It then follows that

$$P_{\mathcal{E}_3} = E_{h_{1,r}, h_{2,r}} [P_{\mathcal{E}_3|h_{1,r}, h_{2,r}}] \stackrel{\leq}{\leq} \rho^{-d_{\mathcal{E}_3}}. \quad (\text{A.31})$$

where $E[\cdot]$ denotes the expectation operation and

$$d_{\mathcal{E}_3} = \inf_{v_{1,r}, v_{2,r} \geq 0} d_{\mathcal{E}_3|h_{1,r}, h_{2,r}} + v_{1,r} + v_{2,r} \quad (\text{A.32})$$

$$= 3(1-r)^+ \quad (\text{A.33})$$

From (A.5), it is then straightforward to conclude that the ML error probability of two-user MARC with MAF is upper bounded by

$$P_{\mathcal{E}} \stackrel{\leq}{\leq} \rho^{-d_{\mathcal{E}}},$$

where

$$\begin{aligned} d_{\mathcal{E}} &\geq \min\{d_{\mathcal{E}_1}, d_{\mathcal{E}_3}\} \\ &= \min\left\{\left(1 - \frac{r}{2}\right)^+ + (1-r)^+, 3(1-r)^+\right\} \\ &= \begin{cases} 2 - \frac{3r}{2} & \text{for } 0 \leq r \leq 2/3 \\ 3(1-r) & \text{for } 2/3 \leq r \leq 1 \end{cases}. \end{aligned}$$

A.3 Upper Bound

As Lemma 5 of [81] shows, the outage probability is a lower bound on the error probability; hence, it will provide an upper bound on the DMT.

Following the outline of [68] and [7], we split the joint outage event \mathcal{O} into mutually exclusive outage events \mathcal{O}_1 , \mathcal{O}_2 and \mathcal{O}_3 , *i.e.*,

$$\mathcal{O} = \mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{O}_3, \quad (\text{A.34})$$

where \mathcal{O}_1 (\mathcal{O}_2) represents the outage event that only user 1 (2) is in outage, and \mathcal{O}_3 represents the outage event that both of the users are in outage. From (A.34), it is

straightforward to see

$$P_{\mathcal{O}} \geq \max\{P_{\mathcal{O}_1}, P_{\mathcal{O}_2}, P_{\mathcal{O}_3}\}. \quad (\text{A.35})$$

A.3.1 Bounding Single User Outage

First, we bound the outage event that only user 1 is in outage, *i.e.*,

$$P_{\mathcal{O}_1} = P \left[\frac{1}{2} \log(1 + \rho |h_{i,d}|^2 + \rho \frac{|h_{i,d}|^2 + |bh_{r,d}h_{i,r}|^2 + \rho |h_{i,d}|^4}{1 + |bh_{r,d}|^2}) \leq \frac{r}{2} \log \rho \right] \quad (\text{A.36})$$

For large SNR, we can simplify (A.36) as

$$P_{\mathcal{O}_1} \doteq P\{\max[1 - v_{1,d}, 2(1 - v_{1,d}), 1 - v_{r,d} - v_{1,r}] \leq r\} \quad (\text{A.37})$$

$$\doteq \rho^{-d_{\mathcal{O}_1}} \quad (\text{A.38})$$

where the second exponential equality comes from (A.2) and

$$d_{\mathcal{O}_1} = \inf_{(v_{1,d}, v_{r,d}, v_{1,r}) \in \mathcal{G}_1^+} v_{1,d} + v_{r,d} + v_{1,r} \quad (\text{A.39})$$

$$= (1 - r/2)^+ + (1 - r)^+. \quad (\text{A.40})$$

Note that \mathcal{G}_1 is defined in (A.10).

A.3.2 Bounding Two User Outage

Next, we bound the outage event that both users are in outage, *i.e.*,

$$P_{\mathcal{O}_3} = P \left\{ \frac{1}{2} \log \left[1 + \rho(|h_{1,d}|^2 + |h_{2,d}|^2) + \rho \frac{(|h_{1,d}|^2 + |h_{2,d}|^2) + |bh_{r,d}|^2(|h_{1,r}|^2 + |h_{2,r}|^2)}{1 + |bh_{r,d}|^2} \right. \right. \\ \left. \left. + \rho^2 \frac{(|h_{1,d}|^2 + |h_{2,d}|^2)^2 + |bh_{r,d}|^2 |h_{1,d}h_{2,r} - h_{2,d}h_{1,r}|^2}{1 + |bh_{r,d}|^2} \right] \leq r \log \rho \right\} \quad (\text{A.41})$$

Again, bounding (A.41) gets more involved due to the existence of subtraction $|h_{1,d}h_{2,r} - h_{2,d}h_{1,r}|^2$. We can exploit similar techniques as in the development of error probability that both users are decoding incorrectly, *i.e.*, utilizing transformations

(A.20) and (A.21), fixing $h_{1,r}$ and $h_{2,r}$ and taking the conditional average with respect to Θ , Ω and $h_{r,d}$, and then taking the average with respect to $h_{1,r}$ and $h_{2,r}$. However, we provide another technique that is general and simple in providing the lower bound on the outage probability. We note that

$$|h_{1,d}h_{2,r} - h_{2,d}h_{1,r}|^2 \leq 2(|h_{1,d}h_{2,r}|^2 + |h_{2,d}h_{1,r}|^2). \quad (\text{A.42})$$

Hence, $P_{\mathcal{O}_3}$ can be further bounded

$$P_{\mathcal{O}_3} \geq P \left\{ \frac{1}{2} \log \left[1 + \rho(|h_{1,d}|^2 + |h_{2,d}|^2) + \rho \frac{(|h_{1,d}|^2 + |h_{2,d}|^2) + |bh_{r,d}|^2(|h_{1,r}|^2 + |h_{2,r}|^2)}{1 + |bh_{r,d}|^2} \right. \right. \\ \left. \left. + \rho^2 \frac{(|h_{1,d}|^2 + |h_{2,d}|^2)^2 + 2|bh_{r,d}|^2(|h_{1,d}h_{2,r}|^2 + |h_{2,d}h_{1,r}|^2)}{1 + |bh_{r,d}|^2} \right] \leq r \log \rho \right\} \quad (\text{A.43})$$

At high SNR, we can simplify (A.43) as

$$P_{\mathcal{O}_1} \geq P \{ \max [1 - v_{1,d}, 2(1 - v_{1,d}), 1 - v_{r,d} - v_{1,r}, \\ 1 - v_{2,d}, 2(1 - v_{2,d}), 1 - v_{r,d} - v_{2,r}, \\ 2 - v_{r,d} - v_{1,d} - v_{2,r}, 2 - v_{r,d} - v_{2,d} - v_{1,r},] \leq 2r \} \quad (\text{A.44})$$

$$\doteq \rho^{-d_{\mathcal{O}_3}} \quad (\text{A.45})$$

where the exponential equality comes from (A.2) and

$$d_{\mathcal{O}_3} = 3(1 - r)^+. \quad (\text{A.46})$$

Combining (A.44), (A.37) and (A.35), we obtain

$$P_{\mathcal{E}} \geq \rho^{-d_{\mathcal{E}}},$$

where

$$d_{\mathcal{E}} \geq \min \{ d_{\mathcal{O}_1}, d_{\mathcal{O}_3} \} \\ = \min \left\{ \left(1 - \frac{r}{2}\right)^+ + (1 - r)^+, 3(1 - r)^+ \right\}.$$

We conclude our proof by noticing that the upper bound and lower bound on DMT overlaps.

■

APPENDIX B

PROOF OF THEOREM 2

In this appendix, we provide a proof for Theorem 2. As we have provided a detail description on the standard development techniques for DMT in Appendix A, we only provide a sketch of key steps in this appendix.

B.1 Proof of Lower Bound

Proof: We consider the ensemble of i.i.d $\mathcal{C.N.}$ random codes. Specifically, each user generates a codebook \mathcal{C}^i containing $\rho^{r_i \times l}$ codewords, denoted as $X_1^{(i)}, X_2^{(i)}, \dots, X_{\rho^{r_i \times l}}^{(i)}$. Each codeword is a $2 \times l/2$ matrix with $\mathcal{C.N.}(0, 1)$ i.i.d. entries. Once picked, the codebooks are revealed to the receiver. In each block period, the transmitted signals of users are simply chosen from the corresponding codebook \mathcal{C}^i equiprobably according to the message to be transmitted. We focus on the symmetric case, *i.e.*, $r_i = r/N$. In general, for the fading coefficient between node i and j , *i.e.*, $h_{i,j}$, we denote $\nu_{i,j}$ as the exponential order of $1/|h_{i,j}|^2$.

Following methods in [68, 7], we bound the system error probability with the type- \mathcal{S} error probability, *i.e.*,

$$P_{\mathcal{E}} \leq \sum_{\mathcal{S}} P_{\mathcal{E}_{\mathcal{S}}}, \tag{B.1}$$

where: \mathcal{S} is a subset of users; $\mathcal{E}_{\mathcal{S}}$ is the event that the receiver decodes messages of users in set \mathcal{S} incorrectly and decode user in \mathcal{S}^c correctly.

$P_{\mathcal{E}_{\mathcal{S}}}$ is bounded by first eliminating the contribution of the correctly decoded

user, if any, from the received signal, and then bounding the average corresponding pairwise error probabilities for the remaining user(s) [5]. The average pairwise error probability for type- \mathcal{S} error is defined as the probability that an arbitrary codeword from users in \mathcal{S} , *e.g.*, $[X_1^{(i_1)}, \dots, X_1^{(i_{|\mathcal{S}|})}]$, $\mathcal{S} = \{i_1, \dots, i_{|\mathcal{S}|}\}$, is detected as another codeword, *e.g.*, $[X_2^{(i_1)}, \dots, X_2^{(i_{|\mathcal{S}|})}]$, $\mathcal{S} = \{i_1, \dots, i_{|\mathcal{S}|}\}$, averaged over ensemble of codes. The average pairwise probability corresponding to type- \mathcal{S} error is bounded by

$$P_{\mathcal{P}\mathcal{S}|\mathbf{H}} \leq \left[1 + \frac{\rho}{2} \sum_{i \in \mathcal{S}} |h_{i,d}|^2 + \frac{\rho \sum_{i \in \mathcal{S}} (|h_{i,d}|^2 + |bh_{r,d}h_{i,r}|^2)}{1 + |bh_{r,d}|^2} + \frac{\rho^2 (\sum_{i \in \mathcal{S}} |h_{i,d}|^2)^2 + \sum_{i \in \mathcal{S}} |h_{i,d}|^2 \sum_{i \in \mathcal{S}} |bh_{r,d}h_{i,r}|^2 - |\sum_{i \in \mathcal{S}} bh_{r,d}h_{i,r}h_{i,d}^*|^2}{4(1 + |bh_{r,d}|^2)} \right]^{-l/2} \quad (\text{B.2})$$

From Cauchy-Schwarz inequality, we have

$$\sum_{i \in \mathcal{S}} |h_{i,d}|^2 \sum_{i \in \mathcal{S}} |bh_{r,d}h_{i,r}|^2 \geq \left| \sum_{i \in \mathcal{S}} bh_{r,d}h_{i,r}h_{i,d}^* \right|^2$$

Hence, (B.2) can be further bounded by

$$P_{\mathcal{P}\mathcal{S}|\mathbf{H}} \leq \left[1 + \frac{\rho}{2} \sum_{i \in \mathcal{S}} |h_{i,d}|^2 + \frac{\rho \sum_{i \in \mathcal{S}} (|h_{i,d}|^2 + |bh_{r,d}h_{i,r}|^2)}{1 + |bh_{r,d}|^2} + \frac{\rho^2 (\sum_{i \in \mathcal{S}} |h_{i,d}|^2)^2}{4(1 + |bh_{r,d}|^2)} \right]^{-l/2}. \quad (\text{B.3})$$

For high SNR, we apply the union bound to (B.3) and obtain

$$P_{\mathcal{E}\mathcal{S}|\mathbf{H}} \leq \rho^{-\frac{l}{2} \{ \max[\max_{i \in \mathcal{S}} (1 - v_{i,d}), \max_{i \in \mathcal{S}} (1 - v_{r,d} - v_{i,r}), \max_{i,k \in \mathcal{S}} (2 - v_{k,d} - v_{i,d})] - 2 \sum_{i \in \mathcal{S}} r_i \}} \quad (\text{B.4})$$

We follow the standard methods in the development of DMT in averaging (B.4) over channel realizations, *i.e.*,

$$P_{\mathcal{E}\mathcal{S}} \leq \rho^{-d_{\mathcal{S}}}, \quad (\text{B.5})$$

where

$$d_{\mathcal{S}} = \inf_{v_{i,d}, v_{i,r}, v_{r,d} > 0} \sum_{i \in \mathcal{S}} v_{i,d} + v_{i,r} + v_{r,d} \quad (\text{B.6})$$

subject to

$$\begin{aligned} 2 - v_{i,d} - v_{j,d} &\leq 2 \sum_{i \in \mathcal{S}} r_i \quad \text{for } i, j \in \mathcal{S}, \\ 2(1 - v_{i,d}) &\leq 2 \sum_{i \in \mathcal{S}} r_i \quad \text{for } i \in \mathcal{S}, \\ 1 - v_{i,r} - v_{r,d} &\leq 2 \sum_{i \in \mathcal{S}} r_i \quad \text{for } i \in \mathcal{S}. \end{aligned}$$

Assuming $r_1 = r/N$, (B.6) can be simplified as

$$d_{\mathcal{S}} = |\mathcal{S}| \left(1 - \frac{|\mathcal{S}|}{N} r\right)^+ + \left(1 - 2 \frac{|\mathcal{S}|}{N} r\right)^+. \quad (\text{B.7})$$

Combining (B.1) with (B.7) provides a lower bound on the DMT of the symmetric N -user MARC, *i.e.*,

$$d_{\text{MARC}}(r) \geq \min_{\mathcal{S}} |\mathcal{S}| \left(1 - \frac{|\mathcal{S}|}{N} r\right)^+ + \left(1 - 2 \frac{|\mathcal{S}|}{N} r\right)^+, \quad (\text{B.8})$$

$$= \min_{k \in \{1, \dots, N\}} k \left(1 - \frac{k}{N} r\right)^+ + \left(1 - 2 \frac{k}{N} r\right)^+ \quad (\text{B.9})$$

■

B.2 Proof of Upper Bound

Proof: The upper bound of the DMT is obtained by bounding the outage probability. Following methods in [68, 7], we bound the system outage probability with the type- \mathcal{S} outage probability, *i.e.*,

$$P_{\mathcal{O}} \geq \max_{\mathcal{S}} P_{\mathcal{O}_{\mathcal{S}}}, \quad (\text{B.10})$$

where $\mathcal{O}_{\mathcal{S}}$ is the event that the receivers in set \mathcal{S} are in outage and users in \mathcal{S}^c are not in outage.

Type-S outage probability is defined as following:

$$\begin{aligned}
P_{\mathcal{O}_S} &= P \left\{ \frac{1}{2} \log \left[1 + \rho \sum_{i \in \mathcal{S}} |h_{i,d}|^2 + \rho \frac{\sum_{i \in \mathcal{S}} (|h_{i,d}|^2 + |bh_{r,d}h_{i,r}|^2)}{1 + |bh_{r,d}|^2} \right. \right. \\
&\quad \left. \left. + \rho^2 \frac{(\sum_{i \in \mathcal{S}} |h_{i,d}|^2)^2 + \sum_{i \in \mathcal{S}} |h_{i,d}|^2 \sum_{i \in \mathcal{S}} |bh_{r,d}h_{i,r}|^2 - |\sum_{i \in \mathcal{S}} bh_{r,d}h_{i,r}h_{i,d}^*|^2}{1 + |bh_{r,d}|^2} \right] \right. \\
&\quad \left. \leq \sum_{i \in \mathcal{S}} r_i \log \rho \right\}. \tag{B.11}
\end{aligned}$$

When $|\mathcal{S}| = 1$, (B.11) is equivalent to the outage probability of NAF [5]. Hence,

$$P_{\mathcal{O}_S} \geq \rho^{-((1-r/N)^+ + (1-2r/N))} \text{ for } |\mathcal{S}| = 1. \tag{B.12}$$

In the general case of $|\mathcal{S}| > 1$, it is straightforward to see that (B.11) can be bounded by

$$\begin{aligned}
P_{\mathcal{O}_S} &\geq Pr \left\{ \frac{1}{2} \log \left[1 + \rho \sum_{i \in \mathcal{S}} |h_{i,d}|^2 + \rho \frac{\sum_{i \in \mathcal{S}} (|h_{i,d}|^2 + |bh_{r,d}h_{i,r}|^2)}{1 + |bh_{r,d}|^2} \right. \right. \\
&\quad \left. \left. + \rho^2 \frac{(\sum_{i \in \mathcal{S}} |h_{i,d}|^2)^2 + \sum_{i \in \mathcal{S}} |h_{i,d}|^2 \sum_{i \in \mathcal{S}} |bh_{r,d}h_{i,r}|^2}{1 + |bh_{r,d}|^2} \right] \right. \\
&\quad \left. \leq \sum_{i \in \mathcal{S}} r_i \log \rho \right\}. \tag{B.13}
\end{aligned}$$

Following the standard techniques in [81, 5], to bound (B.13) results in

$$P_{\mathcal{O}_S} \geq \rho^{-(1+|\mathcal{S}|)(1-\frac{|\mathcal{S}|}{N}r)^+}, \text{ for } |\mathcal{S}| > 1. \tag{B.14}$$

Combining (B.12), (B.14) and (B.10) leads to the upper bound on the DMT of MAF,

$$d_{MAF}(r) \leq \min\{((1-r/N)^+ + (1-2r/N)), \min_{k=2,\dots,N} (1+k)(1-kr/N)^+\}. \tag{B.15}$$

■

APPENDIX C

PROOF OF PROPOSITION 1

Proof: We define the following function

$$\begin{aligned}
 f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) &= \lambda_{l_1,k} \left[\log(1 + |h_{l_1,k}|^2 \rho_{l_1,k}^* / w_{l_1,k}^*) \right. \\
 &\quad \left. - \log e \frac{|h_{l_1,k}|^2 \rho_{l_1,k}^*}{w_{l_1,k}^*} \frac{1}{(1 + |h_{l_1,k}|^2 \rho_{l_1,k}^* / w_{l_1,k}^*)} \right] \\
 &\quad - \lambda_{l_2,k} \left[\log(1 + |h_{l_2,k}|^2 \rho_{l_2,k}^* / w_{l_2,k}^*) \right. \\
 &\quad \left. - \log e \frac{|h_{l_2,k}|^2 \rho_{l_2,k}^*}{w_{l_2,k}^*} \frac{1}{(1 + |h_{l_2,k}|^2 \rho_{l_2,k}^* / w_{l_2,k}^*)} \right]. \quad (\text{C.1})
 \end{aligned}$$

The discussions following (5.12) of Chapter 5 clearly suggest that a subcarrier k is assigned to link l_1 if $f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) > 0$, and to link l_2 if $f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) < 0$.

Recall the definition of per-carrier effective SNR in (5.22), we can rewrite (C.1) into

$$\begin{aligned}
 f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) &= \lambda_{l_1,k} \left[\log(1 + \hat{\rho}_{l_1,k}) - \log e \frac{\hat{\rho}_{l_1,k}}{1 + \hat{\rho}_{l_1,k}} \right] \\
 &\quad - \lambda_{l_2,k} \left[\log(1 + \hat{\rho}_{l_2,k}) - \log e \frac{\hat{\rho}_{l_2,k}}{1 + \hat{\rho}_{l_2,k}} \right]. \quad (\text{C.2})
 \end{aligned}$$

Making use of Assumption 2 and (5.21), we further simplify (C.2) as

$$f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) = \log \frac{|h_{l_1,k}|^{2\lambda_{l_1,k}}}{|h_{l_2,k}|^{2\lambda_{l_2,k}}} + \lambda_{l_1,k} \log \frac{\log e \lambda_{l_1,k}}{e\tau_1} - \lambda_{l_2,k} \log \frac{\log e \lambda_{l_2,k}}{e\tau_2} \quad (\text{C.3})$$

Recall Assumption 1, we can see from (C.3) that $f(|h_{l_1,k}|^2, |h_{l_2,k}|^2)$ monotonically decreases as the subcarrier index k increases. Hence, we know there can only exist one L such that $f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) > 0$ for $k < L$ and $f(|h_{l_1,k}|^2, |h_{l_2,k}|^2) < 0$ for $k > L$.

■

BIBLIOGRAPHY

- [1] *IEEE 802.16j: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*. 2006.
- [2] Mohammad R. Aref. *Information flow in relay networks*. PhD thesis, Stanford Univ., Stanford, CA, 1980.
- [3] Erdal Arıkan. Some complexity results about packet radio networks. *IEEE Trans. Inform. Theory*, 30(4):681–685, Jul. 1984.
- [4] A. Salman Avestimehr and David N.C. Tse. Outage Capacity of the Fading Relay Channel in the Low SNR Regime. *IEEE Trans. Inform. Theory*, 53(4):1401–1415, Apr. 2007.
- [5] Kambiz Azarian and Hesham El Gamal. On the Achievable Diversity-Multiplexing Tradeoff in Half-Duplex Cooperative Channels. *IEEE Trans. Inform. Theory*, 51(12):4152–4172, Dec. 2005.
- [6] Kambiz Azarian and Hesham El Gamal. The Throughput-Reliability Trade-off in MIMO Channels. *IEEE Trans. Inform. Theory*, 53(2):488–501, Feb. 2007.
- [7] Kambiz Azarian, Hesham El Gamal, and Philip Schniter. On the Optimality of the ARQ-DDF Protocol. *IEEE Trans. Inform. Theory*. accepted subject to revisions.
- [8] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.
- [9] Ezio Biglieri, John Proakis, and Shlomo Shamai (Shitz). Fading Channels: Information-Theoretic and Communications Aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2692, Oct. 1998.
- [10] Helmut Bölcskei, Rohit U. Nabar, Özgür Oyman, and Arogyaswami J. Paulraj. Capacity Scaling Laws in MIMO Relay Networks. *IEEE Trans. Wireless Commun.*, 5(6):1433–1444, Jun. 2006.
- [11] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [12] John Boyer. Cooperative Connectivity Models and Bounds for Wireless Relay Networks. PhD thesis, Carleton University, Ottawa, Ontario, Canada, 2007.
- [13] Raffaele Bruno, Marco Conti, and Enrico Gregori. Mesh Networks: Commodity Multihop Ad Hoc Networks. *IEEE Communications Magazine*, 43(3):123–131, Mar. 2005.

- [14] Arnab Chakrabarti, Alexandre de Baynast, Ashutosh Sabharwal, and Behnaam Aazhang. Low Density Parity Check Codes for the Relay Channel. *IEEE J. Select. Areas Commun.*, 25(2):280–291, Feb. 2007.
- [15] Deqiang Chen, Kambiz Azarian, and J. Nicholas Laneman. A Case for Amplify-Forward Relaying in the Block-Fading Multiaccess Channel. *IEEE Trans. Inform. Theory*, Jan. 2007. submitted for publication.
- [16] Deqiang Chen, Martin Haenggi, and J. Nicholas Laneman. Distributed Spectrum-Efficient Routing Algorithms in Wireless Networks. *IEEE Trans. Wireless Commun.*, Apr. 2007. submitted for publication.
- [17] Robert M. Corless, G. H. Gonnet, D. E. G. Hare, David J. Jeffrey, and Donald E. Knuth. On the Lambert W Function. *Adv. Computational Maths*, 5:329–359, 1996.
- [18] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford SteinMilton Abramowitz, and Irene A. Stegun. *Introduction to Algorithms*. The MIT Press and McGraw-Hill, 2001.
- [19] M. Scott Corson, J.P. Macker, and G.H. Cirincione. Internet-based mobile ad hoc networking. *IEEE Internet Computing*, 1999.
- [20] Douglas S. J. De Couto, Daniel Aguayo, John Bicket, and Robert Morris. A High-Throughput Path Metric for Multi-Hop Wireless Routing. In *Proc. ACM MOBICOM*, 2003.
- [21] Thomas M. Cover and Abbas A. El Gamal. Capacity Theorems for the Relay Channel. *IEEE Trans. Inform. Theory*, 25(5):572–584, Sep. 1979.
- [22] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.
- [23] Olivier Dousse, Massimo Franceschetti, and Patric Thiran. On the throughput scaling of wireless relay networks. *IEEE Trans. Inform. Theory*, 52(6):2756–2761, Jun. 2006.
- [24] Richard Draves, Jitendra Padhye, and Brian Zill. Comparison of Routing Metrics for Static Multi-Hop Wireless Networks. In *Proc. ACM SIGCOMM*, 2004.
- [25] Richard Draves, Jitendra Padhye, and Brian Zill. Routing in Multi-radio, Multi-hop Wireless Mesh Networks. In *Proc. ACM MOBICOM*, 2004.
- [26] Chip Elliott and Bob Heile. Self-Organizing, Self-Healing Wireless Networks. In *Proc. IEEE Int’l Conf. on Personal Wireless Comm.*, 2000.
- [27] Anthony Ephremides and Bruce Hajek. Information Theory and Communication Networks: An Unconsummated Union. *IEEE Trans. Inform. Theory*, 44(6):2416–2434, Oct. 1998.
- [28] Jr. G. David Forney and Gottfried Ungerboeck. Modulation and Coding for Linear Gaussian Channels. *IEEE Trans. Inform. Theory*, 44(6):2384–2415, Oct. 1998.

- [29] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.
- [30] Michael Gastpar and Martin Vetterli. On the Capacity of Large Gaussian Relay Networks. *IEEE Trans. Inform. Theory*, 51(3):765–779, Mar. 2005.
- [31] Piyush Gupta and P.R. Kumar. The Capacity of Wireless Networks. *IEEE Trans. Inform. Theory*, 46(2):388–404, Mar. 2000.
- [32] Martin Haenggi. The Impact of Power Amplifier Characteristics on Routing in Random Wireless Networks. In *Proc. IEEE Global Comm. Conf. (GLOBECOM)*, 2003.
- [33] Martin Haenggi. On Routing in Random Rayleigh Fading Networks. *IEEE Trans. Wireless Commun.*, 4:1553–1562, Jul. 2005.
- [34] Martin Haenggi and Daniele Puccinelli. Routing in Ad Hoc Networks: A Case for Long Hops. *IEEE Communications Magazine*, 43:93–101, Oct. 2005.
- [35] Bruce Hajek and Galen Sasaki. Link Scheduling in Polynomial Time. *IEEE Trans. Inform. Theory*, 34(5):910–917, Sep. 1988.
- [36] Christoph Hausl and Philippe Dupraz. Joint Network-Channel Coding for the Multiple-Access Relay Channel. In *Proc. of Intern. Workshop on Wireless Ad-hoc and Sensor Networks (IWVAN)*, 2006.
- [37] Xiaoyan Hong, Kaixin Xu, and Mario Gerla. Scalable Routing Protocols for Mobile Ad Hoc Networks. *IEEE Network Magazine*, 2002.
- [38] Nihar Jindal, Jeffrey G. Andrews, and Steven Weber. Bandwidth-SINR Trade-offs in Spatial Networks. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2007.
- [39] R. Knopp and P. A. Humblet. Information Capacity and Power Control in Single-cell Multiuser Communications. In *Proc. IEEE Int. Conf. Communications (ICC)*, 1995.
- [40] Gerhard Kramer. Models and Theory for Relay Channels with Receive Constraints. In *Proc. Allerton Conf. Communications, Control, and Computing*, 2004.
- [41] Gerhard Kramer, Michael Gastpar, and Piyush Gupta. Cooperative Strategies and Capacity Theorems for Relay Networks. *IEEE Trans. Inform. Theory*, 51(9):3037–3063, Sep. 2005.
- [42] Gerhard Kramer and A. J. van Wijngaarden. On the White Gaussian Multiple-Access Relay Channel. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, Jun. 2000.
- [43] J. Nicholas Laneman. *Cooperative Diversity in Wireless Networks: Algorithms and Architectures*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, Aug. 2002.

- [44] J Nicholas Laneman, David N. C. Tse, and Gregory W. Wornell. Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior. *IEEE Trans. Inform. Theory*, 50(12):3062–3080, Dec. 2004.
- [45] Sung-Ju Lee, Julian Hsu, Russell Hayashida, Mario Gerla, and Rajive Bagrodia. Selecting A Routing Strategy for Your Ad Hoc network. *Computer Communications*, 26(7):723–733, 2003.
- [46] Ye (Geoffrey) Li and Gordon Stuber. *Orthogonal Frequency Division Multiplexing For Wireless Communications*. Springer, 2006.
- [47] P. Johansson M. Frodigh and P. Larsson. Wireless Ad Hoc Networking: The Art of Networking Without A Network. *Ericsson Review*, 2000.
- [48] Özgür Oyman. OFDM2A: A Centralized Resource Allocation Policy for Cellular Multi-hop Networks. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2006.
- [49] Özgür Oyman and Sundeep Sandhu. A Shannon-Theoretic Perspective on Fading Multihop Networks. In *Proc. Conf. Inform. Sci. and Syst. (CISS)*, 2006.
- [50] Özgür Oyman and Sundeep Sandhu. Non-Ergodic Power-Bandwidth Tradeoff in Linear Multi-hop Networks. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2006.
- [51] Daniel P. Palomar and Mung Chiang. A Tutorial on Decomposition Methods for Network Utility Maximization. 24(8):1439–1451, Aug. 2006.
- [52] Matthias Patzold. *Mobile Fading Channels*. John Wiley & Sons, Ltd., 2002.
- [53] Edward C. Posner and Arthur L. Rubin. The Capacity of Digital Links in Tandem. *IEEE Trans. Inform. Theory*, 30(3):464–470, May 1984.
- [54] John G. Proakis. *Digital Communications*. McGraw-Hill, fourth edition, 2001.
- [55] Theodore Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, 2001.
- [56] Alejandro Ribeiro, Nikolaos D. Sidiropoulos, Georgios B. Giannakis, and Yingqun Yu. Achieving Wireline Random Access Throughput in Wireless Networking Via User Cooperation. *IEEE Trans. Inform. Theory*, 53(2):732–758, Feb. 2007.
- [57] Lalitha Sankaranarayanan, Gerhard Kramer, and Narayan B. Mandayam. Capacity Theorems for the Multiple-Access Relay Channel. In *Proc. Allerton Conf. Communications, Control, and Computing*, 2004.
- [58] Lalitha Sankaranarayanan, Gerhard Kramer, and Narayan B. Mandayam. Cooperation vs. Hierarchy: An Information-theoretic Comparison. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2005.

- [59] Lalitha Sankaranarayanan, Gerhard Kramer, and Narayan B. Mandayam. Cooperative Diversity in Wireless Networks: A Geometry-Inclusive Analysis. In *Proc. Allerton Conf. Communications, Control, and Computing*, 2005.
- [60] Andrew Sendonaris, Elza Erkip, and Behnaam Aazhang. User Cooperation Diversity, Parts I & II. 15(11):1927–1948, Nov. 2003.
- [61] Claude Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 1948.
- [62] Marcin Sikora, J. Nicholas Laneman, Martin Haenggi, Jr. Daniel J. Costello, and Tom E. Fuja. Bandwidth- and Power-Efficient Routing in Linear Wireless Networks. *Joint Special Issue of IEEE Trans. Inform. Theory and IEEE Trans. Networking*, 52(6):2624–2633, Jun. 2006.
- [63] João Luís Sobrinho. Algebra and Algorithms for QoS Path Computation and Hop-by-Hop Routing in the Internet. *IEEE/ACM Trans. Networking*, 10(4):541–550, Aug. 2002.
- [64] João Luís Sobrinho. An Algebraic Theory of Dynamic Network Routing. *IEEE/ACM Trans. Networking*, 13(5):1160–1173, Oct. 2005.
- [65] Gordon L. Stuber. *Principle of Mobile Communications*. Kluwer, second edition, 2002.
- [66] Andrew S. Tanenbaum. *Computer Networks*. Prentice Hall PTR, 2002.
- [67] David N. C. Tse and Stephen V. Hanly. Multiaccess Fading Channels – Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities. *IEEE Trans. Inform. Theory*, 44(7):2796–2815, Nov. 1998.
- [68] David N.C. Tse, Pramod Viswanath, and Lizhong Zheng. Diversity-Multiplexing Tradeoff in Multiple-Access Channels. *IEEE Trans. Inform. Theory*, 50(9):1859–1874, Sep. 2004.
- [69] Gottfried Ungerboeck. Channel Coding with Multilevel/phase Signals. *IEEE Trans. Inform. Theory*, 28(1):55–67, Jan. 1982.
- [70] Edward C. van der Meulen. Three-Terminal Communication Channels. *Adv. Appl. Prob.*, 3:120–154, 1971.
- [71] Edward C. van der Meulen. A Survey of Multi-Way Channels in Information Theory : 1961–1976. *IEEE Trans. Inform. Theory*, 23(1):1–37, Jan. 1977.
- [72] Sergio Verdu. Spectral Efficiency in the Wideband Regime. *IEEE Trans. Inform. Theory*, 48(6):1319–1343, Jun. 2002.
- [73] Cheong Yui Wong, R.S. Cheng, K.B. Lataief, and R.D. Murch. Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation. *IEEE J. Select. Areas Commun.*, 17(10):1747–1757, Oct. 1999.
- [74] Aaron D. Wyner. The Rate-Distortion Function for Source Coding with Side Information at the Decoder-II: General Sources. *Inf. Control*, 38:60–80, 1978.

- [75] Liang-Liang Xie and P.R. Kumar. A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation. *IEEE Trans. Inform. Theory*, 50(5):748–767, May 2004.
- [76] Min Xie and Martin Haenggi. A Study of the Correlations between Channel and Traffic Statistics in Multihop Networks. *IEEE Trans. on Vehicular Technology*, 2007. To be published.
- [77] Edmund M. Yeh and Randall A. Berry. Throughput Optimal Control of Cooperative Relay Networks. *IEEE Trans. Inform. Theory*, 53(10):3827–3833, Oct. 2007.
- [78] Wei Yu and John M. Cioffi. FDMA Capacity of Gaussian Multiple-Access Channels With ISI. *IEEE Trans. Commun.*, 50(1):102–111, Jan. 2002.
- [79] Melda Yuksel and Elza Erkip. Multiple-Antenna Cooperative Wireless Systems: A Diversity-Multiplexing Tradeoff Perspective. *IEEE Trans. Inform. Theory*, 53(10):3371–3393, Oct. 2007.
- [80] Wenyi Zhang and Urbashi Mitra. Multihopping Strategies: An Error-Exponent Comparison. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2007.
- [81] Lizhong Zheng and David N. C. Tse. Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels. *IEEE Trans. Inform. Theory*, 49(5):1073–1096, May 2003.