Cooperative Diversity in Wireless Networks: Algorithms and Architectures

by

J. Nicholas Laneman

B.S. E.E., Washington University (1995)
B.S. C.S., Washington University (1995)
S.M. E.E., Massachusetts Institute of Technology (1997)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2002

Author .....
Department of Electrical Engineering and Computer Science
August 1, 2002

Certified by .....
Gregory W. Wornell
Professor
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Cooperative Diversity in Wireless Networks:

Algorithms and Architectures

by

J. Nicholas Laneman

Submitted to the Department of Electrical Engineering and Computer Science
on August 1, 2002, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

To effectively combat multipath fading across multiple protocol layers in wireless networks, this dissertation develops energy-efficient algorithms that employ certain kinds of cooperation among terminals, and illustrates how one might incorporate these algorithms into various network architectures. In these techniques, sets of terminals relay signals for each other to create a virtual antenna array, trading off the costs—in power, bandwidth, and complexity—for the greater benefits gained by exploiting spatial diversity in the channel. By contrast, classical network architectures only employ point-to-point transmission and thus forego these benefits.

After summarizing a model for the wireless channel, we present various practical cooperative diversity algorithms based upon different types of relay processing and re-encoding, both with and without limited feedback from the ultimate receivers. Using information-theoretic tools, we show that all these algorithms can achieve full spatial diversity, as if each terminal had as many transmit antennas as the entire set of cooperating terminals. Such diversity gains translate into greatly improved robustness to fading for the same transmit power, or substantially reduced transmit power for the same level of performance. For example, with two cooperating terminals, power savings as much as 12 dB (a factor of sixteen) are possible for outage probabilities around one in a thousand. Finally, we discuss how the required level of complexity in the terminals makes different algorithms suitable for particular network architectures that arise in, for example, current cellular and ad-hoc networks.

Thesis Supervisor: Gregory W. Wornell
Title: Professor

Acknowledgments

Numerous people have supported me during the development of this dissertation, and my graduate experience more generally. A few words' mention here cannot adequately capture all my appreciation.

My advisor, Greg Wornell, deserves particular attention and many, many thanks. He has helped me reach several milestones as an academic, and prepared me to reach many more. I want to thank him for his inquisitiveness, insight, and inspiration. I have benefited tremendously from his countless efforts and undaunting dedication to the intellectual and personal growth of his graduate students. He has left an indelible mark.

I would also like to thank the other members of my dissertation committee, Hari Balakrishnan, Muriel Medard, and David N.C. Tse, for many useful interactions and for contributing their broad perspective in refining the ideas in this dissertation.

The financial support of the National Science Foundation, the U.S. Army Research Laboratory, and the Hewlett-Packard Company, through the HP-MIT Alliance, are also gratefully acknowledged.

The Digital Signal Processing Group (DSPG) has been a truly enriching and collegial environment for me as a graduate student, one I can only hope to emulate in the future. I thank Al Oppenheim and Greg Wornell for continually attracting such extraordinary students. Of the many who have contributed significantly to my M.I.T. experience over the years, I offer a word of particular appreciation to several from more recent years. Thanks go to Richard Barron, Aaron Cohen, Stark Draper, Emin Martinian, and Wade Torres for intense technical discussions, photo critiques, Toscanini runs, *Ishmael* debates, and more than a few basketball games; thanks also to Yonina Eldar and Andrew Russell for markedly non-technical discussions.

Of many other friends and colleagues, I want to thank Brian Hearing and Steve Ludwick for a number of well-timed and unforgettable hiking trips. Thanks also go to Sae-Young Chung, Thierry Klein, and Edmund Yeh for playing foosball matches instead of doing research at Bell Labs, and to Jay Gelman for playing golf instead of doing research at Lincoln Labs.

Never having had a brother, I have sought out several replacements through the years. None have been better to me than Giovanni Aliberti. I want to thank Vanni for his computer hacking and general wisdom, for taking me under his wing, and for making Mishie and I part of his family. Along with Darlete, Victor, Clarissa, and Martin, he makes leaving Boston especially bittersweet.

As always, I thank my family for inspiring me to pursue an academic career: Mom and Dad for giving the first lectures I attended, and Jenny and Jill for patiently listening to my earliest tutorials. I thank them, and the rest of my extended family, for "letting me do what I needed to do." Their encouragement, support, and understanding through these many years have meant more I can ever express. I GOT IT DONE!

Last, my most tender and sincere thanks go to my loving wife, Mishie. Aside from perhaps myself, no one has been more impacted by this dissertation and the latter years of my graduate program. I want to thank her for supporting me (and the two of us) in innumerable ways. This dissertation is not good enough to dedicate to her, but the one we are writing together will be.

" . . . And what's the good of diversity?"

"I don't know. It's certainly more . . . interesting."

— from *Ishmael*, by Daniel Quinn

Contents

1 Introduction **19**

1.1 Cooperative Diversity 19

1.2 Motivating Example 20

1.3 Layered Architectures and Cross-Layer Design 22

1.4 Outline of the Dissertation 24

2 High-Level System Model **27**

2.1 Radio Hardware and Constraints 29

2.2 Wireless Channel Impairments 30

2.2.1 Multipath Propagation: Path-Loss and Fading 30

2.2.2 Interference and Other Issues 33

2.3 Network Architectures 34

2.3.1 Prevalent Wireless Network Architectures 35

2.3.2 New Architectures Incorporating Cooperation 37

3 Background and Related Literature **41**

3.1 Relay Channels and Extensions 42

3.2 Fading Channel Capacity 45

3.2.1 Shannon Capacity 45

3.2.2 Capacity-vs.-Outage 48

3.2.3 Average Capacity 49

3.3 Multi-Antenna Systems 49

3.3.1 Fundamental Performance Limits 52

3.3.2 Space-Time Codes 54

3.4 Wireless Networks 54

3.4.1 Infrastructure Networks . . . 54
3.4.2 Ad-Hoc Networks . . . 55

4 Cooperative Diversity with Full Temporal Diversity 59

4.1 Model and Definitions . . . 60
4.1.1 Block Diagram . . . 61
4.1.2 Parameterization . . . 62
4.1.3 Special Cases and Coding Strategies . . . 63
4.1.4 Definitions . . . 65
4.2 Converse: Outer Bound on the Capacity Region . . . 68
4.3 Achievability: Inner Bounds on the Capacity Region . . . 71
4.3.1 Non-Cooperative Transmission . . . 71
4.3.2 Decode-and-Forward Transmission . . . 74
4.4 Rayleigh Multipath Fading . . . 75

5 Cooperative Diversity without Temporal Diversity 79

5.1 System Model . . . 81
5.1.1 Medium Access . . . 82
5.1.2 Equivalent Channel Models . . . 83
5.1.3 Parameterizations . . . 85
5.2 Cooperative Transmission Protocols . . . 86
5.2.1 Fixed Protocols . . . 87
5.2.2 Adaptive Protocols . . . 88
5.2.3 Protocols with Limited Feedback . . . 89
5.3 Outage Behavior . . . 90
5.3.1 Direct Transmission . . . 91
5.3.2 Fixed Protocols . . . 92
5.3.3 Adaptive Protocols . . . 94
5.3.4 Bounds for Cooperative Diversity Transmission . . . 96
5.3.5 Protocols with Limited Feedback . . . 98
5.4 Discussion . . . 100
5.4.1 Asymmetric Networks . . . 100
5.4.2 Symmetric Networks . . . 104

6 Cooperative Diversity in Networks **109**

6.1 Fully Cooperative Networks 109

6.1.1 System Model 112

6.1.2 Repetition-Based Cooperative Diversity 117

6.1.3 Space-Time Coded Cooperative Diversity 121

6.1.4 Diversity-Multiplexing Tradeoff 126

6.2 Partially Cooperative Networks 127

6.2.1 Centralized Partitioning for Infrastructure Networks 129

6.2.2 Clustering in Ad-Hoc Networks 135

6.2.3 Comments on Layering Issues 136

7 Conclusions **139**

7.1 Contributions 139

7.2 Future Research 141

A Coding Theorem for Cooperative Diversity **145**

A.1 Definitions 145

A.2 Preliminaries 147

A.3 Decode-and-Forward Transmission 148

A.3.1 Codebook Generation 149

A.3.2 Encoding 150

A.3.3 Backwards Decoding 151

A.3.4 Probability of Error and Achievable Rates 152

B Asymptotic CDF Approximations **159**

B.1 Results for Chapter 5 159

B.2 Results for Chapter 6 168

B.2.1 The Basic Result 168

B.2.2 Repetition Decode-and-Forward Cooperative Diversity 171

B.2.3 Space-Time Coded Cooperative Diversity 172

C Mutual Information Calculations **175**

C.1 Amplify-and-Forward Mutual Information 175

C.2 Input Distributions for Transmit Diversity Bound 176

List of Figures

1-1 Illustration of radio signal transmit paths in an example wireless network with two terminals transmitting information and two terminals receiving information. 21

1-2 Layered protocol architecture. 23

2-1 Block diagram for a wireless network model having four terminals. 28

2-2 Radio block diagram. 29

2-3 Discrete-time, baseband-equivalent channel model for signal received by radio j. 31

2-4 Protocol stack for a general layered network architecture. 35

2-5 Various relaying configurations that arise in wireless networks: (a) classical relay channel, (b) parallel relay channel, (c) multiple-access channel with relaying, (d) broadcast channel with relaying, (e) interference channel with relaying. 38

3-1 Block diagrams relating a point-to-point physical array (a) and a multi-user virtual array (b) arising from cooperative diversity transmission. 50

3-2 Multi-antenna system model. 51

4-1 Block diagram for a multiple-access channel with cooperative diversity. . . . 61

4-2 Block diagram for a relay channel. 64

4-3 Illustration of shape of various outer and inner bounds on the capacity region of the Gaussian multiple access channel with cooperative diversity. Here the SNRs take values $s_{0,1} = s_{0,2} = 3$, and $s_{2,1} = s_{1,2} = 15$. 72

5-1 Illustration of radio signal transmit paths in an example wireless network with two terminals transmitting information and two terminals receiving information. 81

5-2 Example time-division channel allocations for (a) direct transmission with interference, (b) orthogonal direct transmission, and (c) orthogonal cooperative diversity transmission protocols. We focus on orthogonal transmissions of the form (b) and (c) throughout the chapter. 83

5-3 SNR loss for cooperative diversity protocols (solid) and orthogonal transmit diversity bounds (dashed) relative to the (unconstrained) transmit diversity bound (0 dB). 101

5-4 Outage event boundaries for amplify-and-forward (solid) and adaptive decode-and-forward (dashed and dash-dotted) transmission as functions of the realized fading coefficient $|a_{s,r}|^2$ between the cooperating terminals. Outage events are to the left and below the respective outage event boundaries. Successively lower solid curves correspond to amplify-and-forward with increasing values of $|a_{s,r}|^2$. The dashed curve corresponds to the outage event for adaptive decode-and-forward when the relay can fully decode, *i.e.*, $\mathtt{SNR}_{\text{norm}}|a_{s,r}|^2 \geq 2$, and the relay repeats, while the dash-dotted curve corresponds to the outage event of adaptive decode-and-forward when the relay cannot fully decode, *i.e.*, $\mathtt{SNR}_{\text{norm}}|a_{s,r}|^2 < 2$, and the source repeats. Note that the dash-dotted curve also corresponds to the outage event for direct transmission. 103

5-5 Outage probabilities vs. $\mathtt{SNR}_{\text{norm}}$, small $\mathtt{R}$ regime, for statistically symmetric networks, *i.e.*, $\sigma_{i,j}^2 = 1$. The outage probability curve for amplify-and-forward transmission was obtained via Monte-Carlo simulation, while the other curves are computed from analytical expressions. The dashed curve corresponds to the transmit diversity bounds in this low spectral efficiency regime. 106

5-6 Diversity order Δ vs. $\mathtt{R}_{\text{norm}}$ for direct transmission and cooperative diversity protocols. 108

6-1 Illustration of the two-phases of repetition-based and space-time coded cooperative diversity algorithms. In the first phase, the source broadcasts to the destination as well as potential relays. Decoding relays are shaded. In the second phase, the decoding relays either repeat on orthogonal subchannels or utilize a space-time code to simultaneously transmit to the destination. . 111

6-2 Non-cooperative medium-access control. Example source allocations among m transmitting terminals across orthogonal frequency channels. 113

6-3 Repetition-based medium-access control. Example source channel allocations across frequency and relay subchannel allocations across time for repetition-based cooperative diversity among m terminals. 114

6-4 Space-time coded medium-access control. Example channel allocations across frequency and time for m transmitting terminals. For source s, $\mathcal{D}(s)$ denotes the set of decoding relays participating in a space-time code during the second phase. 114

6-5 Comparison of numeric integration of the outage probability (solid lines) to calculation of the outage probability approximation (6.10) (dashed lines) vs. normalized SNR for different network sizes $m = 1, 2, \ldots, 10$. Successively lower curves at high SNR correspond to larger networks. For simplicity of exposition, we have plotted the case of $\mathtt{R} = 1$ b/s/Hz and $\lambda_{i,j} = 1$; more generally the plot can be readily updated to incorporate a model of the network geometry. 120

6-6 Comparison of numeric integration of the outage probability (solid lines) to calculation of the outage probability approximation (6.19) (dashed lines) vs. normalized SNR for different network sizes $m = 1, 2, \ldots, 10$. Successively lower curves at high SNR correspond to larger networks. For simplicity of exposition, we have plotted the case of $\mathtt{R} = 1$ b/s/Hz and $\lambda_{i,j} = 1$; more generally the plot can be readily updated to incorporate a model of the network geometry. 124

6-7 Diversity order $\Delta(\mathtt{R}_{\text{norm}})$ for non-cooperative transmission, repetition-based cooperative diversity, and space-time coded cooperative diversity. As $\mathtt{R}_{\text{norm}} \rightarrow 0$, all cooperative diversity protocols provide full spatial diversity order m, the number of cooperating terminals. Relative to direct transmission, space-time coded cooperative diversity can be effectively utilized for a much broader range of $\mathtt{R}_{\text{norm}}$ than repetition-coded cooperative diversity, especially as m becomes large. 128

6-8 Matching algorithm performance in terms of average outage probability vs. received SNR (normalized for direct transmission). 133

6-9 Matching algorithm results for an example network: (a) minimal matching, (b) greedy matching. Terminals are indicated by circles, and matched terminals are connected with lines. 134

6-10 Clustering with (a) direct transmission and (b) cooperative diversity transmission. 135

A-1 Block-Markov encoding structure for decode-and-forward transmission. . . . 150

List of Tables

2.1 Form of $\mathsf{a}_{j,i}[l;k]$ for a several important channel conditions. 32

5.1 Summary of outage probability approximations. To capture the salient trade-offs between signal-to-noise ratio SNR, spectral efficiency R b/s/Hz, and diversity gain of the various protocols, the results are specialized to the case of statistically symmetric networks with fading variances $\sigma_{i,j}^2 = 1$. 105

Chapter 1

Introduction

Exploding demand for a growing number of wireless applications has fueled significant development of wireless networks, especially several generations of cellular voice and data networks and, more recently, ad-hoc data networks for wireless computer, home, and personal networking. Radio hardware and wireless services grow more efficient and cost effective as system designers better understand the channel environment and multi-user communications in general, and technological advances in integrated circuits and radio-frequency electronics increasingly allow for more sophisticated signal processing and channel coding algorithms. However, compared to point-to-point links, it seems we are only beginning to understand the fundamental performance limits of wireless networks and practical ways for approaching them. Moreover, given their impact on society as well as other technologies, wireless communications and networking remain important areas of research.

1.1 Cooperative Diversity

Taking advantage of the rich wireless propagation environment across multiple protocol layers in a network architecture offers numerous opportunities to dramatically improve network performance. In this dissertation, we develop an energy-efficient class of cross-layer network algorithms called *cooperative diversity* that exploit the broadcast nature and inherent spatial diversity of the channel. Through cooperative diversity, sets of wireless terminals benefit by relaying messages for each other to propagate redundant signals over multiple paths in the network. This redundancy allows the ultimate receivers to essentially average channel variations resulting from fading, shadowing, and other forms of interference.

By contrast, classical network architectures only employ a single path through the network and thus forego these benefits.

We develop various cooperative diversity algorithms that have the relays (1) simply amplify what each receives, or (2) fully decode, re-encode, and re-transmit each other's messages. In addition, we evaluate algorithms based upon repetition codes or more general space-time codes, as well as algorithms with and without limited feedback from the ultimate receivers. We demonstrate that cooperative diversity can provide full spatial diversity, as if each terminal had as many transmit antennas as the entire set of cooperating terminals. Such diversity gains translate into greatly improved robustness to fading for the same transmit power, or substantially reduced transmit power for the same level of performance. For example, power savings on the order of 12 dB (a factor of sixteen) are possible for outage probabilities around one in a thousand. Although applicable to any wireless setting, these algorithms are most beneficial when other forms of diversity—such as temporal coding, spread-spectrum, and multi-antenna systems—cannot be exploited.

From an architectural perspective, we illustrate how repetition and space-time coded cooperative diversity are each amenable to different settings. Repetition-based schemes require relatively low complexity in the terminals, but require more complexity in the network for deciding which terminals cooperate in order for the algorithms to be effective; thus, these algorithms are well-suited to infrastructure networks, *e.g.*, cellular, satellite, and certain wireless local area network (LAN) configurations, in which terminals communicate directly to a super-terminal that selects the cooperating groups. To manage complexity in the super-terminal, we use our analytical results to develop a variety of grouping algorithms based upon set partitioning and weighting matching in graphs. By contrast, space-time coded cooperative diversity requires more complexity in the terminals, but readily extends to distributed implementation; thus, these algorithms are well-suited to ad-hoc networks, and especially ad-hoc networks with clusters. We also briefly discuss various layering and other architectural issues.

1.2 Motivating Example

To illustrate the main concepts, consider the example wireless network in Fig. 1-1, in which terminals T_1 and T_2 transmit to terminals T_3 and T_4, respectively. This example might

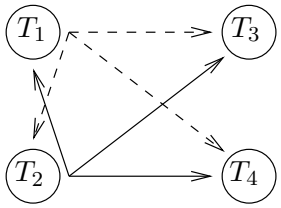


Figure 1-1: Illustration of radio signal transmit paths in an example wireless network with two terminals transmitting information and two terminals receiving information.

correspond to a snapshot of a wireless network in which a higher-level network protocol has allocated bandwidth to two users for transmission to their intended destinations or next hops. For example, in the context of a cellular network, T_1 and T_2 might correspond to terminal handsets and $T_3 = T_4$ might correspond to the basestation. As another example, in the context of a wireless LAN, the case $T_3 \neq T_4$ might correspond to an ad-hoc configuration among the terminals, while the case $T_3 = T_4$ might correspond to an infrastructure configuration, with T_3 serving as an access point. The key property of the wireless medium that allows for cooperative diversity between the transmitting radios is its broadcast nature: transmitted signals can, in principle, be received and processed by any of a number of terminals. Thus, instead of transmitting independently to their intended destinations, T_1 and T_2 can listen to each other's transmissions and jointly communicate their information. Although these extra observations of the transmitted signals are available for free (except, possibly, for the cost of additional receive hardware) wireless network protocols often ignore or discard them.

In the most general case, T_1 and T_2 can share their resources to cooperatively transmit their information to their respective destinations, corresponding to a wireless multiple-access channel with relaying for $T_3 = T_4$, and to a wireless interference channel with relaying for $T_3 \neq T_4$. At one extreme, corresponding to a wireless relay channel, the transmitting terminals can focus all their resources, in terms of power and bandwidth, on transmitting the information of T_1. In this case, T_1 acts as the "source" of the information, and T_2 serves as a "relay". Such an approach might provide diversity in a wireless setting because, even if the channel quality between T_1 and T_3 is poor, the information might be successfully transmitted through T_2. Similarly, T_1 and T_2 can focus their resources on transmitting the information of T_2, corresponding to another wireless relay channel.

To date, these channel models have been only partially addressed in the literature, and mainly within the information theory community. For general memoryless channels, the capacity region without cooperation is known for multiple-access channels, and remains an open problem for interference channels, although several achievable rate regions have been demonstrated [19]. Several notions of channel capacity, including Shannon, delay-limited, and information outage capacities, have been treated for wireless multiple-access channels [83, 34, 52]. Multiple-access channels with varying degrees of cooperation between the transmitting terminals have also been examined [88, 91, 89, 90]. Reference [17] examines certain relay channels, without specifically addressing wireless channels, and constructs several coding schemes for achieving reliable communication over such channels. Multiple-relay extensions have also been examined [66, 65, 33, 29, 28, 27]. As we will see, our work on cooperative diversity for wireless networks blends many techniques and insights from these references.

Recently, [68, 69, 70] considers cooperative diversity for a wireless multiple-access channel with relaying. The transmitters and receiver employ a two-user generalization of the cooperation scheme developed in [17], and, under the condition that the signal-to-noise ratio (SNR) between the transmitting terminals is high, cooperative diversity of this form is shown to enlarge the achievable rate region (including increasing the sum rate for the case in which both the transmitters and receivers possess knowledge of the channel conditions) in ergodic settings, as well as improve outage performance under strict delay constraints or in non-ergodic settings, when compared to the multiple-access channel without cooperative diversity [69, 70]. We develop some new insights for ergodic settings, but the majority of our results are for non-ergodic settings.

1.3 Layered Architectures and Cross-Layer Design

Several key properties of wireless environments make design of wireless networks particularly complex and challenging. First, radio signals experience significant attenuation, called *path-loss*, as well as self-interference, called *fading*, induced by multipath propagation through a lossy medium. Generally speaking, these channel distortions require increasing power, bandwidth, and receiver complexity to reliably communicate over longer distances. At the same time, radios in a wireless network share a common transmission medium, *e.g.*,

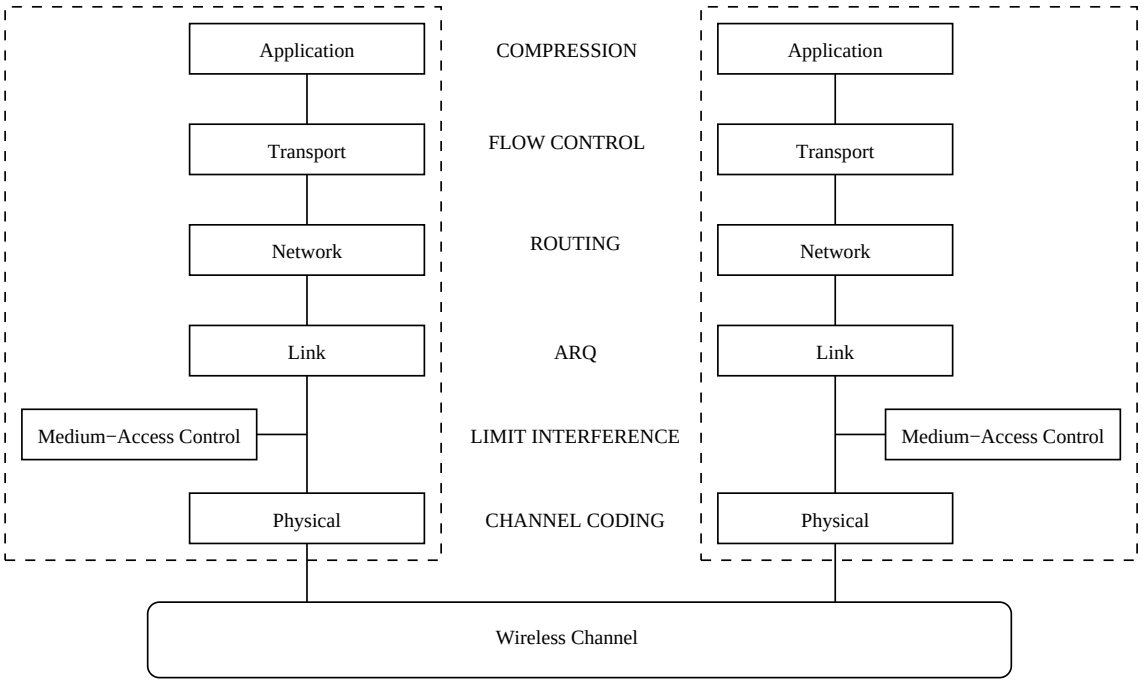


Figure 1-2: Layered protocol architecture.

a fixed amount of wireless spectrum; thus, radio signals are subject to *interference* from other users in the system as well as from other wireless systems operating in the same spectrum. Tradeoffs among required power, bandwidth, and receiver complexity naturally arise because of the interference characteristics of the channel.

As a result of this rich channel environment, wireless system designers are presented with many challenges. These include, for example: reliably transmitting information among radio terminals; mitigating severe channel impairments such as multipath fading and interference from other users; efficiently allocating and utilizing resources such as power and bandwidth; scaling algorithms as the number of terminals in the network grows; and supporting a large and ever-growing number of applications, such as voice, data, and multimedia networking.

Engineers have historically partitioned solutions to these problems into a stack of protocol layers, each serving a particular purpose. Fig. 1-2 illustrates these layers, and indicates the functions they usually serve in the wireless setting. As examples, the Medium-Access Control (MAC) Layer conventionally manages interference in the network, and the Physical (PHY) Layer conventionally combats fading with coding, spread-spectrum, and multiple antennas. Layering promotes development of understanding and technology within each

layer; however, different communities often work on problems at the various layers. For example, the data and computer networking community has developed a variety of standard protocols for flow control in the Transport Layer, routing in the Network Layer, automatic repeat request (ARQ) in the Link Layer, and collision resolution/avoidance in the Medium-Access Control Layer. Traditionally, the communication theory, information theory, and signal processing communities have played a role mainly at the Application Layer in the form of source coding and at the Physical Layer in the form of channel coding.

From a broader perspective, one can ask whether the particular layers and allocation of functions shown in Fig. 1-2 are appropriate, especially fading mitigation in the PHY Layer and interference management in the MAC Layer, and whether there is a more natural set of abstractions. Indeed, many new results seem to diffuse across the traditional protocol layers, especially in the wireless setting. This phenomenon is perhaps attributable to the rich channel environment, and presents opportunities for cross-fertilization of ideas among various research communities, from computer and data networking, to communications, information theory, and signal processing. We view cooperative diversity as involving various aspects of the physical, medium-access control, and network layers. More generally, returning to Fig. 1-2, the channel and source coding communities appear to be expanding up and down the protocol stack, respectively. Many interesting problem formulations have either appeared or re-emerged in the last ten years, bearing direct applications in wireless settings. These include, for example, various network source coding problems [26, 6, 8, 21], and network channel coding, medium access, and power control problems [9, 31, 32, 46, 83]. Moreover, the data and computer networking communities appear to be expanding throughout the layer hierarchy; see, for example, the work in [7, 15, 36, 37, 75] and references therein.

1.4 Outline of the Dissertation

This dissertation continues as follows. Chapter 2 describes a fairly general system model for considering cooperative diversity in wireless networks. Chapter 3 summarizes a variety of background literature that contributes insights and natural comparison points for our study. Our main discussion and results are contained in Chapters 4-6, with detailed mathematical development deferred to Appendices A-C. Chapter 4 treats cooperative diversity in ergodic settings, when full temporal diversity can also be exploited by the coding strategy. Chapter 5

treats cooperative diversity in non-ergodic settings, in particular when no temporal diversity can be exploited by the coding strategy. Both Chapters 4 and 5 examine the case of two cooperating terminals; Chapter 6 extends the algorithms of Chapter 5 to more than two cooperating terminals and discusses a variety of related networking issues. Finally, Chapter 7 summarizes our conclusions and points to areas for future research.

Chapter 2

High-Level System Model

This chapter summarizes the key ingredients in a network model for examining cooperative diversity and related problems. A useful model is rich enough to capture the significant effects observed in practice, yet tractable enough to lend itself to analysis and design. Any model incorporates simplifying assumptions, and these must be clearly stated and reasonably justified.

The model we describe here is quite general, allowing an uninitiated reader to develop a sense of context for the specific models that we emphasize later in the dissertation. Significant features of our specific models include: radio hardware constraints such as half-duplex operation, frequency-selective and slowly varying Rayleigh multipath fading, channel state information available only to the receivers, and Gaussian noise and other forms of interference.

The general multi-terminal network model of [19] serves as the basis for our model, with the addition of wireless channel effects such as path-loss, fading, and interference. Consider a collection of M radio terminals seeking to communicate information-bearing signals $\mathsf{w}_{i,j}$, *e.g.*, voice, music, images, binary data. Each terminal transmits a signal, denoted by $\mathbf{x}_i$, and receives a signal, denoted by $\mathbf{y}_j$. We leave the details of the structure of the transmit and receive signals to the specific instances of networks that we consider in later chapters. As we discuss in Section 2.3, many networks transmit digital representations of $\mathsf{w}_{i,j}$, thereby separating source and channel coding.

Fig. 2-1 depicts a block diagram for a wireless network with four terminals. It will be helpful to keep this general diagram in mind as we specialize it in later chapters.

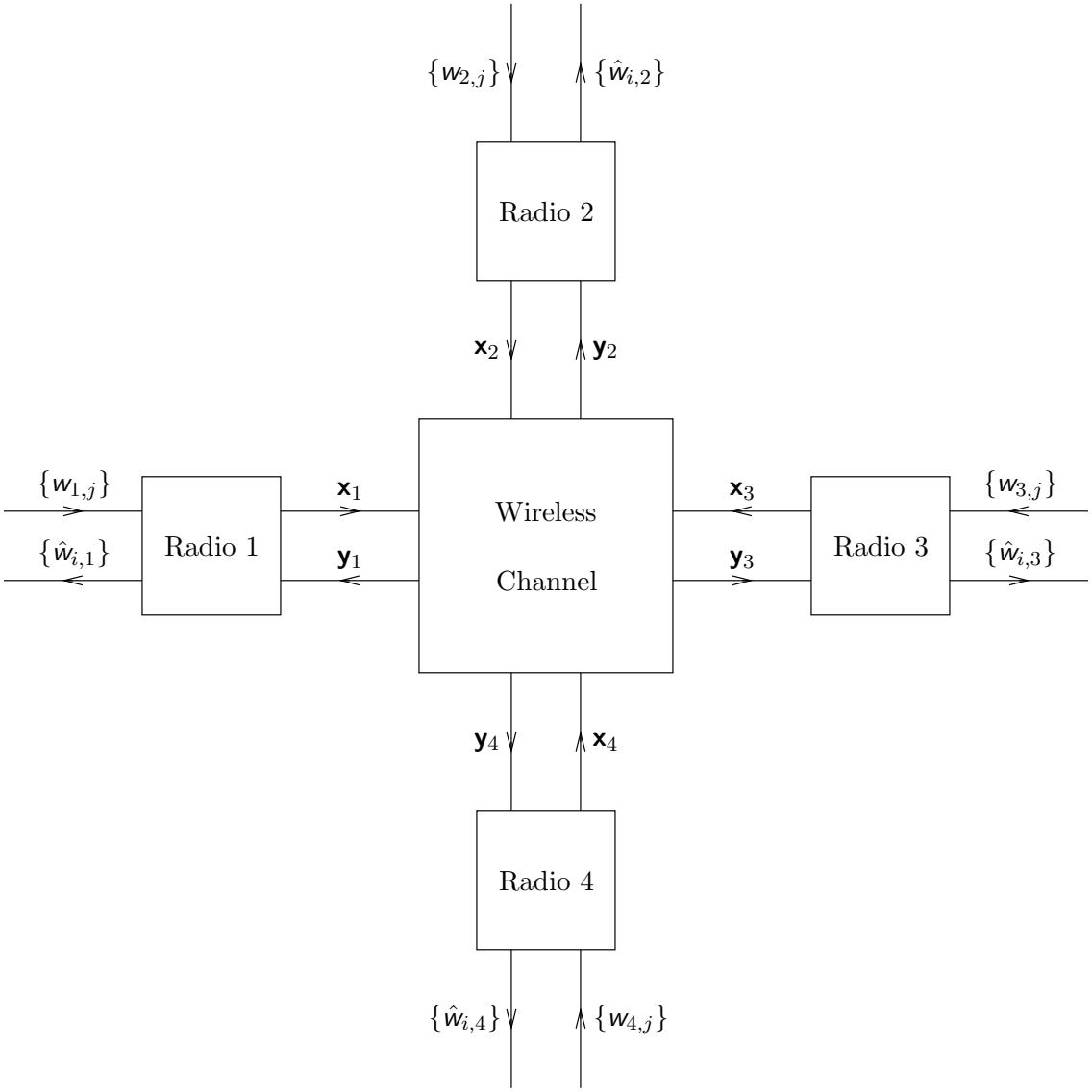


Figure 2-1: Block diagram for a wireless network model having four terminals.

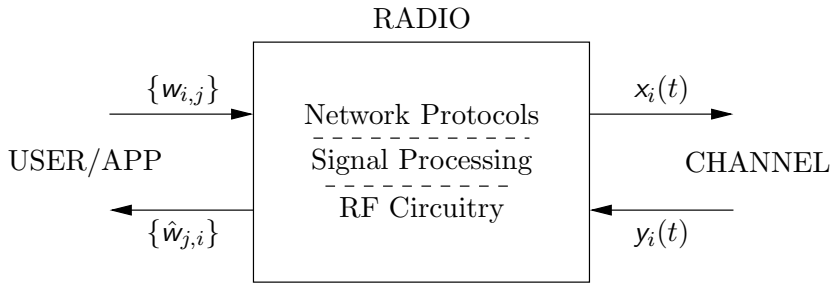


Figure 2-2: Radio block diagram.

Among many problem formulations for wireless communication, a fundamental one addresses the limiting tradeoffs among resources, computational complexity, and transmission quality, *e.g.*, end-to-end distortion, block error rates, and so forth. Just as important are questions of how to practically approach these fundamental tradeoffs. Problem formulations such as these can be made more specific by further developing the model. To this end, Section 2.1 describes radio hardware and its constraints. Section 2.2 describes the salient characteristics of wireless channels, including path-loss, fading, and interference. Finally, Section 2.3 discusses how the general problem is often simplified by imposing layered network architectures.

2.1 Radio Hardware and Constraints

In a general wireless setting, each terminal exchanges information with any of a number of other terminals in the system. Enabling such functionality requires a collection of radio devices that can each be viewed according to Fig. 2-2. Each radio consists of radio-frequency (RF) analog circuitry, and associated signal processing hardware and software, for emitting and observing information-bearing signals over the wireless channel, as well as distributed algorithms, or protocols, for coordinating the transmissions among the radios. In some systems, radios employ multi-antenna elements for increased capacity and improved robustness.

Regulatory restrictions and practical limitations on radio implementation lead to several system constraints in our system model. Among other possibilities, for example, regulatory

bodies often place an average power constraint

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} |x_i[k]|^2 \leq P_i, \tag{2.1}$$

on the transmitted signals. Furthermore, because of the near-far effect[1], it appears necessary to preclude radios from simultaneously transmitting and receiving on the same channel. This restriction constrains a radio so that

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_i[k] y_i[k] = 0 \ , \tag{2.2}$$

where the orthogonality can be imposed via time- or frequency-division between transmission and reception. Other constraints imposed by regulatory restrictions and implementation limitations include peak power and bandwidth constraints.

We note that we focus throughout the dissertation on algorithms that allow reductions in transmit power, often considered to be the dominant source of power consumption in wireless systems. Reductions in receiver power consumption can also improve terminal and network lifetime.

2.2 Wireless Channel Impairments

Our system model for wireless networks cannot be complete without capturing the salient effects of the wireless channels over which they operate. Indeed, many design decisions depend upon the particular channel conditions that prevail for a given application. In this section, we describe the significant channel distortions affecting wireless transmissions, and provide a fairly general mathematical description for use in our system models.

2.2.1 Multipath Propagation: Path-Loss and Fading

Wireless transmissions are severely degraded by the effects of so-called *multipath propagation*. A signal emitted by a radio antenna propagates, *e.g.*, in all directions (omnidirectionally) or, if the antenna is directed, only in a somewhat more restricted set of directions. Multipath arises because the propagated signal reflects off, refracts through,

[1]The *near-far* effect in this setting refers to a terminal's transmit signal drowning out the signals of other terminals at its receiver input because of path-loss and circuit isolation issues.

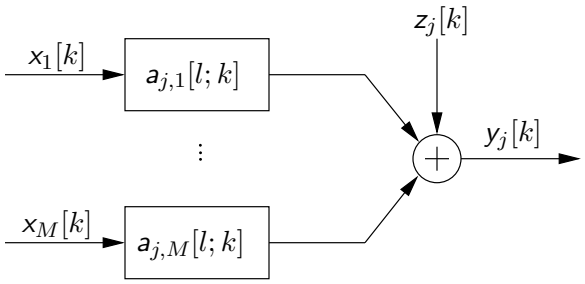


Figure 2-3: Discrete-time, baseband-equivalent channel model for signal received by radio j.

and diffuses around scattering objects in the channel environment. Example obstructions include, *e.g.*, buildings, trees, and cars in outdoor settings, and walls, furniture, and people in indoor settings. Scattering and propagation over longer distances increasingly attenuates signal power, an effect called *path-loss*. Thus, a radio receiver observes multiple attenuated and time-delayed versions of the transmitted signal, that are further corrupted by additive receiver thermal noise and other forms of interference. The copies of the transmitted signal might add constructively, thereby increasing the signal-to-noise ratio (SNR), or destructively, thereby decreasing the SNR. With relative motion of the transmitters, receivers, and scatterers in the channel environment, SNR fluctuations occur across both time and frequency, and are generally called *fading*. We model the effects of path-loss and fading as a time-varying linear filter.

Although radio transmissions are often continuous-time signals centered at carrier frequencies ranging from kHz to GHz, *i.e.*, passband signals, when the signals are bandlimited, it is often conceptually convenient to model them as discrete-time signals centered at 0 Hz, *i.e.*, baseband signals. Similarly, we model the continuous-time, passband channel effects with an associated discrete-time, baseband channel. Baseband-equivalent models are convenient because they suppress the issues of frequency up- and down-conversion, and discrete-time models are appealing because architectures designed for them can be efficiently implemented in digital signal processing (DSP) hardware.

Fig. 2-3 shows a fairly general, discrete-time, baseband-equivalent channel model for each of the received signals in a wireless network consisting of M terminals. For a given contiguous transmission bandwidth W, we use a baseband-equivalent, discrete-time channel model with W (complex) channel uses per second. Transmitter i emits signal $x_i[k]$, and

	Time Selective	Time Nonselective
Frequency Selective	$a_{j,i}[l;k]$	$a_{j,i}[l;0]\,\delta[k]$
Frequency Nonselective	$a_{j,i}[0;k]\,\delta[l]$	$a_{j,i}[0;0]\,\delta[l]\,\delta[k]$

Table 2.1: Form of $a_{j,i}[l;k]$ for a several important channel conditions.

receiver j observes signal $y_j[k]$. The effects of multipath propagation on $x_i[k]$ are modeled as convolution with the time-varying, discrete-time, linear filter $a_{j,i}[l;k]$, so that the received signals are modeled by the relationship

$$y_j[k] = \sum_{i=1}^{M} \sum_{l} a_{j,i}[l;k]\, x_i[k-l] + z_j[k], \qquad j = 1, 2, \ldots, M. \tag{2.3}$$

Here $z_j[k]$ captures the effects of receiver thermal noise and other forms of interference.

Multipath propagation manifests itself in a variety of ways depending upon the form of $a_{j,i}[l;k]$. For example, if $a_{j,i}[l;k] = a_{j,i}[\Delta;k]\,\delta[l-\Delta]$—a single, time-varying tap with delay Δ—the channel experienced by signal $x_i[k]$ in transmission to radio j is called time selective and frequency nonselective. Table 2.1 characterizes the form of $a_{j,i}[l;k]$ for several other important classes of wireless channels. The prevailing kind of channel can result from system constraints, *e.g.*, bandwidth limitations of regulatory bodies such as the FCC, or design choices, *e.g.*, allowing users to be mobile at varying speeds, or employing spread-spectrum signals [42, 60, 61].

Statistical Characterization

Because the scattering environment is often too complex for precise physical modeling to be tractable, system designers frequently employ statistical models for characterizing the channel effects. Such models are developed based upon the physics of radio-wave propagation and augmented with data obtained from measurements of real-world channels [42, 61].

Under the well-known Rayleigh fading model, the fading coefficients $a_{j,i}[l;k]$ are modeled as zero-mean, stationary complex jointly Gaussian random sequences in k that are independent for different values of i and j (with radio separations greater than roughly half the carrier wavelength, about 30 cm for a carrier frequency of 1 GHz) and sometimes independent for different values of l (called the *uncorrelated scattering* model). Temporal

correlation models have also been developed, *e.g.*, the Jakes model [61, 42].

We assign path-loss between radios i and j using models based upon the network geometry. For example, field measurements suggest path-loss models proportional to $G/d_{i,j}^{\nu}$, where $d_{i,j}$ is the distance between radio i and radio j, G captures the effects of antenna gain and carrier wavelength, and ν is a constant whose measured value typically lies in the range $3 \leq \nu \leq 5$ [61].

Channel State Information

An important issue affecting the design and analysis of transmissions protocols is channel state information, *i.e.*, how much radios know about each channel realization throughout the network. For example, using training signals, *e.g.*, pilot tones or symbols, the receivers may estimate the multipath coefficients affecting their respective received signals. Such channel measurement and estimation is reasonable when the channels are not over parameterized, *e.g.*, systems with small numbers of users transmitting at the same time in the same bandwidth within a given local area, channels exhibiting only a few significant non-zero taps, and slow enough temporal variations that allow estimation to provide accurate estimates.

Once channel state information is acquired at the distributed radio receivers, protocol designs can feed this information back to the transmitters. Feedback allows the transmitters to adapt their transmissions to the realized channel in effect, often leading to performance improvements when accurate channel state information is obtainable.

2.2.2 Interference and Other Issues

In addition to the salient channel effects such as path-loss and fading, and radio model variants such as multiple antennas at the transmitters and receivers, general wireless systems exhibit instances of many simpler channel models, including: multiple-access, *i.e.*, several transmitters conveying information to a common receiver; broadcast, *i.e.*, one transmitter convening information to several separate receivers; interference, *i.e.*, several transmitters conveying information to several separate receivers; and two-way, *i.e.*, two radios conveying information two one another. Each of these simpler examples exhibit different types of *interference* among multiple transmitted and/or received signals, and all of these types of interference can arise in a general wireless network.

Furthermore, any of these simpler models can incorporate *feedback*, *i.e.*, (partial) knowledge of the received signals at the transmitter, and *relaying*, *i.e.*, terminals without information to transmit or receive assist other transmitting and receiving terminals. These possibilities correspond to different ways for the terminals to interact. In addition to these interference and structural variations, networks might incorporate source-channel interactions that introduce additional possibilities. One such example is voice-activity detection in the IS-95 system [61]. Indeed, general wireless networks represent a huge space for designing and optimizing systems for various applications.

2.3 Network Architectures

Many different wireless network architectures have appeared in practice due to the broad array of applications and the varying extent to which providers desire to leverage existing infrastructure. For example, wireless network architectures for delivering voice or data and multimedia services tend to be quite different.

Often, these complicated systems are simplified along several dimensions to make the resulting problem formulations hierarchical and generally more tractable. For example, in a cellular setting, the coverage area is divided into "cells", each with its own basestation and a given amount of channel bandwidth. If the cell bandwidth is less than the total system bandwidth, then the system employs *frequency-reuse* to limit interference between cells. Within a cell, the bandwidth is often divided between a multiple-access uplink (mobiles to basestation) and broadcast downlink (basestation to mobiles). Finally, even the uplink and downlink channels are further divided into essentially orthogonal point-to-point channels to reduce receiver complexity. Feedback is somewhat limited, and relaying has not been employed in these systems to date. We call a set of restrictions on the structure of the network an *architecture*; thus, we have described a standard cellular architecture for delivering voice traffic.

Although design and implementation of wireless networks, and communication networks in general, can be approached in a variety of ways, many architectures are partitioned into a set of protocol layers [10, 61] as shown in Fig. 1-2. We include Fig. 1-2 here in abbreviated form as Fig. 2-4 for convenience. These layers range from the highest-layer (farthest from the wireless channel, most abstracted), the application layer, to the lowest layer (closest

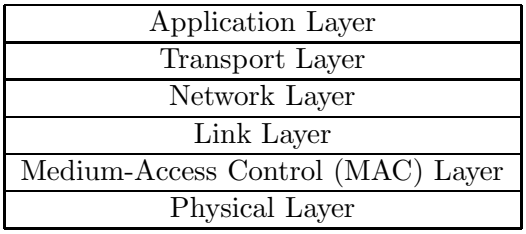


Figure 2-4: Protocol stack for a general layered network architecture.

to the wireless channel, least abstracted), the physical layer. Roughly speaking, the tasks allocated to each layer are as follows. The *application layer* generates or handles user signals $\{w_{i,j}\}$, and conveys them through an interface to the transport layer. The *transport layer* often performs packet sequencing, end-to-end retransmission, and flow control. The *network layer* routes messages through the network over a set of point-to-point links created by the link layer. The *link layer*, and its associated *medium-access control (MAC) sub-layer*, maintains a set of virtual point-to-point communication links built on top of the physical layer. Finally, the *physical layer* incorporates a majority of the analog circuitry and signal processing described in Section 2.1 and provides for transmission of signals $\{\mathbf{x}_i\}$ and reception and processing of signals $\{\mathbf{y}_j\}$ over the wireless channel.

The layered approach is convenient because it allows for abstraction barriers in the system. In particular, it allows for the design of general networks, from the network layer down, that support a variety of applications. The simplest example of this type is the separation of source and channel coding on a point-to-point link, where the interface between the two subsystems is bits. Furthermore, standardized layered architectures allow different manufacturers to provide components for different layers or complete radio implementations that inter-operate.

2.3.1 Prevalent Wireless Network Architectures

It will be convenient throughout the dissertation to classify network architectures into two broad classes called *infrastructure* and *ad-hoc* networks, respectively. Each of these has gone by different names in the past, but we borrow our terminology from current wireless local area network (LAN) standards [35] that allow for both configurations. We now describe these two classes of architectures.

Infrastructure Networks

In an infrastructure network, low-power, possibly mobile radio terminals connect via local, high-power, usually stationary radios, called *access points*, that are themselves connected via a backbone network. Typically, the backbone network is an existing wire-line network, such as the public switch telephone network or the Internet. Examples of such infrastructure networks include: current cellular networks, in which the basestations act as access points, for voice and data services; satellite networks, in which case the satellites act as access points, also for voice and data services, and certain wireless LANs, from which we have borrowed the term "access point". Two distinctive features of infrastructure networks are the power and processing asymmetries between radio terminals and access points, and the fact that all communication occurs through at least one access point. Even closely located radio terminals do not communicate directly, but instead communicate through their local access point.

Nominally, the architecture of an infrastructure network is organized as follows. The network layer assigns mobile radios to access points, using some information about channel conditions. Once a mobile is assigned to an access point, all communication occurs with that access point; thus, routing is not important within a cell. Sometimes, mobility requires handoff between access points, and this function is also handled by the network layer. This assignment of mobiles to access points creates a collection of "cells"; hence the name "cellular". The medium access control sub-layer at each access point allocates available channel bandwidth to mobiles assigned to it. Frequency reuse among the cells and orthogonal uplink (mobile to basestation) and downlink (basestation to mobile) transmissions limit interference. The physical layer codes and modulates for point-to-point transmission to and from the access point. Intra- and inter-cell interference is typically treated as noise, although multi-user detection can be employed to combat intra-cell interference.

Ad-Hoc Networks

In an ad-hoc (or packet radio, or peer-to-peer) network, radio terminals generally have more symmetric power and processing capabilities, and they do not leverage access points or a backbone network, as in infrastructure networks. Examples of ad-hoc networks include CB and amateur radio, emerging wireless LANs, and wireless sensor networks. In principle, the

distinctive features of ad-hoc networks include their potentially fast and ad-hoc deployment as well as their resulting robustness to loss of radio terminals. For these reasons, ad-hoc networks initially found application primarily in military settings, but are more recently penetrating certain commercial arenas such as home networking.

The layered architecture of ad-hoc networks is based largely upon wire-line store-and-forward networks. Terminals communicate primarily with nearby terminals, and exchange neighbor information to enable routing throughout the network. In contrast to infrastructure networks that employ direct wireless transmission between mobiles and a basestation, ad-hoc networks have gravitated toward cascade transmission—often called *multihop routing* in the ad-hoc networking literature—between source and destination terminals via several intermediate terminals. Cascade transmission potentially conserves energy by combating path-loss and limiting interference in the network.

Another architectural device frequently considered for ad-hoc networks is clustering. Clustering arises in a variety of forms in large, dense ad-hoc networks. (See [15, 37] and the references therein.) In essence, a clustering algorithm partitions a large ad-hoc network into a set of *clusters*, each centered around a *clusterhead*. Terminals communicate directly to their associated clusterhead, and routing is usually performed between clusterheads. In this sense, clustering mimics some of the features of infrastructure networks: clusters correspond to cells and clusterheads correspond to access points. However, in ad-hoc settings the clusters and clusterheads may be varying as the network operates, the clusterheads themselves can have information to transmit, and the clusterhead network must share the wireless bandwidth.

2.3.2 New Architectures Incorporating Cooperation

Many options arise for cooperative diversity in wireless settings. Fig. 2-5 depicts block diagrams for a number of these options. The configurations in Fig. 2-5 specialize to well known channel models when cooperation is not employed. For example, as we discuss in Section 3.4, the classical relay channel in Fig. 2-5(a) specializes to direct transmission when the relay is removed, and cascade transmission when the destination cannot receive (or ignores) the source transmission. The configurations in Fig. 2-5(c)-(e) specialize to the classical multiple-access channel, broadcast channel, and interference channel, respectively. Of these configurations, we focus throughout the dissertation on the multiple-access and

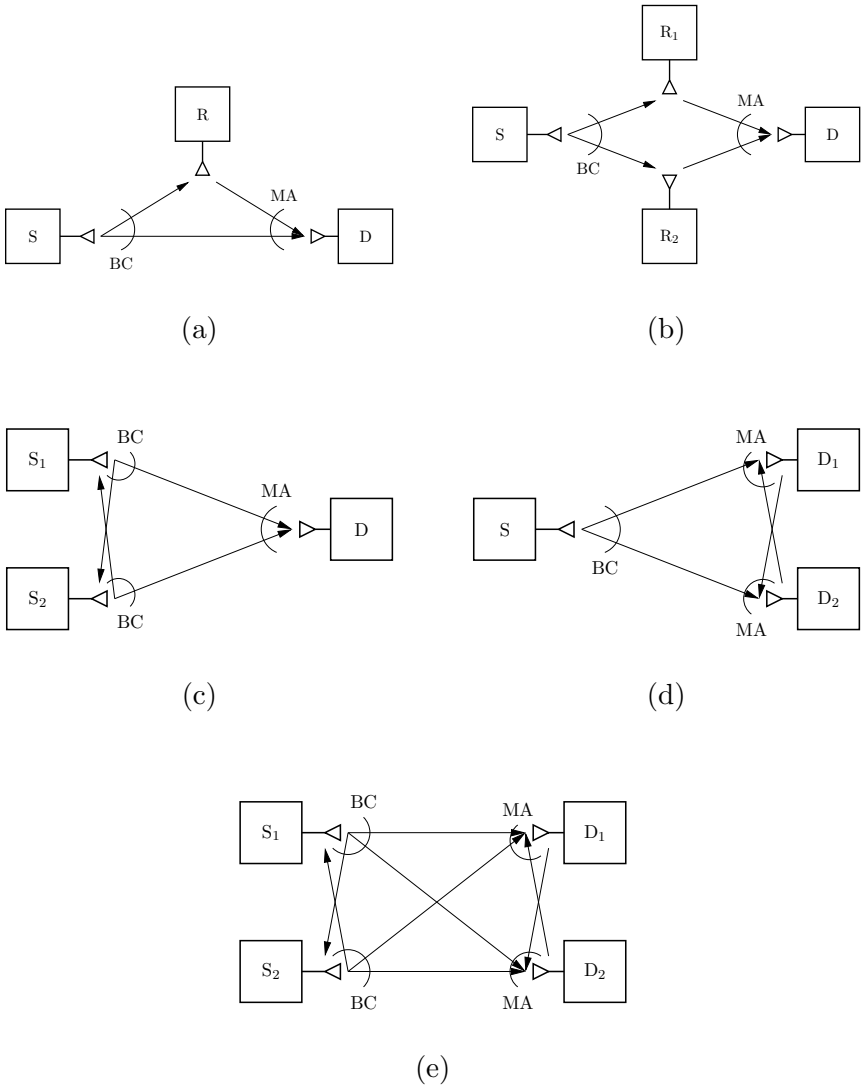


Figure 2-5: Various relaying configurations that arise in wireless networks: (a) classical relay channel, (b) parallel relay channel, (c) multiple-access channel with relaying, (d) broadcast channel with relaying, (e) interference channel with relaying.

interference channel versions of Fig. 2-5(c) and Fig. 2-5(e), respectively. We include the other possibilities for completeness as well as to encourage their study.

Chapter 3

Background and Related Literature

This chapter summarizes important references from a broad array of literature that relate to the problems studied by the dissertation. Our objective is to make the reader aware of the many considerations involved, highlight the particular scenarios that we study throughout the dissertation, and encourage further work in the area. Because cooperative diversity is a network problem, it can be viewed as living across several of the layers in a layered network architecture as discussed in Section 2.3. While the ideas build significantly from work on the relay channel within the information theory community, there are several other bodies of research to build from and relate to, including single-user multi-antenna systems and various results for ad-hoc networks, specifically in multihop routing.

Section 3.1 begins by summarizing work on the classical relay channel, the simplest cooperative example in Fig. 2-5. Evaluating the utility of cooperation in wireless systems starts with considering the issue of fading. Section 3.2 summarizes relevant notions of channel capacity in fading environments. It then seems natural to compare performance of cooperative transmission and reception protocols for creating virtual arrays to the performance of physical antenna arrays. Section 3.3 surveys important results from the physical array literature. Finally, Section 3.4 cites several interesting results that have studied wireless networks with an eye toward breaking barriers within the standard layered network architectures in order to improve performance.

In principle, fundamental performance limits for general wireless systems can be developed. Most generally, results would indicate the limiting tradeoffs between achievable channel rates and distortions on the information-bearing signals, without regard to com-

putational complexity. However, surprisingly many of even the simplest network channel models have not been yet fully characterized. For example, the capacity for general relay channels, broadcast channels, and interference channels remain unknown [19]. Since a general wireless system generalizes these models, as well as all the models shown Fig. 2-5, there is much work to be done.

3.1 Relay Channels and Extensions

Relay channels and their multi-terminal extensions are central to our study of cooperative diversity. Much of the work on these channel models to date has focused on discrete or additive white Gaussian noise channels, and examined performance in terms of the well-known Shannon capacity (or capacity region) [19]. Only the more recent work has considered the issue of multipath fading, which is another central issue of this dissertation. We also focus on the multi-terminal aspects of relaying or cooperative diversity problems, in particular, the multiple-access channel case and, to some extent, the interference channel case.

The classical relay channel models a class of three terminal communication channels (*cf.* Fig. 2-5(a)), originally introduced and examined by van der Meulen [84, 85], and subsequently studied by a number of authors, primarily from the information theory community. The distinctive property of relay channels in general is that certain terminals, called "relays", receive, process, and re-transmit some information bearing signal(s) of interest in order to improve performance of the system. As we illustrated in Fig. 2-5, in some cases extra terminals in the network, without information to transmit or receive, serve as relays, while in other cases transmitting and/or receiving terminals can cooperate by serving as relays for one another.

Cover and El Gamal [17] examine certain non-faded relay channels, developing lower and upper bounds on the channel capacity via random coding and converse arguments, respectively. Generally these lower and upper bounds do not coincide, except in the class of *degraded* relay channels [17]. While the class of degraded relay channels is mathematically convenient, we stress that none of the wireless channels found in practice fall into this class.

The lower bounds on capacity, *i.e.*, achievable rates, are obtained via three structurally different random coding schemes, referred to in [17] as *facilitation*, *cooperation*, and *obser-*

vation, respectively.[1]. The facilitation scheme is nothing special: the relay does not actively help the source, but rather, facilitates the source transmission by inducing as little interference as possible. The cooperation and observation schemes are more involved, as we now describe.

In the cooperation scheme of [17], the relay fully decodes the source message, and retransmits some information about that signal to the destination. More precisely, the relay encodes the bin index of the previous source message, from a random binning of the source messages as in well-known Slepian-Wolf coding [19]. The source transmits the superposition of a new encoded message and the encoded bin index of the previous message, in a block-Markov fashion. The destination suitably combines the source and relay transmissions, possibly coherently combining the identical bin index transmissions, in order to achieve higher rates than with the direct transmission alone. We note that practical implementations of this cooperation scheme can be obtained with suitable configurations of multi-level codes [87].

Of course, full decoding at the relay can, in some circumstances, be a limiting factor; the rates achieved using this form of cooperation are no greater than the capacity of direct transmission from the source to the relay. As one alternative in such circumstances, Cover and El Gamal propose the observation scheme, in which the relay encodes a quantized version of its received signal. The destination combines information about the relay received signal with its own in order to form a better estimate of the source message. For Gaussian noise channels, the destination can essentially average to two observations of the source message, thereby reducing the noise.

Broadly speaking, we can expect cooperation (resp. observation) to be most beneficial when the channel between the source and relay (resp. relay and destination) is particularly good. For intermediate regimes, Cover and El Gamal propose superposition of the two schemes in order to maximize the achievable rates.

Of the remaining configurations depicted in Fig. 2-5, only parallel relay channels (*cf.* Fig. 2-5(b)) and multiple-access channels with relaying (*cf.* Fig. 2-5(c)) have received attention in the literature. Schein and Gallager [66] introduced the parallel relay channel model in an attempt to make the classical relay channel symmetric. Schein [65] considers a

[1]The names facilitation and cooperation were introduced in [17], but the authors did not give a name to their third approach. We use the name *observation* throughout the dissertation for convenience

number of coding techniques for various regimes, and develops tighter converse results for certain discrete alphabet channels based upon ideas from distributed source coding [8].

Willems and others [88, 91, 89, 90] have examined the multiple-access channel with varying degrees of cooperation and generalized feedback between the transmitting terminals. Kramer and Wijngaarden [48] study a multiple-access channel model in which the mobiles share a common relay between themselves and the basestation. Sendonaris *et. al* [68, 69, 70] consider cooperative diversity for a multiple-access channel with relaying and fading, building upon the earlier work of Willems [91]. The transmitters and receiver employ a two-user generalization of the cooperation scheme developed in [17], and, under the condition that the signal-to-noise ratio (SNR) between the transmitting terminals is high, cooperative diversity of this form is shown to enlarge the achievable rate region (including increasing the sum rate for the case in which the transmitters and receivers possess knowledge of the fading) for ergodic fading, as well as improve outage performance under strict delay constraints or for non-ergodic fading [69, 70]. Most recently, a variety of results for extensions to multiple relays have appeared in the work of Gupta, Gastpar, and others [33, 29, 28, 27].

These studies offer techniques for analyzing certain multi-terminal communications problems, and suggest coding and decoding strategies that can be appropriately reduced into practice. There appear to be two general classes of approaches to relay processing and re-transmission. In one class, the relay decodes the source message and re-transmits some information about the message. We refer to techniques in this class broadly as *decode-and-forward* schemes. For example, the relay might decode the message and simply repeat the transmission, as in regenerative repeaters, or it might transmit additional parity bits about the message, as in the cooperation scheme of [17]. In the other class, the relay tries to convey a representation of its received signal to the destination, so that the destination can effectively combine two receive signals and decode the message. We refer to this second class broadly as *observe-and-forward* schemes. For example, the relay might simply amplify its received signal, as in amplifying repeaters [66], or it might quantize or rate-distortion code its received signal and encode for transmission to the destination [17]. As we have mentioned, superposition of the two classes was proposed in [17].

3.2 Fading Channel Capacity

Because of the wide variety of conditions under which wireless systems—even single-user point-to-point links—can operate, several notions of capacity have been developed for fading channels. These include Shannon (ergodic) capacity, capacity-vs.-outage (with delay-limited capacity as a special case), and average capacity. A very nice and more in-depth review of these notions of capacity is given by Berry [9].

Two issues that significantly influence the capacity notions are the extent to which the fading varies during a coding interval, or the degree to which temporal diversity may be exploited, and the amount of channel state information available at the transmitters and receivers. Of the various settings that we mention in this section, we stress that we focus throughout the rest of the dissertation scenarios in which accurate channel state information is available at the receivers but not the transmitters. While we address ergodic fading environments in which temporal diversity can be fully exploited in Chapter 4, we focus on non-ergodic fading environments in which no temporal diversity can be exploited throughout the remainder of the dissertation in Chapters 5 and 6.

To briefly describe the various notions of capacity, consider a single-user additive white Gaussian noise (AWGN) channel with frequency nonselective fading. We model the channel in complex baseband-equivalent form as

$$y[n] = a[n]\,x[n] + z[n]\ , \tag{3.1}$$

where $x[n]$ is the transmitted signal, $a[n]$ captures the effects of multipath fading, and $z[n]$ captures the effects of receiver thermal noise and other forms of interference.

3.2.1 Shannon Capacity

In this section, we discuss Shannon capacity for several cases of the model in (3.1).

Ergodic Fading, Full Temporal Diversity

When $a[n]$ corresponds to a stationary and ergodic process, its ergodic structure emerges if coding is performed over long blocklengths, and Shannon capacity becomes a useful measure of the maximum rate of reliable communication over the channel. In terms of diversity, we

note that these temporal variations allow the coding strategy to fully exploit temporal diversity.

Reliable communication is in the sense of codeword error probabilities approaching zero asymptotically. Several different quantities for Shannon capacity (and their associated coding and decoding schemes) arise depending upon whether the receiver or transmitter obtain fading state information.

For example, if only the receiver measures the fading process to high accuracy, then the fading can be viewed as an additional channel output, and the mutual information between input and output may be written as

$$\begin{aligned} I(\mathsf{x};\mathsf{y},\mathsf{a}) &= I(\mathsf{x};\mathsf{a}) + I(\mathsf{x};\mathsf{y}|\mathsf{a}) \\ &= I(\mathsf{x};\mathsf{y}|\mathsf{a}) \\ &= \mathrm{E}\left[I(\mathsf{x};\mathsf{y}|\mathsf{a}=a)\right] \ , \end{aligned} \tag{3.2}$$

where the first equality results from the well-known chain rule for mutual information, the second equality results from the fact that the transmit signal x is independent of the fading process a, and the third equality results from the definition of conditional mutual information. In the case of $\mathsf{z}[n]$ being i.i.d. complex Gaussian with variance N_0 and $\mathsf{x}[n]$ being i.i.d. complex Gaussian[2] with variance P, the mutual information in (3.2) becomes the channel capacity [22]

$$C_{\mathrm{CSIR}} = \mathrm{E}\left[\log\left(1+\frac{|\mathsf{a}|^2\,P}{N_0}\right)\right] \ , \tag{3.3}$$

which can be computed using the stationary distribution of the fading process $\mathsf{a}[n]$.

If the receiver can share channel state information with the transmitter, *e.g.*, by means of a separate feedback channel, then the transmitter can adapt $\mathsf{x}[n]$ to the channel states. A simple adaptation rule would be to conserve power by not transmitting when the channel SNR falls below a certain threshold, and transmit with high power when the SNR lies above the threshold. Any adaptations must be performed subject to appropriate average or peak power constraints. More generally, Shannon capacity is achieved in the Gaussian noise case by a "water-pouring" power allocation [19] over the fading states [30]. In this case, the

[2]The choice of $\mathsf{x}[n]$ complex Gaussian achieves capacity under an average power constraint of P.

Shannon capacity expression can be written in the form [30]

$$C_{\text{CSIR,CSIT}} = \max_{P(\cdot)} \mathrm{E}\left[\log\left(1 + \frac{|\mathsf{a}|^2 P(\mathsf{a})}{N_0}\right)\right] , \tag{3.4}$$

where $P(\cdot)$ represents the power allocation function, subject to, *e.g.*, an average power constraint $\mathrm{E}[P(\mathsf{a})] \leq P$. The case of state information available to the transmitter causally can be addressed [9], but the resulting capacity expressions are too complicated to lend much insight.

The case of channel state information available to neither to the transmitter nor the receiver has also been examined. Abou-Faycal *et. al* [1, 2] were the first to examine this case. While no convenient expressions for capacity were obtained, the authors prove the interesting result that capacity achieving input distributions are discrete and have value zero as a high probability input.

Finally, we note that all of the above results suggest that fading channel capacity with additive white Gaussian noise increases as log(SNR) for high SNR. Another interesting set of results suggest that fading channel capacity instead increases as log(log(SNR)). While we focus throughout the dissertation on the prior framework, it is important to keep these emerging results in mind.

Non-Ergodic Fading, No Temporal Diversity

In certain circumstances, delay constraints on the system may prevent the fading process from revealing its ergodic structure within the coding interval. Or more generally, the fading process may be non-ergodic, as in the case of completely stationary environments whose geometries are unknown at system design time. In terms of diversity, we note that lack of temporal variations in the channel prevent the coding strategy from exploiting temporal diversity. In these cases Shannon capacity is not a useful performance measure for system design because it is often zero.

To see why this is so, let us consider the non-ergodic fading process $\mathsf{a}[n] = \mathsf{a}$ for all n in the model (3.1). In this case, the wireless channel is appropriately modeled as a family of channels indexed by the value of $\mathsf{a} = a$, *i.e.*, a *compound channel* [50]. The Shannon

capacity of the compound channel can be shown to be

$$C_{\text{compound}} = \max_{p_{\mathsf{x}}(x)} \inf_{a} I(\mathsf{x}; \mathsf{y}|a) \tag{3.5}$$

The problem with Shannon capacity in this context is that, for families in which a can be very close to 0, as with Rayleigh fading, the Shannon capacity (3.5) is arbitrarily small or zero. Thus, Shannon capacity is not a useful tool for system design in such scenarios.

Essentially, Shannon capacity breaks down because we cannot guarantee reliable communication of any fixed, non-zero rate *a priori* since the realized channel SNR may not support that rate. There are several approaches to developing useful performance measures in such cases, and we briefly describe capacity-vs.-outage and average capacity next. Both rely on a probability distribution over the family of channels, corresponding to a *composite channel* framework [50].

3.2.2 Capacity-vs.-Outage

The notion of capacity-vs.-outage examines the tradeoff between a fixed rate and the probability that rate is achievable over the composite channel. For example, continuing with our non-ergodic Gaussian fading channel discussed above, for a fixed rate R certain channel realizations will support the rate, *i.e.*, those with

$$\log\left(1 + \frac{|a|^2 P}{N_0}\right) \geq R \ ,$$

and other channel realizations will not support the rate, *i.e.*, those with

$$\log\left(1 + \frac{|a|^2 P}{N_0}\right) < R \ .$$

The event $\log\left(1 + |\mathsf{a}|^2 P/N_0\right) < R$ is referred to as an *outage event*, and the probability of this event is referred to as the *outage probability* of the channel. Since generally R_0 achievable implies R achievable for all $R \leq R_0$, we expect the outage probability to be a non-decreasing function of R. In the Gaussian case, this can be readily shown because the outage probability for successively larger values of R corresponds to the cumulative distribution function of the fading random variable for successively larger arguments. Finally, the *capacity-vs.-outage* is defined to be the maximum rate with outage probability less than

some level. Delay-limited capacity is the special case of capacity-vs.-outage corresponding to zero outage.

Capacity-vs.-outage was introduced by Ozarow, Shamai, and Wyner [59] to examine the performance of certain cellular systems with delay constraints. It is intimately related to the more general and precise ϵ-capacity framework of Verdu and Han [86], and this relationship was solidified in the work of Caire, Taricco, and Biglieri [13]. Both [59, 13] extend the notion (beyond our simple example outlined in the above discussion) to handle block-fading models with delay constraints limiting the number of blocks available for transmission.

3.2.3 Average Capacity

In the capacity-vs.-outage framework, coding and modulation are performed at some fixed and pre-specified rate, and given the channel realization, this rate is either achievable or unachievable. Another option for coding and modulation would be to code for a monotonically increasing set of rates, *e.g.*, by means of general superposition codes for the broadcast channel [19]. Depending upon the channel realization, rates only up to a certain point are achievable. In this case, one can design the coding scheme to maximize the expected achievable rate. This average capacity framework may only be useful if paired with appropriate source-coding techniques, such as successive refinement coding [62]. Such an approach was originally proposed for fading channels by Shamai [71].

3.3 Multi-Antenna Systems

Recently there has been great interest in the use of multi-antenna physical arrays at the transmitters and/or receivers in a wireless system. Physical arrays offer space diversity to combat fading, or when sufficient knowledge of the channel conditions are available at both the transmitter and receiver, offer beamforming to combat both fading and interference from other terminals, and other wireless systems in the same band. As a result, physical arrays increase capacity and improve robustness to fading. Motivated by these possible gains, a great deal of research effort has focused on design of practical space-time codes and their associated decoding algorithms. Several studies have shown that, aside from suitable encoding and decoding algorithms, the key to leveraging spatial diversity with physical arrays is to have separation among the antennas on the order of several (3-10)

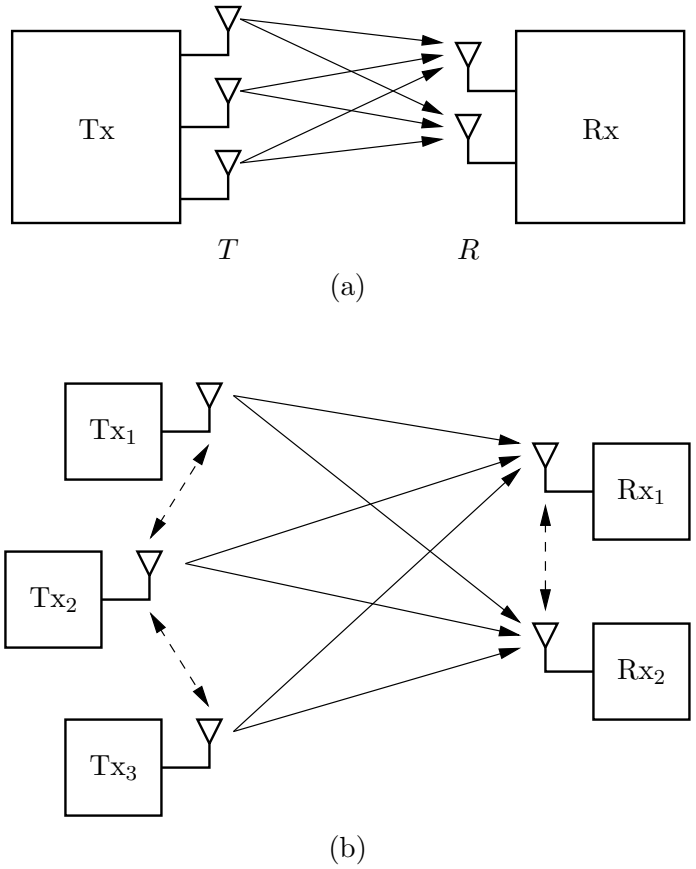


Figure 3-1: Block diagrams relating a point-to-point physical array (a) and a multi-user virtual array (b) arising from cooperative diversity transmission.

wavelengths of the carrier frequency so that the fading coefficients are uncorrelated. As carrier frequencies increase, this constraint becomes less restrictive; however, terminal size also decreases with time and circuit integration, thereby limiting the number of antennas that can be effectively placed in a transmitter or receiver.

For systems in which size constraints limit the number of antennas that can be placed in the transmitters or receivers, our research examines issues associated with creating a virtual array by allowing multiple users to cooperate and effectively share their antennas. Fig. 3-1 compares block diagrams for physical and virtual arrays. While multi-antenna array problems are generally treated at the physical layer, virtual arrays can be dealt with at a variety of layers, including interaction across layers.

Clearly, much can be gained from comparing virtual arrays to physical arrays, as in Fig. 3-1. As we exploit in Chapters 4 and 5, the performance of physical array systems

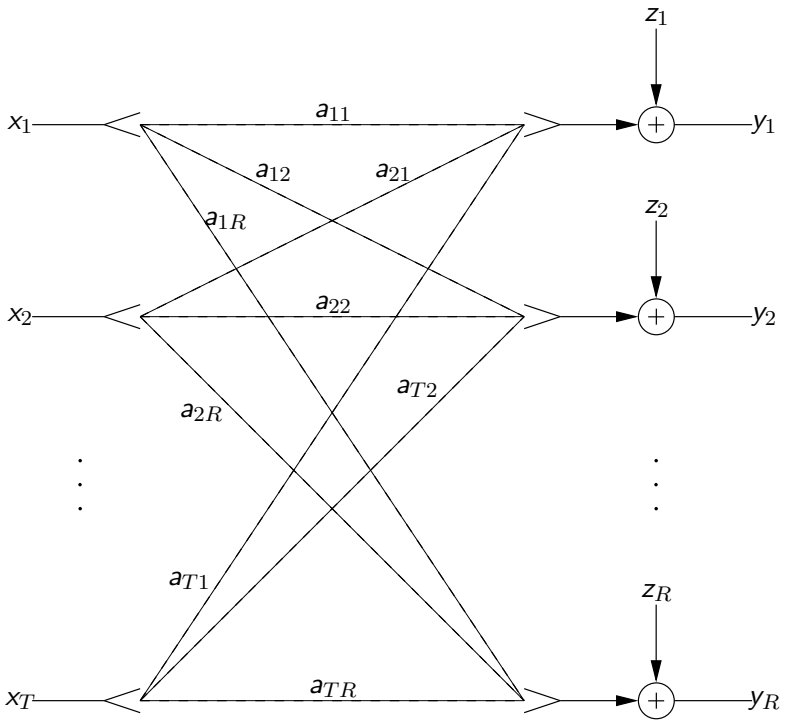


Figure 3-2: Multi-antenna system model.

provides useful performance bounds for virtual array systems. Furthermore, space-time code designs for physical arrays can be readily adapted to cooperative settings.

Fig. 3-2 shows a general model for multi-antenna systems utilizing T transmit and R receive antennas. The model can be expressed in vector form as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z} \tag{3.6}$$

where $\mathbf{A}$ is a $R \times T$ matrix, and $\mathbf{y}$ and $\mathbf{z}$ (resp. $\mathbf{x}$) are column vectors of size $R \times 1$ (resp. $T \times 1$). Here the element $[\mathbf{A}]_{r,t} = a_{r,t}$ captures the effects of multipath fading between transmit antenna t and receiver antenna r, while z_r captures the effects of receiver thermal noise and other forms of interference. Note that the multi-antenna model of Fig. 3-2 is a special case of the general wireless network (*cf.* Chapter 2) consisting of a single transmitter and receiver, with vector inputs and outputs, respectively.

There has been great and growing interest in channels of the form shown in Fig. 3-2. Initially, attention focused on systems with multiple receiver antennas and their associated

diversity combining algorithms, *e.g.*, maximum-ratio and selection combining, and array processing techniques, *e.g.*, beamforming and interference mitigation, [42, 60], but more recently systems employing multiple transmitter antennas, possibly with multiple receiver antennas, have been emphasized. Transmit antenna arrays generally require more sophisticated algorithms than receive antenna arrays alone, both because different signals can be transmitted from the multiple antennas and because these signals superimpose at the receiver antennas. Substantial energy has focused on characterizing the ultimate limits on performance for multi-antenna systems, and designing practical coding and decoding algorithms that approach these limits.

3.3.1 Fundamental Performance Limits

Of late, there has been substantial work characterizing the limiting performance of multi-antenna systems under a variety of fading conditions. For example, as discussed in Section 3.2.1, for systems without delay constraints and with sufficient fading variability (ergodicity), within the coding interval, classical Shannon theory provides the capacity of the channel. The Shannon, or ergodic, capacity for the channel model in Fig. 3-2 has been developed for several different cases of channel state information available to the transmitter and/or receiver: no channel state information [1, 2, 55, 94]; channel state information available to the receiver only [79, 57]; state information available to both transmitter and receiver [79].

Shannon (Ergodic) Capacity

The ergodic capacity results to date suggest that dramatic increases in capacity are possible using multi-antenna systems. For example, for the case of channel state information available to the receiver only, the ergodic capacity increases by $\min\{T, R\}$ b/s/Hz for each additional 3 dB of SNR, in the high SNR regime [79].[3]

For the case of no channel state information at either the transmitter or receiver, the channel capacity depends upon the number of transmit and receive antennas as well as the coherence time K of the channel, defined to be the number of samples for which the

[3]The ability to employ beamforming when channel state information is additionally available at the transmitter further provides an SNR gain, but the slope of the capacity function remains $\min\{T, R\}$ b/s/Hz for each additional 3 dB of SNR in the high SNR regime [79, 11].

channel remains constant in the assumed block fading model before it changes to another independent realization. In this case, the ergodic capacity has been shown to increase as $T'(1 - T'/K)$ b/s/Hz for each additional 3 dB of SNR in the high SNR regime, where $T' = \min\{T, R, \lfloor K/2 \rfloor\}$ [94]. The slope is maximized by employing $T = \lfloor K/2 \rfloor$ transmit antennas, assuming $R \geq T$, and in fact degrades if more than this number of transmit antennas is utilized. In this case, the capacity increases as $T/2$ b/s/Hz for each additional 3 dB of SNR.

As a point of reference, the capacity of an AWGN channel (without fading) increases by only 1 b/s/Hz for each additional 3 dB of SNR in the high SNR regime. Thus, quite large spectral efficiencies can, in principle, be achieved using multi-antenna systems. Adding antenna elements, along with suitable transmitter coding and receiver processing methods, is akin to adding cabling in a wire-line setting.

Capacity-vs.-Outage, Delay-Limited Capacity

As discussed in Section 3.2.2, for systems with tighter delay constraints, the channel may not exhibit its ergodic nature within a coding interval, so that the Shannon capacity is zero. In such cases, alternative performance metrics such as capacity-vs.-outage/outage probability [59, 12] or delay-limited capacity [34] can be employed to evaluate the efficacy of multi-antenna schemes. The work of Foschini and Gans [25] treats the case of burst transmissions and examines multi-antenna outage probability. Telatar [79] performs a similar analysis, and provides analytic forms of the outage probabilities. Narula, Trott, and Wornell [57] also utilize outage to compare the performance of practical multi-antenna schemes, several of which are mentioned below. Finally, Biglieri, Caire, and Taricco [13, 11] examine the outage probability when coding across a finite number of independent channel realizations. Foschini and Gans, Telatar, and Narula *et. al* treat the case of channel state information available to the receiver only, while Biglieri *et. al* treat the case of channel state information available to both the transmitter and receiver. Again, capacity increases as $\min\{T, R\}$ b/s/Hz for each additional 3 dB of SNR in the high SNR regime; this result is observed for capacity-vs.-outage in [25, 79], and for delay-limited capacity in [11].

In addition to increasing capacity, multi-antenna systems can be viewed as improving system robustness to fading conditions. Specifically, dramatic decreases in outage probability for a fixed rate have been observed [25, 79, 57].

3.3.2 Space-Time Codes

To approach the high spectral efficiencies forecast by the results in Section 3.3.1, many authors have examined suitable coding and decoding methods, called *space-time codes*, for multi-antenna systems. Early schemes consisted of scalar-coded methods, *e.g.*, repetition diversity over orthogonal frequency bands as well as bandwidth-conserving schemes such as time-shifting and phase-sweeping diversity. See [57] and the references therein. Notable developments in the area of vector-coding for multi-antenna systems have come from the work of Foschini and Gans [24, 25] on the BLAST system; the work of Alamouti [5] on simple block codes that achieve full diversity and have particularly simple, linear decoding algorithms; the work of Tarokh *et. al* [77, 76] on space-time trellis and block codes; and the work of Hochwald, Marzetta, and others [39, 40, 38] on non-coherent and differential space-time coding and modulation methods.

3.4 Wireless Networks

As we have emphasized, opportunities for exploiting cooperative diversity arise primarily in *networks* of radio terminals, and, except in the simplest cases (*e.g.*, the three-terminal relay channel described above), these networks require rather involved architectures in order to function. Furthermore, depending upon the application and amount of pre-existing infrastructure, these architectures can take on very different forms. We now summarize important results obtained for both infrastructure networks and ad-hoc networks.

3.4.1 Infrastructure Networks

There is a vast array of literature on specific cellular systems, *e.g.*, TDMA (GSM, IS-136), CDMA (IS-95). From the perspective of fundamental performance limits at the physical layer, infrastructure networks are treated primarily at the cell level, typically treating inter-cell interference as noise. Specifically, the uplink and downlink transmission in the cell are often modeled as multiple-access and broadcast channels, respectively, with or without multiple-antennas at the basestation, channel state information at the basestation and mobiles, and inter-cell interference through the cell.

Early fundamental work by Wyner [92] on cellular uplink models treated systems without fading and dealt with inter-cell interference by allowing all the basestations to coopera-

tively decode transmissions from the mobiles. It was determined that intra-cell TDMA was sufficient for optimality, but that inter-cell TDMA degrades performance. Building upon this model, Shamai and Wyner [72, 73] incorporate fading measured only at the receivers into the model, and examine lower complexity receivers, and their associated inter-cell interference issues, in which only one or two basestations decode the transmissions of mobiles.

Other work incorporates fading measured by the receivers and fed back to the transmitters, allowing more sophisticated forms of power control in the network. Knopp and Humblet [46] showed in the symmetric case, and Tse and Hanly [83] in the more general case, that to maximize the total uplink throughput, in terms of Shannon capacity, it is sufficient to have only the user with the strongest channel to the basestation transmit at any given time—a multi-user generalization of Goldsmith and Varaiya's results [30] on single-user water-pouring over fading channel states. Moreover, such power control schemes offer a form of multiuser diversity in that, as the number of users increase, it is more likely that at least one of them has a strong channel to the basestation. This multiuser diversity translates into the total throughput of a multiple-access channel with fading being much larger than that of a non-faded channel. These results have been extended to downlink channels [53] and other forms of channel capacity such as delay-limited capacity and capacity-vs.-outage [34, 54, 52].

3.4.2 Ad-Hoc Networks

Ad-hoc networks, under the name packet radio networks, were introduced by [43, 44] as a wireless extension of packet switching in wire-line networks. Later research, particularly [51] and references therein, re-examined the issues in light of technological advances. Numerous authors have compared direct, or singlehop, transmission and cascade, or multihop, transmission under a variety of channel conditions and perspectives; see [80] and references therein. Still more authors have examined issues such as scheduling, routing, and organizational problems associated with these networks. A thorough review of this work is given by Kassab [45]. Ad-hoc networks have traditionally been developed within the computer and data networking communities, and recent work in the information theory community has contributed some fundamental performance and scaling laws.

In a landmark paper [32], Gupta and Kumar prove that certain fixed ad-hoc networks containing M stationary terminals have total throughputs per terminal that decay to zero

with increasing M in a constant area. Specifically, under no interference, or limited interference, protocols utilizing only direct or cascade transmission, the capacity per terminal decreases on the order of $1/\sqrt{M}$ for carefully constructed scenarios—terminal locations, traffic patterns, and transmit powers optimally chosen—and on the order of $1/\sqrt{M \log M}$ for more random scenarios—terminals locations and traffic patterns random, fixed power. The fixed protocol maximizes transport capacity (bit-meters/second) by having terminals transmit to their nearest neighbors. It is interesting that Shepard [75] draws essentially the same conclusion, that terminals should transmit only to their nearest neighbors, by examining the asymptotic behavior of the interference from non-nearest neighbors. In Gupta and Kumar's analysis, nearest neighbors are on the order of $1/\sqrt{M}$ away in normalized distance. To travel a normalized distance of 1 to its destination, each packet must travel over on the order of $\sqrt{M}$ hops. This growing number of hops in the protocol limits the throughput per terminal. These results suggest that, without more sophisticated network protocols, providing high data rates requires fixed networks to consist of small numbers of terminals, or utilize extremely large bandwidths as in Shepard [75].

Building upon [32], a clever paper by Grossglauser and Tse [31] examines highly mobile ad-hoc networks and proves that a suitable cascade transmission policy provides throughput of order 1 per terminal as the network density increases, provided long delays are tolerable to the application. The basic idea of the mobile protocol is as follows. At any given time, a terminal transmits packets—either its own or packets to be forwarded on behalf of another terminal—only to the receiving terminal to which it is closest. As the terminal density increases, the distances over which the transmissions occur become very small, so that very little power is required and interference is minimal. Over time, every terminal carries queued packets for every other terminal, and each packet is only forwarded once, when an intermediate terminal is close to the intended destination. Thus the protocol offers a so-called multi-user diversity effect: each time the destination terminal receives it will very likely be near either the original source terminal or an intermediate terminal carrying packets for the source.

We emphasize that the models, protocols, and results of [32] and [31] represent certain endpoints in the spectrum of possible conditions in which ad-hoc networks might operate in practice. Emerging research appears to be addressing some of the interior points. First, while both papers address the throughput capacity per terminal as the number of

terminals within a constant area becomes large, recent work by Toumpis and Goldsmith [82] has formulated an achievable rate region for multihop routing among a given finite number of terminals. Interestingly, [82] introduces rate vectors with negative rates corresponding to forwarding information for another terminal. We also note that Toumpis and Goldsmith have obtained asymptotic results [81] similar to [32]. Second, both [32] and [31] address path-loss and deal with inter-user interference; however, both leave out the issue of multipath fading and only consider transmission formats based upon direct and cascade transmission. In addition to our work on relay and cooperative transmission, Gupta and Kumar appear to be incorporating ideas from the classical relay channel into their work [33] to address fading more explicitly. Third, along the mobility dimension, [32] examines completely stationary networks, while [31] examines completely mobile networks. Fourth, both works allow the delay to become arbitrarily large. To our knowledge, no work has addressed either of these last two issues in a comprehensive way.

Chapter 4

Cooperative Diversity with Full Temporal Diversity

In this chapter, we specialize our wireless network channel models of Chapter 2 to multiple-access channels with cooperative diversity (*cf.* Fig. 2-5(c)). As in Chapter 2, we focus on channels with additive white Gaussian noise, with and without multipath Rayleigh fading. Our objective is to characterize fundamental performance limits when the channels are ergodic, so that full temporal diversity can be exploited, by examining the well-known Shannon capacity region [19]. The capacity region represents the largest set of transmission rates that can be reliably communicated over the channel, in the sense of asymptotically negligible codeword error probability with long codewords and unconstrained decoding complexity.

Our basic models are similar to those employed by Sendonaris *et. al* [68, 67, 69, 70], who were the first to examine cooperative diversity in the wireless setting. Sendonaris *et. al* examined the case of channel phase information being available to the transmitters, and demonstrate that cooperative diversity increases the sum-rate over non-cooperative transmission. We demonstrate that the degree of channel knowledge available to the transmitters significantly influences the relative utility of cooperative diversity over non-cooperative transmission in such settings. In particular, for the case of channel state information available only to the receivers emphasized throughout the dissertation, we show that cooperative diversity does not increase the maximum sum-rate over non-cooperative transmission.

In addition, we discuss how the multiple-access channel with cooperative diversity generalizes various relay channel models. The classical relay channel [17], parallel relay channel

[66, 65], and multiple relay channel [33, 29, 28, 27] models all can be viewed as multiple-access channels with only one of the terminals sending information. Because the performance advantages in terms of sum capacity for the multiple-access case must be *shared* by the cooperating terminals, various results for relay channels can be interpreted as focusing all of network resources on a particular terminal, which reaps all of the associated gains.

More generally, we point out that multiple-access channels with cooperative diversity are special cases of multiple-access channels with generalized feedback, a model originally developed by Carleial [14] and also studied by Willems [88, 91]. Within the generalized feedback setting, channel outputs are available at the encoders as well as at the decoder. These "feedback" outputs are generally different, and may have varying relationships with the channel output at the decoder. In the cooperative diversity setting, all three channel outputs are conditionally independent given the inputs. Thus, the feedback signals in our setting are more a means for the cooperating encoders to observe each others transmissions, and suitably adapt their own transmissions, instead of a means for listening to what the decoder receives. In this sense, "cooperative diversity" seems a more appropriate term than "feedback".

And outline of this chapter is as follows. First, we setup the mathematical framework for treating the multiple-access channel with cooperative diversity throughout the rest of the chapter. Second, we focus on the Gaussian case without fading, and develop insights from inner and outer bounds on the capacity region. Third, with this substantial development for the Gaussian case without fading, extension to the Gaussian case with fading is fairly straightforward using well-established results. Consequently, we keep the discussion on fading brief and focus on the central issue how having channel state information available at the transmitters impacts the relative utility of cooperative diversity over non-cooperative transmission.

4.1 Model and Definitions

In this section, we setup the mathematical framework for the rest of the chapter by specifying our model for Gaussian multiple-access channels with cooperative diversity. We point out several special cases of the model that have received attention in the literature, and we qualitatively describe coding strategies that offer good performance in certain regimes.

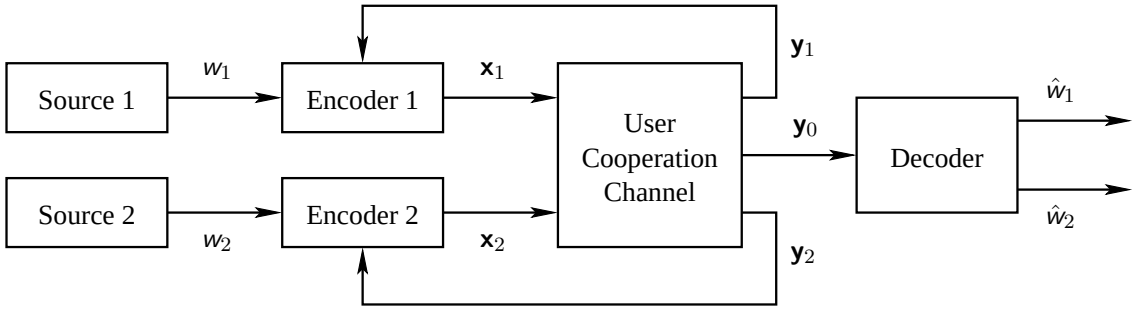


Figure 4-1: Block diagram for a multiple-access channel with cooperative diversity.

Finally, we define various objects such as the random channel codebooks, probability of error, achievable rates, and the capacity region. For convenience, we repeat some of the more relevant material from Chapter 2, and simplify notation wherever possible.

4.1.1 Block Diagram

Fig. 4-1 shows a general block diagram of a two-user multiple-access channel with cooperative diversity. Although the model can be naturally extended to more than two cooperating terminals, we restrict our attention to the two-user cases for simplicity of exposition.

Two sources generate independent messages w_1 and w_2. These messages are encoded into codewords $\mathbf{x}_1$ and $\mathbf{x}_2$ that can also depend causally on the channel outputs (generalized feedback signals) $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. The decoder observes the channel output $\mathbf{y}_0$ and estimates the source messages as $\hat{w}_1$ and $\hat{w}_2$, respectively.

In the complex Gaussian case, a discrete-time, baseband equivalent channel model has, for each sample time k,

$$\begin{bmatrix} y_0[k] \\ y_1[k] \\ y_2[k] \end{bmatrix} = \begin{bmatrix} a_{0,1}[k] & a_{0,2}[k] \\ a_{1,1}[k] & a_{1,2}[k] \\ a_{2,1}[k] & a_{2,2}[k] \end{bmatrix} \cdot \begin{bmatrix} x_1[k] \\ x_2[k] \end{bmatrix} + \begin{bmatrix} z_0[k] \\ z_1[k] \\ z_2[k] \end{bmatrix} . \tag{4.1}$$

We point out that, throughout this chapter, we relax the half-duplex constraint discussed in Section 2.1 and allow the encoders to simultaneously transmit and receive. This explains, why $x_1[k]$ (resp. $x_2[k]$) affects $y_1[k]$ (resp. $x_2[k]$) in (4.1). In principle, each encoder knows its channel input and can remove the effects of the input from the corresponding channel

output. Thus, in effect,

$$y_1[k] = a_{1,2}[k]\, x_2[k] + z_1[k] \ , \tag{4.2}$$

and similarly for $y_2[k]$.

As in Chapter 2, $a_{j,i}[k]$ captures the effects of attenuation and multipath fading between input i and output j, and $z_j[k]$ captures the effects of additive noise and other forms of interference. In this section, we treat the case of $a_{j,i}[k]$ being fixed coefficients $a_{j,i}$ that are known to the encoders and decoder, while in later sections we model $a_{j,i}[k]$ as stationary and ergodic random processes known to at least the decoder and possibly the encoders. Throughout we model $z_j[k]$ as mutually independent, zero-mean, circularly-symmetric complex Gaussian white noise processes, each with variance N_j.[1]

4.1.2 Parameterization

In the sequel, we consider n consecutive channel uses of the channel, where n is large. Where appropriate we group collections of samples of the same signal into (possibly random) vectors. For example, we write

$$\mathbf{x}_i[l] = \begin{bmatrix} x_i[l] & x_i[l+1] & \dots x_i[l+n-1] \end{bmatrix}^T \ , \tag{4.3}$$

and similarly for $\mathbf{y}_j[l]$ and $\mathbf{z}_j[l]$. When the block index can be inferred from the context, we drop it for notational compactness.

Because of transmit power constraints in the system, the transmit signals are constrained to satisfy the average power constraints

$$\frac{1}{n} \sum_{k=1}^{n} |x_i[k]|^2 \leq P_i \ , \tag{4.4}$$

with high probability for large n. With this parameterization, it is useful to express our results in terms of the signal-to-noise ratios (SNRs)

$$s_{j,i} = |a_{j,i}|^2 \, \frac{P_i}{N_j} \ . \tag{4.5}$$

[1]Recall that a unit-variance random variable u being circularly-symmetric means $u = u_R + j u_I$, where the real and imaginary parts, u_R and u_I, respectively, are uncorrelated and each have variance 1/2.

In later sections of this chapter, when the fading coefficients are random variables $\mathsf{a}_{j,i}$, we denote the corresponding SNR random variables by $\mathsf{s}_{j,i}$ and parameterize performance by the average SNRs $\mathrm{E}[\mathsf{s}_{j,i}]$.

4.1.3 Special Cases and Coding Strategies

Given the channel model and parameters described above, we now illustrate various special cases of the network in Fig. 4-1, including the multiple-access channel (without cooperative diversity), the relay channel, and multi-relay channels. These special cases are important because work by other researchers on these problems contributes insights to our understanding of the more general problem. In particular, work on these special cases provide certain coding strategies, some of which we employ directly or modify in this and later chapters. Furthermore, examination of the more general problem in the sequel lends certain insights about the special cases.

Multiple-Access Channel

The well-known Gaussian multiple-access channel [19] is an immediate special case of the general channel model in Fig. 4-1. It arises if, for example, $\mathbf{y}_j \equiv 0$, $j = 1, 2$. Because the encoder channel outputs are useless, the encoders cannot cooperate in any fashion and instead transmit independently. When used for the more general channel, we refer to independent signaling as *non-cooperative transmission*. As we might expect, non-cooperative transmission is effective in the general model when $s_{1,2}$ and $s_{2,1}$ are extremely small.

Relay Channel

The well-known Gaussian relay channel without feedback [19] is another immediate special case of the general channel model in Fig. 4-1. It arises if, for example, encoder 2 transmits no information of its own, *i.e.*, $w_2 \equiv 0$, and encoder 1 has no channel output, *i.e.*, $\mathbf{y}_1 \equiv 0$. For convenience, Fig. 4-2 shows Fig. 4-1 redrawn under these conditions, with Encoder 1 relabeled simply "Encoder", and Encoder 2 relabeled "Relay".

As we discussed in Section 3.1, there are three basic coding strategies that have been developed for the general Gaussian relay channel by Cover and El Gamal [17]. In the first strategy, which Cover and El Gamal call "facilitation" in [17] and we call "direct" transmission throughout the dissertation, the relay chooses a channel input to maximize

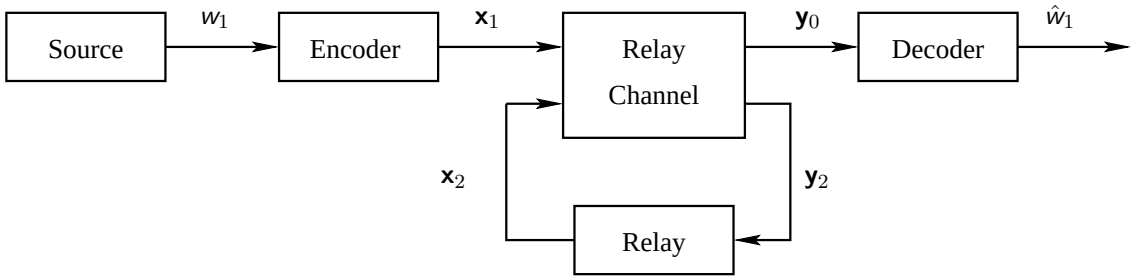


Figure 4-2: Block diagram for a relay channel.

the amount of information that can be transmitted between the encoder and decoder. In essence, the relay tries to minimize the amount of interference it causes the encoder. Direct transmission becomes appealing when either $s_{2,1}$ or $s_{0,2}$ is very small. In the second strategy, which Cover and El Gamal call "cooperation" in [17] and we call "decode-and-forward" throughout the dissertation, the relay fully decodes the source message w_1 from its channel output $\mathbf{y}_2$. With the message available to both the encoder and relay, the two cooperatively transmit correlated signals that coherently combine at the decoder. Decode-and-forward is appealing when $s_{0,2}$ is not too small and $s_{2,1}$ is large. Finally, in the third strategy, which Cover and El Gamal do not name and we call "observe-and-forward" throughout the dissertation, the relay communicates a representation $\hat{\mathbf{y}}_2$ of its channel output $\mathbf{y}_2$ to the decoder, in effect giving the decoder another observation from the channel with which to fuse its own observation $\mathbf{y}_0$. The observation is generally useful when $s_{2,1}$ is not too small and $s_{0,2}$ is large enough that $\hat{\mathbf{y}}_2$ is a high-quality representation of $\mathbf{y}_2$.

Multiple-Relay Channels

More recent extensions of the relay channel are also special cases of the Gaussian multiple-access channel with cooperative diversity, but with typically more than two users. For example, the multi-relay channel model of [33, 29, 28, 27] with M relays is a special case of the $(M+1)$-user multiple-access channel with cooperative diversity, and has structure similar to that shown in Fig. 4-2 but with M relays instead of one. The parallel relay channel [66, 65] is a special case of the multi-relay channel in which the encoder signal does not affect the decoder channel output, *e.g.*, $a_{0,1}[k] \equiv 0$. All these cases consider only a single information source.

In addition to coding strategies based upon generalizations of decode-and-forward transmission [33] and generalizations and combinations of decode-and-forward and observe-and-forward transmission [27], the authors in [66, 65, 28] also consider coding strategies in which the relays simply amplify what they receive subject to their power constraint. These "amplify-and-forward" strategies, as we refer to them throughout the dissertation, are also shown to be appealing in certain regimes. For example, [29, 28] shows that amplify-and-forward asymptotically achieves capacity of the Gaussian multi-relay channel as $M \to \infty$.

4.1.4 Definitions

We confirm some of the intuitive statements made in the previous section by partially characterizing, in the next few sections, the so-called *capacity region* [19, Chapter 14] of the Gaussian multiple-access-channel with cooperative diversity. In this section, we provide the necessary definitions. As we might expect, since the cooperative diversity scenario is a special case of the multiple-access channel with generalized feedback, our definitions mirror those of Willems [91]. More generally, we assume throughout the sequel that the reader is familiar with the concepts and properties of entropy, differential entropy, and mutual information as developed in Cover and Thomas [19].

A (two-user) *memoryless Gaussian multiple-access channel with cooperative diversity* consists of the following:

- Channel inputs x_1 and x_2, with corresponding alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, and channel outputs y_0, y_1, and y_2, with corresponding alphabets $\mathcal{Y}_0$, $\mathcal{Y}_1$, and $\mathcal{Y}_2$. Due to complex-valued additive Gaussian noise in the channel model (4.1), $\mathcal{Y}_0$,$\mathcal{Y}_1$, and $\mathcal{Y}_2$ are the complex plane $\mathbb{C}$. Generally, $\mathcal{X}_1$ and $\mathcal{X}_2$ will also be the complex plane $\mathbb{C}$, though this is not necessary.

- Channel probability law mapping inputs to outputs

$$p_{y_0,y_1,y_2|x_1,x_2}(y_0,y_1,y_2|x_1,x_2) = \prod_{i=0}^{2} p_{y_i|x_1,x_2}(y_i|x_1,x_2) \ . \tag{4.6}$$

  By memoryless, we mean that the probability law for n consecutive uses of the channel

is the product

$$p_{\mathbf{y}_0,\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}_1,\mathbf{x}_2}(\mathbf{y}_0,\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}_1,\mathbf{x}_2) = \prod_{k=1}^{n} p_{y_0[k],y_1[k],y_2[k]|x_1[k],x_2[k]}(y_0[k],y_1[k],y_2[k]|x_1[k],x_2[k]) \ . \quad (4.7)$$

A *communication strategy* for the Gaussian multiple-access channel with cooperative diversity consists of the following:

- Messages $w_i \in \mathcal{M}_i = \{1, 2, \ldots, M_i\}$, $i = 1, 2$, distributed uniformly and independently. The rates in bits per channel use are then

$$\frac{1}{n} \log_2 M_i \ , \qquad i = 1, 2 \ . \quad (4.8)$$

- Encoding functions, one for each encoder for each time k, that map the encoder's message w_i and past channel observations $y_i[1], \ldots, y_i[k-1]$ into its transmitted signal x_i, $i = 1, 2$. That is

$$x_i[k] = f_{i,k}(w_i, y_i[1], y_i[2], \ldots, y_i[k-1]) \ , \qquad i = 1, 2 \ . \quad (4.9)$$

  Note that each encoder is a *causal* function of the encoder's message w_i and channel output $\mathbf{y}_i$, $i = 1, 2$.

- A decoding function mapping the channel output vector $\mathbf{y}_0$ into $\mathcal{M}_1 \times \mathcal{M}_2$, *i.e.*,

$$(\hat{w}_1, \hat{w}_2) = g(\mathbf{y}_0) \ . \quad (4.10)$$

One way to characterize system performance for a particular communication strategy over a given channel is in terms of the *average probability of error* in the decoder. This quantity is defined in the standard way as follows.

Definition 1 *The average probability of error for a communication strategy operating over a Gaussian multiple-access channel with cooperative diversity is*

$$P_e^{(n)} = \Pr\left[g(\mathbf{y}_0) \neq (w_1, w_2)\right] \ . \quad (4.11)$$

Note that the probability defined in (1) is for error events in a *block* of n channel uses. Generally it is the case that $P_e^{(n)}$ is decreasing in n for fixed transmission rates, and increasing in transmission rates for fixed n. A reasonable objective for a system designer is to determine communication strategies that have minimal average probability of error for a given blocklength; however, solving such constructive problems becomes intractable as n becomes large. Alternatively, using extensions of arguments originally due to Shannon [74], we can ascertain whether there exist communication strategies with fixed transmission rates having negligible average probability of error as n becomes large. Although we cannot characterize the strategies explicitly, *i.e.*, the arguments are not constructive, we can determine certain conditions on the rates that guarantee the existence of such communication strategies.

Definition 2 *A pair of transmission rates (R_1, R_2) is said to be achievable on a Gaussian multiple-access channel with cooperative diversity if there exists a sequence of communication strategies operating over the channel with*

$$M_i^{(n)} = 2^{\lceil nR_i \rceil} \ , \qquad i = 1, 2 \ ,$$

and

$$P_e^{(n)} \to 0 \ , \qquad \text{as } n \to \infty \ .$$

Given this conventional definition for achievable rates, it is natural to consider the largest set of achievable rate pairs that can be reliably transmitted over the channel. This set of rate pairs is the well-known *capacity region*, and is defined as follows.

Definition 3 *The capacity region of a Gaussian multiple-access channel with cooperative diversity is the closure of the set of achievable rates.*

As in many other multi-terminal communication settings [19, Chapter 14], the capacity region has a compact and useful interpretation: rates inside the capacity region can be reliably transmitted over the channel, in the sense of Definition 2, by communication strategies

that go largely unspecified; however, certain structural properties of the effective strategies can be ascertained. Rates outside the capacity region cannot be reliably transmitted over the channel using any communication strategy.

4.2 Converse: Outer Bound on the Capacity Region

To develop an outer bound on the capacity region for the Gaussian multiple-access channel with cooperative diversity, we specialize the well-known cut-set bounds to the Gaussian case. To our knowledge, this outer bound has not been explicitly stated or examined for the multiple-access channel with cooperative diversity, but it does bear a striking similarity to the corresponding converse developed by Ozarow [58] for the multiple-access channel with noiseless feedback. Specifically, we have the following theorem.

Theorem 1 *For the Gaussian memoryless multiple-access channel with cooperative diversity, if the rate pair (R_1, R_2) is achievable, then there exists a $0 \leq |\rho| \leq 1$ such that*[2]

$$R_1 \leq \log\left(1 + [1 - |\rho|^2][s_{0,1} + s_{2,1}]\right) \tag{4.12}$$

$$R_2 \leq \log\left(1 + [1 - |\rho|^2][s_{0,2} + s_{1,2}]\right) \tag{4.13}$$

$$R_1 + R_2 \leq \log\left(1 + s_{0,1} + s_{0,2} + 2|\rho|\sqrt{s_{0,1}s_{0,2}}\right) \ . \tag{4.14}$$

We note that a proof of Theorem 1 follows along the same lines as the proof of the cut-set bounds [19, Theorem 14.10.1], with the additional steps of appealing to the fact that the Gaussian distribution maximizes entropy subject to a covariance constraint, utilizing convexity properties of the logarithm, and applying the power constraints. Due to the similarity of the multiple-access channels with cooperative diversity and with feedback, the detailed converse proof given by Ozarow [58] also applies with y_i substituted for y_0 at the respective encoders. For these reasons, we do not provide a complete proof here.

To see how, for example, (4.12) results, consider the cut-set corresponding to transmission from encoder 1 and reception at both encoder 2 and the decoder, *i.e.*, a "broadcast" cut-set. According to the cut-set bound [19, Theorem 14.10.1], if R_1 is achievable then

[2] *As throughout the dissertation, all logarithms are to base-2 unless stated otherwise.*

there exists a joint probability distribution $p_{x_1,x_2}(x_1, x_2)$ on x_1 and x_2 such that

$$R_1 < I(x_1; y_0, y_2|x_2) \ . \tag{4.15}$$

For the additive white Gaussian noise case in (4.1), we define covariance matrices

$$\Lambda_{x_1,x_2} = \begin{bmatrix} \lambda_{x_1} & \rho\sqrt{\lambda_{x_1}\lambda_{x_2}} \\ \rho^*\sqrt{\lambda_{x_1}\lambda_{x_2}} & \lambda_{x_2} \end{bmatrix} \ , \qquad \Lambda_{z_0,z_1,z_2} = \begin{bmatrix} N_0 & 0 & 0 \\ 0 & N_1 & 0 \\ 0 & 0 & N_2 \end{bmatrix} \ ,$$

$$\Lambda_{y_0,y_1,y_2} = A\Lambda_{x_1,x_2}A^\dagger + \Lambda_{z_0,z_1,z_2} \ ,$$

where $\lambda_{x_i} = \mathrm{Var}\left[x_i\right]$, $i = 1, 2$, ρ is the (complex) correlation coefficient between x_1 and x_2, and $[A]_{j,i} = a_{j,i}$.

Then, following arguments similar to those in [17, 58], we have

$$\begin{aligned} I(x_1; y_0, y_2|x_2) &= h(y_0, y_2|x_2) - h(y_0, y_2|x_2, x_1) \\ &= h(y_0, y_2|x_2) - h(z_0, z_2) \\ &= \mathrm{E}\left[h(y_0, y_2|x_2 = x_2)\right] - \log\det(\pi e)\Lambda_{z_0,z_2} \\ &\leq \mathrm{E}\left[\log\det(\pi e)\Lambda_{y_0,y_2|x_2=x_2}\right] - \log\det(\pi e)\Lambda_{z_0,z_2} \\ &= \mathrm{E}\left[\log\det\Lambda_{y_0,y_2|x_2=x_2}\Lambda_{z_0,z_2}^{-1}\right] \\ &= \mathrm{E}\left[\log\left(1 + \lambda_{x_1|x_2=x_2}\left[\frac{|a_{0,1}|^2}{N_0} + \frac{|a_{2,1}|^2}{N_2}\right]\right)\right] \\ &\leq \log\left(1 + \mathrm{E}\left[\lambda_{x_1|x_2=x_2}\right]\left[\frac{|a_{0,1}|^2}{N_0} + \frac{|a_{2,1}|^2}{N_2}\right]\right) \ . \end{aligned} \tag{4.16}$$

The first inequality follows from the fact that the circularly-symmetric complex Gaussian distribution maximizes entropy [78], and the second inequality follows from the well-known Jensen's inequality.

Finally, for any pair of random variables, standard estimation results tell us the conditional variance $\lambda_{x_1|x_2=x_2}$ of x_1 around the conditional mean estimate is no greater than the

conditional variance of x_1 around the linear estimate $\hat{x}_1 = (\rho/\lambda_{x_2})x_2$. Thus

$$\begin{aligned} \mathrm{E}\left[\lambda_{x_1|x_2=x_2}\right] &\leq \mathrm{E}\left[\left|x_1 - \frac{\rho}{\lambda_{x_2}}x_2\right|^2\right] \\ &= \lambda_{x_1}(1-|\rho|^2)\ . \end{aligned} \tag{4.17}$$

Substituting (4.17) into (4.16), and applying the power constraint $\lambda_{x_1} \leq \mathrm{E}\left[|x_1|^2\right] \leq P_1$, yields the desired result (4.12). The result (4.13) follows in the same fashion.

To see why (4.14) is true, consider the cut-set corresponding to transmission from encoders 1 and 2 and reception at the decoder, *i.e.*, the multiple-access cut-set. If $R_1 + R_2$ is achievable, then the corresponding cut-set bound implies there exists a joint probability distribution $p_{x_1,x_2}(x_1, x_2)$ on x_1 and x_2 such that

$$R_1 + R_2 < I(x_1, x_2; y_0)\ . \tag{4.18}$$

Again, following arguments similar to those in [17, 58] for the Gaussian case, we have

$$\begin{aligned} I(x_1, x_2; y_0) &= h(y_0) - h(y_0|x_1, x_2) \\ &= h(y_0) - h(z_0) \\ &= h(y_0) - \log(\pi e N_0) \\ &\leq \log(\pi e \mathrm{Var}\left[y_0\right]) - \log(\pi e N_0) \\ &= \log(\pi e[\mathbf{a}\Lambda_{x_1,x_2}\mathbf{a}^\dagger + N_0]) - \log(\pi e N_0) \\ &= \log(1 + [\mathbf{a}\Lambda_{x_1,x_2}\mathbf{a}^\dagger]/N_0) \\ &= \log\left(1 + \frac{\lambda_{x_1}|a_{0,1}|^2}{N_0} + \frac{\lambda_{x_2}|a_{0,2}|^2}{N_0}\right. \\ &\qquad \left. + 2\cos(\measuredangle\rho + \measuredangle a_{0,1} - \measuredangle a_{0,2})|\rho|\sqrt{\frac{\lambda_{x_1}|a_{0,1}|^2}{N_0}\frac{\lambda_{x_2}|a_{0,2}|^2}{N_0}}\right) \\ &\leq \log\left(1 + \frac{\lambda_{x_1}|a_{0,1}|^2}{N_0} + \frac{\lambda_{x_2}|a_{0,2}|^2}{N_0} + 2|\rho|\sqrt{\frac{\lambda_{x_1}|a_{0,1}|^2}{N_0}\frac{\lambda_{x_2}|a_{0,2}|^2}{N_0}}\right) \\ &\leq \log(1 + s_{0,1} + s_{0,2} + 2|\rho|\sqrt{s_{0,1}s_{0,2}})\ . \end{aligned} \tag{4.19}$$

where $\mathbf{a} = [a_{0,1} \; a_{0,2}]$. The first inequality again follows from the fact that circularly symmetric complex Gaussians maximize entropy subject to a covariance constraint. The second inequality follows from matching the phase of the correlation coefficient so that $\measuredangle\rho + \measuredangle a_{0,1} - \measuredangle a_{0,2} = 0$, ensuring that the transmitted signals coherently combine at the decoder. The final inequality follows from substitution of each of the power constraints $\lambda_{\mathsf{x}_i} \leq \mathrm{E}\left[|\mathsf{x}_i|^2\right] \leq P_i$, $i = 1, 2$.

Fig. 4-3 illustrates the shape of the outer bound on the capacity region obtained from Theorem 1, as well as the achievable rate regions described in more detail in the next section.

4.3 Achievability: Inner Bounds on the Capacity Region

In this section, we describe several communication strategies for the Gaussian multiple-access channel with cooperative diversity. We compare the corresponding sets of achievable rates of these strategies with the outer bound obtained in Theorem 1. In particular, we compare the sum-rates in high and low SNR regimes. We focus on non-cooperative transmission and decode-and-forward transmission. A complete treatment of observe-and-forward transmission requires better understanding of distributed source coding. A nice survey of this area along with recent progress can be found in [21].

4.3.1 Non-Cooperative Transmission

If the encoders do not exploit their observations y_1 and y_2 from the channel due to complexity or legacy issues, or these observations are too noisy to be very useful, the transmissions can take the form of non-cooperative transmission. In this case, the system model reduces to the classical multiple-access channel, for which the set of achievable rates is well-known [19, Section 14.3.1].

Theorem 2 *The set of achievable rates for non-cooperative transmission over a memoryless Gaussian multiple-access channel with cooperative diversity is given by the set of all* (R_1, R_2)

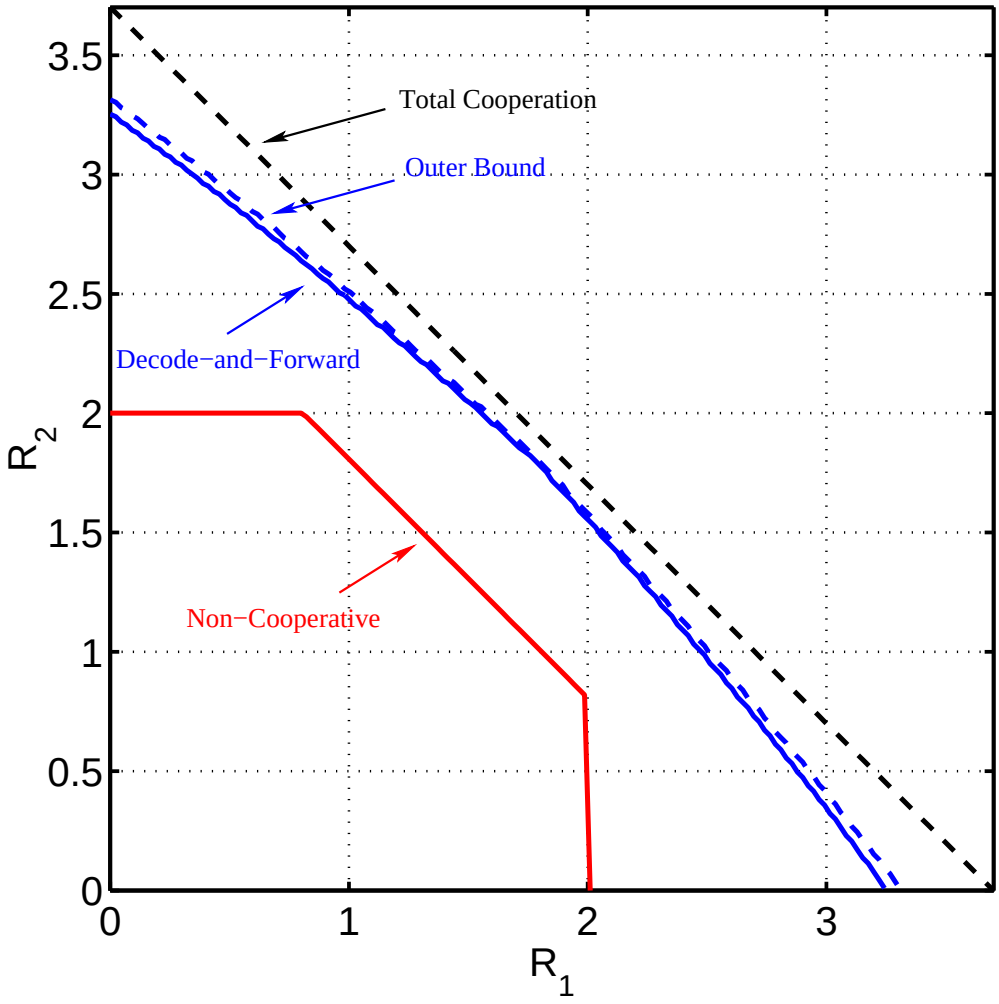


Figure 4-3: Illustration of shape of various outer and inner bounds on the capacity region of the Gaussian multiple access channel with cooperative diversity. Here the SNRs take values $s_{0,1} = s_{0,2} = 3$, and $s_{2,1} = s_{1,2} = 15$.

satisfying

$$R_1 < \log\left(1 + s_{0,1}\right) \ , \tag{4.20}$$

$$R_2 < \log\left(1 + s_{0,2}\right) \ , \tag{4.21}$$

$$R_1 + R_2 < \log\left(1 + s_{0,1} + s_{0,2}\right) \ . \tag{4.22}$$

Fig. 4-3 also illustrates the pentagon region given by Theorem 2. To compare the achievable sum rate (4.22) with the outer bound (4.14), let $\overline{s} = (s_{0,1} + s_{0,2})/2$ be the arithmetic mean of the SNRs from the encoders to the decoder. Then non-cooperative transmission achieves sum-rate $\log(1+2\overline{s})$. The outer bound on the sum rate (4.22) satisfies

$$\log(1 + 2\overline{s} + 2|\rho|\sqrt{s_{0,1}s_{0,2}}) \leq \log(1 + 2\overline{s} + 2|\rho|\overline{s}) \tag{4.23}$$

$$\leq \log(1 + 4\overline{s}) \ , \tag{4.24}$$

where the first inequality follows from the fact that the geometric mean of positive numbers is no greater than their arithmetic mean, and the second inequality follows from the fact that the magnitude of the correlation coefficient ρ is no greater than one.

Now we observe that as $\overline{s} \to 0$, the ratio

$$\log(1 + 4\overline{s}) / \log(1 + 2\overline{s}) \to 2 \ , \tag{4.25}$$

so that cooperative diversity increases the sum rate by at most a factor of two for low average SNR. On the other hand, as $\overline{s} \to \infty$, the difference

$$\log(1 + 4\overline{s}) - \log(1 + 2\overline{s}) \to 1 \ , \tag{4.26}$$

so that cooperative diversity increases the sum rate by at most 1 b/s/Hz for high average SNR. For M encoders, one can show that cooperative diversity provides sum rate no more than $\log(1 + M^2\overline{s})$, while non-cooperative transmission provides sum rate at most $\log(1 + M\overline{s})$, leading to gains in sum rate of at most a factor of M for low SNR and an additional $\log(M)$ b/s/Hz for high SNR. In the following section, we illustrate some regimes in which such capacity increases can be realized.

4.3.2 Decode-and-Forward Transmission

The communication strategy described in this section has the encoders exploit their channel outputs by decoding each other's transmissions and forwarding cooperative refinement information in order to assist the decoder.

The following theorem is proven in Appendix A.

Theorem 3 *The set of achievable rates for decode-and-forward transmission over a memoryless Gaussian multiple-access channel with cooperative diversity is given by the set of all* (R_1, R_2) *satisfying*

$$R_1 < \log\left(1 + \alpha_1 s_{2,1}\right) \ , \tag{4.27}$$

$$R_2 < \log\left(1 + \alpha_2 s_{1,2}\right) \ , \tag{4.28}$$

$$R_1 + R_2 < \log\left(1 + s_{0,1} + s_{0,2} + 2\sqrt{(1-\alpha_1)(1-\alpha_2)s_{0,1}s_{0,2}}\right) \ . \tag{4.29}$$

for some $0 \leq \alpha_i \leq 1$, $i = 1, 2$.

The proof in Appendix A utilizes superposition block-Markov coding [18] and backward decoding [90]. It can be viewed as a two-user generalization of the cooperation strategy introduced by Cover and El Gamal [17] for the relay channel, or a simplified version of the strategy developed by Willems, van der Meulen, and Schalkwijk [91] for the multiple access channel with generalized feedback.

To highlight the communication strategy, we describe the structure of the random codebooks used in the proof; details of encoding and decoding are left to Appendix A. Suppose

$$\mathsf{x}_1 = \sqrt{\alpha_1 P_1}\mathsf{v}_1 + \sqrt{(1-\alpha_1)P_1}\mathsf{u} \tag{4.30}$$

$$\mathsf{x}_2 = \sqrt{\alpha_2 P_2}\mathsf{v}_2 + \sqrt{(1-\alpha_2)P_2}\mathsf{u} \tag{4.31}$$

where u and v_i, $i = 1, 2$, are mutually independent complex Gaussian random variables, and $0 \leq \alpha_i \leq 1$, $i = 1, 2$. Then $\mathsf{x}_1 \leftrightarrow \mathsf{u} \leftrightarrow \mathsf{x}_2$ forms a Markov chain, the development in Appendix A applies, and the region in Theorem 3 is obtained by varying α_i, $i = 1, 2$. The signals v_i, $i = 1, 2$ convey fresh information in the current block, and the signal u conveys cooperative refinement information about the fresh information in the previous block. To transmit identical refinement information in the current block, each encoder decodes the

fresh information of the other encoder from the previous block.

This decode-and-forward communication strategy performs well for regimes in which the inter-encoder SNRs are large, as the example in Fig. 4-3 illustrates. In fact, this strategy offers the full capacity increases of cooperative diversity in such regimes. As one example, suppose the $\overline{s} = s_{0,1} = s_{0,2}$ and $s_{1,2}, s_{2,1} \to \infty$. We can then allow $\alpha_1, \alpha_2 \to 0$ in such a way that the sum rate bound (4.29) is always smaller than the sum of the other two bounds (4.27) and (4.28). In the limit, the achievable sum rate approaches $\log(1 + 4\overline{s})$. Thus, for small $\overline{s}$, the sum rate is essentially doubled over non-cooperative transmission, and for large $\overline{s}$ roughly an additional 1 b/s/Hz can be achieved over non-cooperative transmission.

4.4 Rayleigh Multipath Fading

In this section, we briefly consider the scenario in which the fading coefficients $\mathsf{a}_{j,i}[k]$ are i.i.d. random processes in time and mutually independent of one another. Our discussion can be extended to the case of stationary and ergodic fading processes as in [83]. As throughout the rest of the dissertation, we focus on the case of channel state information available to the receivers only. We preclude the possibility of beamforming and power control by the transmitters in our discussion, but stress that this is an important area of future work on the topic.

We will see that the degree of channel state information available at the transmitters can have a dramatic effect on the utility of cooperative diversity over non-cooperative transmission. For example, for the case in which channel state information is unavailable at the transmitters but available to the receivers, the sum rate for cooperative diversity is no greater than the sum rate without cooperation. Achieving gains similar to those obtained in the Gaussian case requires accurate channel phase information at the encoders, as developed in [67]. Whether or not cooperative diversity is beneficial when power control is an option remains an open question.

When the receivers can accurately track the appropriate fading processes, but the transmitters either do not have access to such information or other do not exploit it, as is standard practice in such conditions [78], we can view the channel observations at the encoders as the modified outputs

$$\tilde{\mathsf{y}}_1 = (\mathsf{y}_1, \mathsf{a}_{1,2}) \ , \qquad \tilde{\mathsf{y}}_2 = (\mathsf{y}_2, \mathsf{a}_{2,1}) \ , \tag{4.32}$$

respectively, and the channel observations at the decoder as the modified output

$$\tilde{y}_0 = (y_0, a_{0,1}, a_{0,2}) \ . \tag{4.33}$$

Under this effective channel model, mutual informations between inputs and outputs reduce to appropriate expectations over the fading distributions. For example,

$$\begin{aligned} I(x_1, x_2; \tilde{y}_0) &= I(x_1, x_2; a_{0,1}, a_{0,2}) + I(x_1, x_2; y_0 | a_{0,1}, a_{0,2}) && (4.34) \\ &= I(x_1, x_2; y_0 | a_{0,1}, a_{0,2}) && (4.35) \\ &= \mathrm{E}\left[I(x_1, x_2; y_0 | a_{0,1} = a_{0,1}, a_{0,2} = a_{0,2}) \right] \ , && (4.36) \end{aligned}$$

where the first equality follows form the chain rule for mutual informations, the second equality follows from the fact that the channel inputs are independent of the fading processes when the transmitters have no channel state information, and the final equality follows from the definition of conditional mutual information.

Similarly,

$$\begin{aligned} I(x_1; \tilde{y}_2 | x_2) &= \mathrm{E}\left[I(x_1; y_2 | x_2, a_{2,1} = a_{2,1}) \right] && (4.37) \\ I(x_2; \tilde{y}_1 | x_1) &= \mathrm{E}\left[I(x_2; y_1 | x_1, a_{1,2} = a_{1,2}) \right] \ , && (4.38) \end{aligned}$$

and so forth.

Following arguments similar to those leading to (4.19), we have

$$\mathrm{E}\left[I(x_1, x_2; y_0 | a_{0,1} = a_{0,1}, a_{0,2} = a_{0,2}) \right] \leq \mathrm{E}\left[\log(1 + [\mathbf{a} \Lambda_{x_1,x_2} \mathbf{a}^\dagger]/N_0) \right] \ , \tag{4.39}$$

with equality for x_i circularly-symmetric complex jointly Gaussian inputs, where $\mathbf{a} = [a_{0,1} \ a_{0,2}]$. We may diagonalize the input covariance matrix $\Lambda_{x_1,x_2} = UDU^\dagger$, for some unitary U and diagonal D, with $[D]_{i,i} \leq P_i$, $i = 1, 2$, according to the power constraints. More generally, for the case of i.i.d. Rayleigh fading, the distribution of $\mathbf{a}$ is invariant to a unitary transformation $\mathbf{a}U$, where U is a unitary matrix [78]. This fact implies that expectations of $\mathbf{a} \Lambda_{x_1,x_2} \mathbf{a}^\dagger$ correspond to expectations of $\mathbf{a} D \mathbf{a}^\dagger$, so that (4.39) becomes, after some

substitutions and simplifications,

$$\mathrm{E}\left[I(x_1, x_2; y_0 | a_{0,1} = a_{0,1}, a_{0,2} = a_{0,2})\right] \leq \mathrm{E}\left[\log(1 + s_{0,1} + s_{0,2})\right] \ . \tag{4.40}$$

Note that (4.40) corresponds to several important quantities in our study. Using our converse result (4.14) with the preceding steps, it represents an upper bound on the sum rate of the Gaussian multiple-access channel with cooperative diversity and i.i.d. Rayleigh fading known only to the receivers. Using our achievability result (4.22) with the preceding steps, it represents a lower bound on the sum rate of the Gaussian multiple-access channel without cooperation and i.i.d. Rayleigh fading known only to the receivers. Thus, we can employ independent inputs to achieve the sum-rate of the Gaussian multiple-access channel with cooperative diversity and i.i.d. Rayleigh fading known only to the receivers. In other words, non-cooperative transmission achieves the sum rate under these conditions. Thus, for this particular scenario of channel state information not being available at the transmitters, cooperative diversity is not useful in terms of increasing sum capacity.

By contrast, Sendonaris *et. al* [68, 67, 69, 70] consider the case of channel phase information available the transmitters. This information allows the encoders to appropriately phase their input signals so that the coherently combine at the decoder, as in (4.19). Under these conditions, the sum-rate bound (4.39) becomes

$$\mathrm{E}\left[I(x_1, x_2; y_0 | a_{0,1} = a_{0,1}, a_{0,2} = a_{0,2})\right] \leq \mathrm{E}\left[\log(1 + s_{0,1} + s_{0,2} + 2|\rho|\sqrt{s_{0,2}s_{0,1}})\right] \ . \tag{4.41}$$

As in the Gaussian case, such phase information allows the sum-rate for cooperative diversity to be larger than the sum-rate for non-cooperative transmission. More generally, if the transmitters also obtain amplitude information, sophisticated power control becomes an option [83]. However, the degree to which distributed radio hardware can obtain and exploit accurate phase and amplitude information is not well-known.

Chapter 5

Cooperative Diversity without Temporal Diversity

In Chapter 4, we examined communication strategies and capacity regions for multiple-access channels with cooperative diversity in ergodic fading environments where full temporal diversity can be exploited. We found that, under certain channel conditions, namely, when fading channel state information is available at the receivers but not the transmitters, cooperative diversity does not increase the achievable sum-rate when compared to non-cooperative transmission. In this chapter, we examine cooperative diversity in non-ergodic fading environments, *e.g.*, scenarios in which the fading varies slowly or delay constraints limit the coding interval to a finite number of channel realizations. As we discussed in Chapter 3, outage probability is a useful performance measure in this context. We will see in this chapter that cooperative diversity can dramatically improve outage performance in non-ergodic environments when no temporal diversity can be exploited, with benefits quite similar to those of transmit antenna arrays [57].

This chapter develops low-complexity cooperative diversity protocols that take into account certain implementation constraints in the system, such as orthogonal transmission and half-duplex relaying. In contrast to Chapter 4, where these constraints complicate the analysis, the orthogonal transmission constraint allows for our algorithms to be readily integrated into existing networks and makes the analysis of outage probability more tractable and convenient for exposition. Essentially, by separating the transmissions into orthogonal channels, we eliminate the usual coupling of outage probabilities due to interference [52].

The half-duplex constraints can be dealt with in a straightforward way within this setting.

We describe several cooperative transmission protocols and demonstrate their robustness to fairly general channel conditions. In addition to direct transmission, we examine fixed cooperative protocols in which the relay either amplifies what it receives, or fully decodes, re-encodes, and re-transmits the source message. We refer to these options as *amplify-and-forward* and *decode-and-forward*, respectively, with the caveat that the algorithms developed in this chapter are special cases of the more involved communication strategies discussed in Chapter 4. Our analysis suggests that cooperating terminals may also employ threshold tests on the measured SNR between them, to obtain adaptive protocols that choose the strategy with best performance. In addition, protocols based upon limited feedback from the destination terminal are also developed.

We evaluate performance of these protocols in terms of outage probability [59] in the presence of slow fading and compare to appropriate performance bounds. Each of our cooperative protocols achieve *full* (*i.e.*, second-order in the case of two terminals) diversity; that is, the outage probability decays proportional to $1/\mathtt{SNR}^2$, where $\mathtt{SNR}$ is signal-to-noise ratio of the channel, while it decays proportional to $1/\mathtt{SNR}$ without cooperation. At fixed low rates, the schemes without feedback are at most 1.5 dB from optimal and offer substantial power savings over direct transmission. For sufficiently high rates, direct transmission becomes preferable to the protocols without feedback because they essentially repeat information all the time and are bandwidth inefficient as a result. Protocols that exploit limited feedback overcome this bandwidth inefficiency by repeating only rarely. The degree to which these protocols are optimal among all cooperative schemes remains an open question, especially for high rates. More broadly, the relative attractiveness of amplify-and-forward and decode-and-forward, and adaptive versions thereof, can depend upon the network architecture and implementation considerations.

An outline of the chapter is as follows. In Section 5.1, we describe a system model for the wireless networks under consideration. The model exhibits slow frequency nonselective fading to capture performance when delay constraints are on the order of the coherence time of the channel. While we focus on a pair of cooperating terminals, the orthogonality conditions we impose for integration into existing wireless standards allow the transmission protocols developed in Section 5.2 to be generalized to multiple cooperating terminals and multiple relays. We develop one extension among many possible in Chapter 6. In Section 5.2,

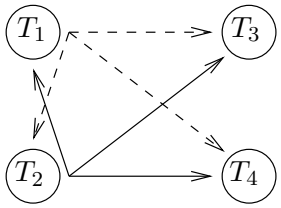


Figure 5-1: Illustration of radio signal transmit paths in an example wireless network with two terminals transmitting information and two terminals receiving information.

we describe fixed amplify-and-forward transmission and decode-and-forward transmission, adaptive versions of these, and protocols that exploit limited feedback from the destination terminal. Section 5.3 characterizes the outage behavior of the various protocols in terms of outage events and outage probabilities, using several results for exponential random variables in Appendix B. Section 5.4 compares the results from a number of perspectives and offers some concluding remarks.

5.1 System Model

We focus on the case of two cooperating terminals communicating to either the same or separate destination terminals. Fig. 5-1 again depicts this setting, shown previously in Fig. 1-1. In our model for the wireless channel in Fig. 5-1, narrowband transmissions suffer the effects of frequency nonselective fading and additive noise. Our analysis in Section 5.3 focuses on the case of slow fading, and measures performance by outage probability, to isolate the benefits of space diversity. Our cooperative protocols can be naturally extended to the kinds of wide-band and highly mobile scenarios in which frequency- and time-selective fading, respectively, are encountered; however, we expect the potential impact of our protocols becomes less substantial as other forms of diversity can be exploited in the system. A detailed study of the relative benefits of cooperative diversity in conjunction with temporal and spectral diversity represents an important area of future research, but remains beyond the scope of the dissertation.

5.1.1 Medium Access

Medium-access control in our model imposes the practical system constraints or orthogonal transmission and half-duplex relaying, as discussed in Chapter 2. In this section, we motivate these constraints more fully and describe the medium-access employed by our algorithms.

As in many current wireless networks, we divide the available bandwidth into orthogonal channels and allocate these channels to the transmitting terminals. The medium-access control (MAC) sublayer typically performs this function. For example, the MAC in many cellular networks seeks to allocate orthogonal channels, *e.g.*, frequency-division, time-division, or code-division, to the terminals in a cell for communicating to the basestation of that cell. As another example, the MAC in the IEEE 802.11 wireless LAN standard uses similar structures for LANs controlled by an access point, or a distributed contention-resolution/collision avoidance algorithm which facilitates random time-division.

We maintain this division into orthogonal channels in the sequel to allow our transmission protocols to be readily integrated into existing networks. As a convenient by-product of this choice, we are able to treat the multiple-access (single receiver) and interference (multiple receivers) cases described in Section 1.2 simultaneously, as a pair of point-to-point channels with signaling between the transmitters. Furthermore, removing the interference between the terminals at the destination radio(s) substantially simplifies the receiver algorithms and the outage analysis for purposes of exposition.

For our cooperative diversity protocols described in Section 5.2, transmitting terminals must also process their received signals; however, as current limitations in radio implementation preclude the terminals from transmitting and receiving at the same time in the same frequency band. Because of severe signal attenuation over the wireless channel, and insufficient electrical isolation between the transmit and receive circuitry, a terminal's transmitted signal drowns out the signals of other terminals at its receiver input.[1] Thus, we further divide each channel into orthogonal subchannels. Fig. 5-2 illustrates our channel allocation for an example time-division approach with two terminals. As Fig. 5-2(c) suggests, and we further develop in Section 5.2, our cooperative diversity protocols can be adaptive: based upon the SNR between the transmitting terminals, they can decide to continue their own

[1] Typically a terminal's transmit signal is $100 - 150$ dB above its received signal.

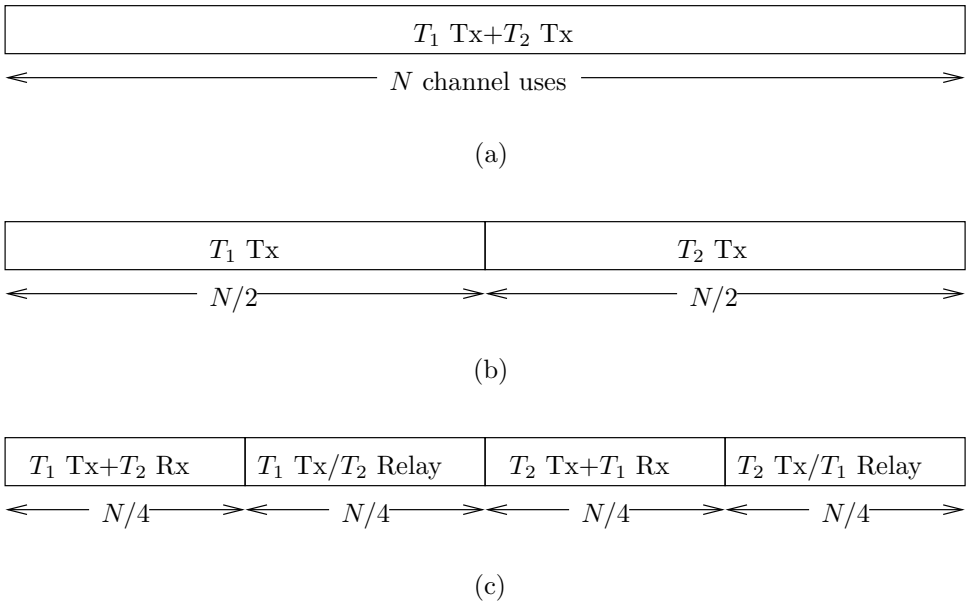


Figure 5-2: Example time-division channel allocations for (a) direct transmission with interference, (b) orthogonal direct transmission, and (c) orthogonal cooperative diversity transmission protocols. We focus on orthogonal transmissions of the form (b) and (c) throughout the chapter.

transmission or relay for one another.

5.1.2 Equivalent Channel Models

Under the above orthogonality constraints, we can now conveniently, and without loss of generality, characterize our channel models using a time-division notation; frequency-division counterparts to this model are straightforward. Due to the symmetry of the channel allocations, we focus on the message of the "source" terminal T_s, which potentially employs terminal T_r as a "relay" terminal, in transmitting to the "destination" terminal T_d, where $s, r \in \{1, 2\}$ and $d(s) \in \{3, 4\}$. We utilize a baseband-equivalent, discrete-time channel model for the continuous-time channel, and we consider N consecutive uses of the channel, where N is a large integer.

For direct transmission, our baseline for comparison, we model the channel as

$$y_d[n] = a_{s,d(s)}\, x_s[n] + z_d[n] \tag{5.1}$$

for, say, $n = 1, \ldots, N/2$, where $x_s[n]$ is the source transmitted signal, and $y_d[n]$ is the

destination received signal. The other terminal transmits for $n = N/2+1, \ldots, N$ as Fig. 5-2(b) depicts. Thus, in the baseline system each terminal utilizes only half of the available degrees of freedom of the channel.

For cooperative diversity transmission, we model the channel during the first half of the block as

$$y_r[n] = a_{s,r}\, x_s[n] + z_r[n] \tag{5.2}$$

$$y_d[n] = a_{s,d(s)}\, x_s[n] + z_d[n] \tag{5.3}$$

for, say, $n = 1, \ldots, N/4$, where $x_s[n]$ is the source transmitted signal and $y_r[n]$ and $y_d[n]$ are the relay and destination received signals, respectively. For the second half of the block, we model the received signal as

$$y_d[n] = a_{r,d(s)}\, x_r[n] + z_d[n] \tag{5.4}$$

for $n = N/4 + 1, \ldots, N/2$, where $x_r[n]$ is the relay transmitted signal and $y_d[n]$ is the destination received signal. A similar setup is employed in the second half of the block, with the roles of the source and relay reversed, as Fig. 5-2(c) depicts. Note that, while again half the degrees of freedom are allocated to each source terminal for transmission to its destination, only a quarter of the degrees of freedom are available for communication to its relay.

In (5.1)-(5.4), $a_{i,j}$ captures the effects of path-loss, shadowing, and frequency nonselective fading, and $z_j[n]$ captures the effects of receiver noise and other forms of interference in the system, where $i \in \{s, r\}$ and $j \in \{r, d\}$. We consider the scenario in which the fading coefficients are known to, *i.e.*, accurately measured by, the appropriate receivers, but not fully known to (or not exploited by) the transmitters. Statistically, we model $a_{i,j}$ as zero-mean, independent, circularly-symmetric complex Gaussian random variables with variances $\sigma_{i,j}^2$, so that the magnitudes $|a_{i,j}|$ are Rayleigh distributed ($|a_{i,j}|^2$ are exponentially distributed with mean $\sigma_{i,j}^2$) and the phases $\measuredangle a_{i,j}$ are uniformly distributed on $[0, 2\pi)$. Furthermore, we model $z_j[n]$ as zero-mean mutually independent, circularly-symmetric, complex Gaussian random sequences with variance N_0.

5.1.3 Parameterizations

Two important parameters of the system are the SNR without fading and the spectral efficiency. We now define these parameters in terms of standard parameters in the continuous-time channel. For a continuous-time channel with bandwidth W Hz available for transmission, the discrete-time model contains W two-dimensional symbols per second (2D/s).

If the transmitting terminals have an average power constraint in the continuous-time channel model of P_c Joules/s, we see that this translates into a discrete-time power constraint of $P = 2P_c/W$ Joules/2D since each terminal transmits in half of the available degrees of freedom, under both direct and cooperative diversity transmission. Thus, the channel model is parameterized by the SNR random variables $\mathtt{SNR}\,|\mathsf{a}_{i,j}|^2$, where

$$\mathtt{SNR} \stackrel{\triangle}{=} \frac{2P_c}{N_0 W} = \frac{P}{N_0} \tag{5.5}$$

is the SNR without fading.

In addition to SNR, transmission schemes are further parameterized by the transmission rate r b/s, or spectral efficiency

$$\mathtt{R} \stackrel{\triangle}{=} 2r/W \text{ b/s/Hz} \tag{5.6}$$

attempted by the transmitting terminals. Note that (5.6) is the transmission rate normalized by the number of degrees of freedom utilized by each terminal, not by the total number of degrees of freedom in the channel.

Nominally, one could parameterize the system by the pair $(\mathtt{SNR}, \mathtt{R})$; however, our results lend more insight, and are substantially more compact, when we parameterize the system by either of the pairs $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ or $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$, where[2]

$$\mathtt{SNR}_{\text{norm}} \stackrel{\triangle}{=} \frac{\mathtt{SNR}}{2^{\mathtt{R}} - 1}\,, \qquad \mathtt{R}_{\text{norm}} \stackrel{\triangle}{=} \frac{\mathtt{R}}{\log\left(1 + \mathtt{SNR}\,\sigma^2_{s,d(s)}\right)}\,. \tag{5.7}$$

For an additive white Gaussian noise (AWGN) channel with bandwidth $(W/2)$ and SNR $\mathtt{SNR}\,\sigma^2_{s,d(s)}$, $\mathtt{SNR}_{\text{norm}} > 1$ is the SNR normalized by the *minimum* SNR required to achieve spectral efficiency $\mathtt{R}$ [23]. Similarly, $\mathtt{R}_{\text{norm}} < 1$ is the spectral efficiency normalized by the

[2]Unless otherwise indicated, logarithms in this chapter are taken to base 2.

maximum achievable spectral efficiency, *i.e.*, channel capacity [93]. In this sense, parameterizations given by $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ and $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$ are duals of one another. In our setting with fading, the two parameterizations yield tradeoffs between different aspects of system performance: results under $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ exhibit a tradeoff between the normalized SNR gain and spectral efficiency of a protocol, while results under $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$ exhibit a tradeoff between the diversity order and normalized spectral efficiency of a protocol.

We note that communication engineers are familiar with parameterizing systems by $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$. Results are often displayed for a given modulation and coding scheme with fixed $\mathtt{R}$ for varying $\mathtt{SNR}_{\text{norm}}$. An example is the bit-error rate of binary phase-shift keying for different $\mathtt{SNR}_{\text{norm}}$. For systems parameterized by $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$, varying $\mathtt{SNR}$ with $\mathtt{R}_{\text{norm}}$ fixed actually implies different coding and modulation schemes with different $\mathtt{R}$. An example could be binary phase-shift keying for low SNR and quadrature phase-shift keying for higher SNR. While this parameterization may seem unnatural for comparing performance within a given family of systems, it is useful for comparing systems in two different families operating at the same $\mathtt{SNR}$ and $\mathtt{R}_{\text{norm}}$.

Note that, although we have parameterized the transmit powers and noise levels to be symmetric throughout the network for purposes of exposition, asymmetries in average SNR and path-loss can be lumped into the fading variances $\sigma_{i,j}^2$. Furthermore, while the tools are powerful enough to consider general rate pairs $(\mathtt{R}_1, \mathtt{R}_2)$, we consider the equal rate point, *i.e.*, $\mathtt{R}_1 = \mathtt{R}_2 = \mathtt{R}$, for purposes of exposition.

5.2 Cooperative Transmission Protocols

In this section, we describe a variety of simple cooperative transmission strategies that can be utilized in the network of Fig. 5-1. These protocols employ different types of processing by the relay terminals, as well as different types of combining at the destination terminals. For fixed cooperative protocols, we allow the relays to either amplify their received signals subject to their power constraint, or to decode, re-encode, and re-transmit the messages. Again, we refer to these two options generally as amplify-and-forward and decode-and-forward, respectively. In addition to fixed strategies, we consider, among many possible adaptive strategies, simple protocols in which the cooperating terminals accurately estimate the realized SNR between them and use this estimate to select a suitable cooper-

ative action; the terminals decide between continuing their own transmission, or relaying the transmissions of the other terminal using amplify-and-forward or decode-and-forward. With direct, decode-and-forward, and adaptive combinations of the two, the radios may employ repetition or more powerful codes. In any of these cases, destination radios can appropriately combine their received signals by exploiting control information in the protocol headers. We stress that adaptation is performed in the absence of feedback from the destination terminal; we also describe one simple protocol that exploits limited feedback from the destination.

5.2.1 Fixed Protocols

Amplify-and-Forward Transmission

For amplify-and-forward transmission, the appropriate channel model is (5.2)–(5.4). The source terminal transmits its information as $x_s[n]$, say, for $n = 1, \ldots, N/4$. During this interval, the relay processes $y_r[n]$, and relays the information by transmitting

$$x_r[n] = \beta \, y_r[n - N/4] \ , \tag{5.8}$$

for $n = N/4 + 1, \ldots, N/2$. To remain within its power constraint (with high probability), an amplifying relay must use gain

$$\beta \leq \sqrt{\frac{P}{|a_{s,r}|^2 \, P + N_0}} \ ,$$

where we allow the amplifier gain to depend upon the fading coefficient $a_{s,r}$ between the source and relay, which the relay estimates to high accuracy. This transmission scheme can be viewed as repetition coding from two separate transmitters, except that the relay transmitter amplifies its own receiver noise. The destination can decode its received signal $y_d[n]$ for $n = 1, \ldots, N/2$ by first appropriately combining the signals from the two subblocks using a suitably designed matched-filter (maximum-ratio combiner) [49].

Decode-and-Forward Transmission

For decode-and-forward transmission, the appropriate channel model is again (5.2)–(5.4). The source terminal transmits its information as $x_s[n]$, say, for $n = 0, \ldots, N/4$. During this

interval, the relay processes $y_r[n]$ by decoding an estimate $\hat{x}_s[n]$ of the source transmitted signal.

Under a repetition coded scheme, the relay transmits the signal

$$x_r[n] = \hat{x}_s[n - N/4]$$

for $n = N/4 + 1, \ldots, N/2$.

Decoding at the relay can take on a variety of forms. For example, the relay might fully decode the source message by estimating the source codeword, or it might employ symbol-by-symbol decoding and allow the destination to perform full decoding. These options allow for trading off performance and complexity at the relay terminal. Because the performance of symbol-by-symbol decoding varies with the choice of coding and modulation, we focus on full decoding in the sequel; symbol-by-symbol decoding of binary transmissions has been treated from uncoded perspective in [49].

5.2.2 Adaptive Protocols

As we might expect, and the analysis in Section 5.3 confirms, fixed decode-and-forward is limited by direct transmission between the source and relay. However, since the fading coefficients are known to the appropriate receivers, $a_{s,r}$ can be measured to high accuracy by the cooperating terminals; thus, they can adapt their transmission format according to the realized value of $a_{s,r}$.

This observation suggests the following class of adaptive algorithms. If the measured $|a_{s,r}|^2$ falls below a certain threshold, the source simply continues its transmission to the destination, in the form of repetition or more powerful codes. If the measured $|a_{s,r}|^2$ lies above the threshold, the relay forwards what it received from the source, using either amplify-and-forward or decode-and-forward, in an attempt to achieve diversity gain.

Adaptive protocols of this form should offer diversity because in either case, two of the fading coefficients must be small in order for the transmission to be lost. Specifically, if $|a_{s,r}|^2$ is small, then $|a_{s,d(s)}|^2$ must also be small for the transmission to be lost when the source continues its transmission. Similarly, if $|a_{s,r}|^2$ is large, then both $|a_{s,d(s)}|^2$ and $|a_{r,d(s)}|^2$ must be small for the transmission to be lost when the relay employs amplify-and-forward or decode-and-forward. We formalize this notion when we consider outage performance of

adaptive protocols in Section 5.3.

5.2.3 Protocols with Limited Feedback

As we will see, the fixed and adaptive protocols described above can make inefficient use of the degrees of freedom of the channel, especially for high transmission rates, because the relays essentially repeat the transmissions all the time. In this section, we describe a very simple protocol that exploits limited feedback from the destination terminal, *e.g.*, a single bit indicating the success or failure of the direct transmission, that we will see can dramatically improve spectral efficiency over the fixed and adaptive protocols. A complete treatment of protocols for this and more general scenarios with feedback is beyond the scope of the dissertation.

As one example, consider the following protocol utilizing feedback and amplify-and-forward transmission. Protocols based upon feedback and decode-and-forward transmission are also possible, but the analysis is more involved and their performance is slightly worse that the following protocol.

We nominally allocate the channels according to Fig. 5-2(b). First, the source transmits its information to the destination at spectral efficiency `R`. The destination indicates success or failure by broadcasting a single bit of feedback to the source and relay, which we assume is detected reliably by at least the relay.[3] If the SNR between the source and destination is sufficiently high, the feedback indicates success of the direct transmission, and the relay does nothing. If the SNR between the source and destination is not sufficiently high for successful direct transmission, the feedback requests that the relay amplify-and-forward what it received from the source. In the latter case, the destination tries to combine the two transmissions. As we will see, protocols of this form make more efficient use of the degrees of freedom of the channel, because they repeat only rarely.

[3]Such an assumption is reasonable if the destination encodes the feedback bit with a very low-rate code. Even if the relay cannot reliably decode, useful protocols can be developed and analyzed. For example, a conservative protocol might have the relay amplify-and-forward what it receives from the source in all cases except when the destination reliably receives the direct transmission *and* the relay reliably decodes the feedback bit.

5.3 Outage Behavior

For fixed fading values, the effective channel models induced by the transmission protocols described in Section 5.2 are variants of well-known channels with additive white Gaussian noise. In this section, we compare the performance of the various transmission protocols in terms of outage events and outage probabilities [59, 57], and focus our attention on performance in the high SNR regime. Outage events specified in terms of the fading random variables $|a_{i,j}|^2$ have useful interpretations in both coded and uncoded settings, but we will develop our results from a coded perspective and determine events in which the realized mutual information of the channel falls below a target transmission rate. We convert this event into an equivalent event defined in terms of the fading coefficients of the channel.

Since the channel average mutual information I is a function of the fading coefficients of the channel, it too is a random variable. The event $I < \mathtt{R}$ that this mutual information random variable falls below some fixed spectral efficiency $\mathtt{R}$, is referred to as an *outage event*, because reliable communication is not possible for realizations in this event.[4] The probability of an outage event, $\Pr[I < \mathtt{R}]$, is referred to as the *outage probability* of the channel.

We note that outage events are independent of the distribution of the underlying random variables, while outage probabilities are intimately tied to them. For example, if the outage event of a scheme at a particular rate is a strict subset of the outage event of another scheme at that rate, then the first scheme has smaller outage probability regardless of the probability distribution on the channel parameters. Furthermore, as we will see, several of our cooperation strategies have similar outage probabilities, but the structure of their outage events is sufficiently different that we might prefer one over the other in various regimes. As a result, both outage events and outage probabilities are useful for characterizing our transmission protocols.

We now develop outage events and outage probabilities for our transmission protocols. To facilitate their comparison in the sequel, we also derive high SNR approximations of the outage probabilities using results from Appendix B.

[4] When specified in terms of the fading random variables of the channel, the outage event is sometimes called an *outage region*.

5.3.1 Direct Transmission

To establish baseline performance, under direct transmission, the source terminal transmits over the channel (5.1). The maximum average mutual information between input and output in this case, achieved by independent and identically-distributed (i.i.d.) zero-mean, circularly-symmetric complex Gaussian inputs, is given by

$$I_D = \log\left(1 + \mathtt{SNR}\,|a_{s,d(s)}|^2\right) \tag{5.9}$$

as a function of the fading coefficient $a_{s,d(s)}$. The outage event for spectral efficiency $\mathtt{R}$ is given by $I_D < \mathtt{R}$ and is equivalent to the event

$$|a_{s,d(s)}|^2 < \frac{2^{\mathtt{R}} - 1}{\mathtt{SNR}} \ . \tag{5.10}$$

For Rayleigh fading, *i.e.*, $|a_{s,d(s)}|^2$ exponentially distributed with parameter $\sigma_{s,d(s)}^{-2}$, the outage probability satisfies[5]

$$\begin{aligned} p_D^{\text{out}}(\mathtt{SNR}, \mathtt{R}) &\stackrel{\triangle}{=} \Pr\left[I_D < \mathtt{R}\right] = \Pr\left[|a_{s,d(s)}|^2 < \frac{2^{\mathtt{R}} - 1}{\mathtt{SNR}}\right] \\ &= 1 - \exp\left(-\frac{2^{\mathtt{R}} - 1}{\mathtt{SNR}\,\sigma_{s,d(s)}^2}\right) \\ &\sim \frac{1}{\sigma_{s,d(s)}^2} \cdot \frac{2^{\mathtt{R}} - 1}{\mathtt{SNR}}, \quad \mathtt{SNR} \text{ large} \ , \end{aligned} \tag{5.11}$$

where we have utilized the results of Fact 1 in Appendix B with $\lambda = 1/\sigma_{s,d(s)}^2$, $t = \mathtt{SNR}$, and $g(t) = (2^{\mathtt{R}} - 1)/t$.

Note the $1/\mathtt{SNR}$ behavior in (5.11), which implies that increasing $\mathtt{SNR}$ by 10 dB reduces the outage probability by only a factor of 10. We will see that our cooperative diversity protocols decrease the outage probability by roughly a factor of 100 when $\mathtt{SNR}$ is increased by 10 dB, for $\mathtt{SNR}$ large.

[5] As we develop more formally in Appendix B, the approximation $f(\mathtt{SNR}) \sim g(\mathtt{SNR})$, $\mathtt{SNR}$ large, is in the sense of $f(\mathtt{SNR})/g(\mathtt{SNR}) \to 1$ as $\mathtt{SNR} \to \infty$. Thus, the approximation can be made as accurate as we like for $\mathtt{SNR}$ sufficiently large.

5.3.2 Fixed Protocols

Amplify-and-Forward Transmission

The amplify-and-forward protocol produces an equivalent one-input, two-output complex Gaussian noise channel with different noise levels in the outputs. As Appendix C.1 details, the maximum average mutual information between the input and the two outputs, achieved by i.i.d.complex Gaussian inputs, is given by

$$I_{AF} = \frac{1}{2}\log\left(1 + \mathtt{SNR}\,|a_{s,d(s)}|^2 + f\left(\mathtt{SNR}\,|a_{s,r}|^2,\ \mathtt{SNR}\,|a_{r,d(s)}|^2\right)\right) \tag{5.12}$$

as a function of the fading coefficients, where

$$f(x,y) \stackrel{\triangle}{=} \frac{xy}{x+y+1}\ . \tag{5.13}$$

The outage event for spectral efficiency $\mathtt{R}$ is given by $I_{AF} < \mathtt{R}$ and is equivalent to the event

$$|a_{s,d(s)}|^2 + \frac{1}{\mathtt{SNR}} f\left(\mathtt{SNR}\,|a_{s,r}|^2,\ \mathtt{SNR}\,|a_{r,d(s)}|^2\right) < \frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}\ . \tag{5.14}$$

For Rayleigh fading, *i.e.*, $|a_{i,j}|^2$ independent and exponentially distributed with parameters $\sigma_{i,j}^{-2}$, analytic calculation of the outage probability becomes involved, but we can approximate its high SNR behavior as

$$p_{AF}^{\text{out}}(\mathtt{SNR},\mathtt{R}) \stackrel{\triangle}{=} \Pr\left[I_{AF} < \mathtt{R}\right] \sim \left(\frac{1}{2\sigma_{s,d(s)}^2}\frac{\sigma_{s,r}^2+\sigma_{r,d(s)}^2}{\sigma_{s,r}^2\sigma_{r,d(s)}^2}\right)\cdot\left(\frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}\right)^2\ ,\quad \mathtt{SNR}\text{ large}\ , \tag{5.15}$$

where we have utilized the results of Claim 1 in Appendix B, with

$$\begin{gathered} u = |a_{s,d(s)}|^2,\quad v = |a_{s,r}|^2,\quad w = |a_{r,d(s)}|^2 \\ \lambda_u = \sigma_{s,d(s)}^{-2},\quad \lambda_v = \sigma_{s,r}^{-2},\quad \lambda_w = \sigma_{r,d(s)}^{-2} \\ g(\epsilon) = (2^{2\mathtt{R}}-1)\epsilon,\quad t = \mathtt{SNR},\quad h(t) = 1/t\ . \end{gathered}$$

The $1/\mathtt{SNR}^2$ behavior in (5.15) indicates that amplify-and-forward achieves full second-order diversity. Thus, increasing $\mathtt{SNR}$ by 10 dB reduces the outage probability by a factor of 100.

Decode-and-Forward Transmissions

To analyze decode-and-forward transmission, we examine a particular decoding structure at the relay. Specifically, we require the relay to fully decode the source message; examination of symbol-by-symbol decoding at the relay becomes involved because it depends upon the particular coding and modulation choices. Requiring both the relay and destination to decode perfectly, the maximum average mutual information for repetition-coded decode-and-forward can be readily shown to be

$$I_{DF} = \frac{1}{2} \min \left\{ \log \left(1 + \mathtt{SNR} \left| a_{s,r} \right|^2 \right), \ \log \left(1 + \mathtt{SNR} \left| a_{s,d(s)} \right|^2 + \mathtt{SNR} \left| a_{r,d(s)} \right|^2 \right) \right\} \tag{5.16}$$

as a function of the fading random variables. The first term in (5.16) represents the maximum rate at which the relay can reliably decode the source message, while the second term in (5.16) represents the maximum rate at which the destination can reliably decode the source message given repeated transmissions from the source and destination. We note that such mutual information forms are typical of relay channels with full decoding at the relay [17].

The outage event for spectral efficiency $\mathtt{R}$ is given by $I_{DF} < \mathtt{R}$ and is equivalent to the event

$$\min \left\{ \left| a_{s,r} \right|^2, \left| a_{s,d(s)} \right|^2 + \left| a_{r,d(s)} \right|^2 \right\} < \frac{2^{2\mathtt{R}} - 1}{\mathtt{SNR}} \ . \tag{5.17}$$

For Rayleigh fading, *i.e.*, $|a_{i,j}|^2$ independent and exponentially distributed with parameters $\sigma_{i,j}^{-2}$, the outage probability for repetition-coded decode-and-forward can be computed according to

$$\begin{aligned} p_{DF}^{\text{out}} \left(\mathtt{SNR}, \mathtt{R} \right) &\stackrel{\triangle}{=} \Pr \left[I_{DF} < \mathtt{R} \right] \\ &= \Pr \left[\left| a_{s,r} \right|^2 < g(\mathtt{SNR}) \right] + \Pr \left[\left| a_{s,r} \right|^2 \geq g(\mathtt{SNR}) \right] \Pr \left[\left| a_{s,d(s)} \right|^2 + \left| a_{r,d(s)} \right|^2 < g(\mathtt{SNR}) \right] \end{aligned} \tag{5.18}$$

where $g(\mathtt{SNR}) = [2^{2\mathtt{R}} - 1]/\mathtt{SNR}$. Although we may readily compute a closed form expression for (5.18), for compactness we examine the large $\mathtt{SNR}$ behavior of (5.18) by computing the

limit

$$\begin{aligned}\frac{1}{g(\mathtt{SNR})}\,p_{DF}^{\text{out}}\left(\mathtt{SNR},\mathtt{R}\right) &= \underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[|a_{s,r}|^2 < g(\mathtt{SNR})\right]}_{\to 1/\sigma_{s,r}^2} \\ &\quad + \underbrace{\Pr\left[|a_{s,r}|^2 \geq g(\mathtt{SNR})\right]}_{\to 1}\,\underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[|a_{s,d(s)}|^2 + |a_{r,d(s)}|^2 < g(\mathtt{SNR})\right]}_{\to 0} \\ &\to 1/\sigma_{s,r}^2\end{aligned}$$

as $\mathtt{SNR} \to \infty$, using the results of Facts 1 and 2 in Appendix B. Thus, we conclude that

$$p_{DF}^{\text{out}}\left(\mathtt{SNR},\mathtt{R}\right) \sim \frac{1}{\sigma_{s,r}^2}\cdot\frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}, \quad \mathtt{SNR}\text{ large.} \tag{5.19}$$

The $1/\mathtt{SNR}$ behavior in (5.19) indicates that fixed decode-and-forward does not offer diversity gains for large $\mathtt{SNR}$, because requiring the relay to fully decode the source transmissions limits the performance of decode-and-forward transmission to that of direct transmission between the source and relay.

5.3.3 Adaptive Protocols

To overcome the shortcomings of decode-and-forward transmission, we described adaptive versions of the amplify-and-forward and decode-and-forward protocols, both of which fall back to direct transmission if $|a_{s,r}|^2$ falls below some threshold.

As an example analysis, we analyze the performance of adaptive decode-and-forward transmission. The mutual information of this adaptive hybrid is somewhat involved to write down, but in the case of repetition coding at the relay, can be readily shown to be

$$I_{ADF} = \begin{cases} \frac{1}{2}\log\left(1+2\,\mathtt{SNR}|a_{s,d(s)}|^2\right) & |a_{s,r}|^2 < g(\mathtt{SNR}) \\ \frac{1}{2}\log\left(1+\mathtt{SNR}|a_{s,d(s)}|^2+\mathtt{SNR}|a_{r,d(s)}|^2\right) & |a_{s,r}|^2 \geq g(\mathtt{SNR}) \end{cases} \tag{5.20}$$

where $g(\mathtt{SNR}) = [2^{2\mathtt{R}}-1]/\mathtt{SNR}$. The first case in (5.20) corresponds to the maximum average mutual information of repetition coding from the source to the destination, hence the extra factor of 2 in the SNR. The second case in (5.20) corresponds to the maximum average mutual information of repetition coding from the source and relay to the destination, assuming

the relay can fully decode the source transmission.

The outage event for spectral efficiency $\mathtt{R}$ is given by $I_{ADF} < \mathtt{R}$ and is equivalent to the event

$$\left(\{|a_{s,r}|^2 < g(\mathtt{SNR})\} \bigcap \{2|a_{s,d(s)}|^2 < g(\mathtt{SNR})\}\right) \bigcup \left(\{|a_{s,r}|^2 \geq g(\mathtt{SNR})\} \bigcap \{|a_{s,r}|^2 + |a_{r,d(s)}|^2 < g(\mathtt{SNR})\}\right) . \quad (5.21)$$

The first (resp. second) event of the union in (5.21) corresponds to the first (resp. second) case in (5.20). We observe that adapting to the realized fading coefficient ensures that the protocol performs no worse than direct transmission, except for the fact that it potentially suffers the bandwidth inefficiency of repetition coding.

Because the events in the union of (5.21) are mutually exclusive, the outage probability becomes the sum

$$\begin{aligned} p_{ADF}^{\text{out}}(\mathtt{SNR}, \mathtt{R}) &\stackrel{\triangle}{=} \Pr\left[I_{ADF} < \mathtt{R}\right] \\ &= \Pr\left[|a_{s,r}|^2 < g(\mathtt{SNR})\right] \Pr\left[2|a_{s,d(s)}|^2 < g(\mathtt{SNR})\right] \\ &\quad + \Pr\left[|a_{s,r}|^2 \geq g(\mathtt{SNR})\right] \Pr\left[|a_{s,d(s)}|^2 + |a_{r,d(s)}|^2 < g(\mathtt{SNR})\right] , \end{aligned} \quad (5.22)$$

and we may readily compute a closed form expression for (5.22). For comparison to our other protocols, we compute the large $\mathtt{SNR}$ behavior of (5.22) by computing the limit

$$\begin{aligned} \frac{1}{g^2(\mathtt{SNR})} p_{ADF}^{\text{out}}(\mathtt{SNR}, \mathtt{R}) &= \underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[|a_{s,r}|^2 < g(\mathtt{SNR})\right]}_{\to 1/\sigma_{s,r}^2} \underbrace{\frac{1}{g(\mathtt{SNR})}\Pr\left[2|a_{s,d(s)}|^2 < g(\mathtt{SNR})\right]}_{\to 1/(2\sigma_{s,d(s)}^2)} \\ &\quad + \underbrace{\Pr\left[|a_{s,r}|^2 \geq g(\mathtt{SNR})\right]}_{\to 1} \underbrace{\frac{1}{g^2(\mathtt{SNR})}\Pr\left[|a_{s,d(s)}|^2 + |a_{r,d(s)}|^2 < g(\mathtt{SNR})\right]}_{\to 1/(2\sigma_{s,d(s)}^2\sigma_{r,d(s)}^2)} \\ &\to \left(\frac{1}{2\sigma_{s,d(s)}^2} \frac{\sigma_{s,r}^2 + \sigma_{r,d(s)}^2}{\sigma_{s,r}^2\sigma_{r,d(s)}^2}\right) \end{aligned} \quad (5.23)$$

as $\mathtt{SNR} \to \infty$, using the results of Facts 1 and 2 of Appendix B. Thus, we conclude that the large $\mathtt{SNR}$ performance of adaptive decode-and-forward transmission is *identical* to that of fixed amplify-and-forward transmission.

Analysis of more general adaptive protocols becomes involved because there are ad-

ditional degrees of freedom in choosing the thresholds for switching between the various options such as direct, amplify-and-forward, and decode-and-forward. A detailed analysis of such protocols is beyond the scope of this chapter.

5.3.4 Bounds for Cooperative Diversity Transmission

We now develop performance limits for fixed and adaptive cooperative diversity transmission protocols. If we suppose that the source and relay know each other's messages *a priori*, then instead of our direct transmission protocol, each would benefit from using a space-time code for two transmit antennas. In this sense, the outage probability of conventional single-user transmit diversity [57] represents an (optimistic) lower bound on the outage probability of our cooperative diversity protocols. The following two sections develop two such bounds: an unconstrained transmit diversity bound, and an orthogonal transmit diversity bound that takes into account orthogonality constraints at the relay.

Transmit Diversity Bound

To utilize a space-time code for each terminal, we allocate the channel as in Fig. 5-2(b). Both terminals transmit in all the degrees of freedom of the channel, so their transmitted power is $P/2$ Joules/2D, half that of direct transmission. The spectral efficiency for each terminal remains $\mathtt{R}$.

For transmit diversity, we model the channel as

$$y_d[n] = \begin{bmatrix} a_{s,d(s)} & a_{r,d(s)} \end{bmatrix} \begin{bmatrix} x_s[n] \\ x_r[n] \end{bmatrix} + z_d[n] \ , \tag{5.24}$$

for, say, $n = 0, \ldots, N/2$. As developed in Appendix C.2, an optimal signaling strategy, in terms of minimizing outage probability in the large $\mathtt{SNR}$ regime, is to encode information using $\begin{bmatrix} x_s & x_r \end{bmatrix}^{\mathrm{T}}$ i.i.d.complex Gaussian, each with power $P/2$. Using this result, the maximum average mutual information as a function of the fading coefficients is given by

$$I_T = \log\left(1 + \frac{\mathtt{SNR}}{2}\left[|a_{s,d(s)}|^2 + |a_{r,d(s)}|^2\right]\right) \ . \tag{5.25}$$

The outage event $I_T < \mathtt{R}$ is equivalent to the event

$$|a_{s,d(s)}|^2 + |a_{r,d(s)}|^2 < \frac{2^{\mathtt{R}} - 1}{(\mathtt{SNR}/2)} \ . \tag{5.26}$$

For $|a_{i,j}|^2$ exponentially distributed with parameters $\sigma_{i,j}^{-2}$, the outage probability satisfies

$$\begin{aligned} p_T^{\text{out}}(\mathtt{SNR}, \mathtt{R}) &\stackrel{\triangle}{=} \Pr[I_T < \mathtt{R}] \\ &\sim \frac{2}{\sigma_{s,d(s)}^2 \sigma_{r,d(s)}^2} \cdot \left(\frac{2^{\mathtt{R}} - 1}{\mathtt{SNR}}\right)^2, \quad \mathtt{SNR} \text{ large} \ , \end{aligned} \tag{5.27}$$

where we have applied the results of Fact 2 in Appendix B.

Orthogonal Transmit Diversity Bound

The transmit diversity bound (5.27) does not take into account the inability of a relay to simultaneously transmit and receive in the same band. To capture this effect, we constrain the transmit diversity scheme to be orthogonal, so that the transmissions of the source and relay are emitted and received over parallel channels.

When the source and relay can cooperate perfectly, an equivalent model to (5.24), incorporating the relay orthogonality constraint, consists of parallel channels

$$y_d[n] = a_{s,d(s)}\, x_s[n] + z_d[n], \qquad n = 0, \ldots, N/4 \tag{5.28}$$

$$y_d[n] = a_{r,d(s)}\, x_r[n] + z_d[n], \qquad n = N/4 + 1, \ldots, N/2 \tag{5.29}$$

This set of parallel channels is utilized half as many times as the corresponding direct transmission channel, so the source must transmit at twice spectral efficiency in order to achieve the same spectral efficiency as direct transmission.

For each fading realization, the maximum average mutual information can be obtained using independent complex Gaussian inputs. Allocating a fraction α of the power to x_s, and the remaining fraction $(1 - \alpha)$ of the power to x_r, the average mutual information is given by

$$I_P = \frac{1}{2} \log\left[\left(1 + 2\alpha \mathtt{SNR} |a_{s,d(s)}|^2\right)\left(1 + 2(1-\alpha)\mathtt{SNR} |a_{r,d(s)}|^2\right)\right] \ , \tag{5.30}$$

The outage event $I_P < \mathtt{R}$ is equivalent to the outage region

$$\alpha|\mathsf{a}_{s,d(s)}|^2 + (1-\alpha)|\mathsf{a}_{r,d(s)}|^2 + 2\alpha(1-\alpha)\mathtt{SNR}|\mathsf{a}_{s,d(s)}|^2|\mathsf{a}_{r,d(s)}|^2 < \frac{2^{2\mathtt{R}}-1}{2\,\mathtt{SNR}} \ . \tag{5.31}$$

As in the case of amplify-and-forward transmission, analytical calculation of the outage probability (5.31) becomes involved; however, we can approximate its high SNR behavior for Rayleigh fading as

$$\begin{aligned} p_P^{\text{out}}(\mathtt{SNR},\mathtt{R}) &\stackrel{\triangle}{=} \Pr[I_P < \mathtt{R}] \\ &\sim \frac{1}{4\alpha(1-\alpha)\sigma^2_{s,d(s)}\sigma^2_{r,d(s)}} \cdot \frac{2^{2\mathtt{R}}[2\mathtt{R}\ln(2)-1]+1}{\mathtt{SNR}^2}, \quad \mathtt{SNR} \text{ large} \ , \end{aligned} \tag{5.32}$$

using the results of Claim 2 in Appendix B, with

$$\begin{gathered} u = \alpha|\mathsf{a}_{s,d(s)}|^2, \quad v = (1-\alpha)|\mathsf{a}_{r,d(s)}|^2 \\ \lambda_u = 1/(\alpha\sigma^2_{s,d(s)}), \quad \lambda_v = 1/((1-\alpha)\sigma^2_{r,d(s)}) \\ \epsilon = [2^{2\mathtt{R}}-1]/(2\,\mathtt{SNR}), \quad t = 2^{2\mathtt{R}}-1 \ . \end{gathered}$$

Clearly (5.32) is minimized for $\alpha = 1/2$, yielding

$$p_P^{\text{out}}(\mathtt{SNR},\mathtt{R}) \sim \frac{1}{\sigma^2_{s,d(s)}\sigma^2_{r,d(s)}} \cdot \frac{2^{2\mathtt{R}}[2\mathtt{R}\ln(2)-1]+1}{\mathtt{SNR}^2}, \quad \mathtt{SNR} \text{ large} \ , \tag{5.33}$$

so that i.i.d.complex Gaussian inputs again minimize outage probability for large $\mathtt{SNR}$. Note that for $\mathtt{R} \to 0$, (5.33) converges to (5.27), the transmit diversity bound without orthogonality constraints. Thus, the orthogonality constraint has little effect for small $\mathtt{R}$, but induces a loss in $\mathtt{SNR}$ loss proportional to

$$\sqrt{\mathtt{R}\ln(2)}$$

with respect to the (unconstrained) transmit diversity bound for large $\mathtt{R}$.

5.3.5 Protocols with Limited Feedback

Outage analysis of protocols with feedback is complicated by, among other things, their variable-rate nature. In addition to outage probability, another relevant quantity in the analysis of protocols with feedback is the *expected* spectral efficiency.

For amplify-and-forward with feedback, the outage probability is given by

$$\begin{aligned} p_{AFF}^{\text{out}}(\mathtt{SNR},\mathtt{R}) =& \Pr\left[|a_{s,d(s)}|^2 \le g(\mathtt{SNR})\right] \\ & \cdot \Pr\left[|a_{s,d(s)}|^2 + \frac{1}{\mathtt{SNR}} f(\mathtt{SNR}|a_{s,r}|^2, \mathtt{SNR}|a_{r,d(s)}|^2) \le g(\mathtt{SNR}) \middle| |a_{s,d(s)}|^2 \le g(\mathtt{SNR})\right] \\ =& \Pr\left[|a_{s,d(s)}|^2 + \frac{1}{\mathtt{SNR}} f(\mathtt{SNR}|a_{s,r}|^2, \mathtt{SNR}|a_{r,d(s)}|^2) \le g(\mathtt{SNR})\right] \ , \end{aligned} \tag{5.34}$$

where $g(\mathtt{SNR}) = [2^{\mathtt{R}}-1]/\mathtt{SNR}$ and where $f(\cdot,\cdot)$ is given in (5.13). The second equality follows from the fact that the intersection of the direct and amplify-and-forward outage events is exactly the amplify-and-forward outage event. Furthermore, the expected spectral efficiency can be computed as

$$\begin{aligned} \overline{\mathtt{R}} &= \mathtt{R} \Pr\left[|a_{s,d(s)}|^2 > \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right] + \frac{\mathtt{R}}{2} \Pr\left[|a_{s,d(s)}|^2 \le \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right] \\ &= \mathtt{R} \exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right) + \frac{\mathtt{R}}{2}\left[1-\exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right)\right] \\ &= \frac{\mathtt{R}}{2}\left[1+\exp\left(-\frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right)\right] \stackrel{\triangle}{=} h_{\mathtt{SNR}}(\mathtt{R}) \ , \end{aligned} \tag{5.35}$$

where the second equality follows from substituting standard exponential results for $|a_{s,d(s)}|^2$.

A fixed value of $\overline{\mathtt{R}}$ can arise from several possible $\mathtt{R}$, depending upon the value of $\mathtt{SNR}$; thus, we see that the pre-image $h_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})$ can contain several points. We define a function $\tilde{h}_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}}) \stackrel{\triangle}{=} \min h_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})$ to capture a useful mapping from $\overline{\mathtt{R}}$ to $\mathtt{R}$; for a given value of $\overline{\mathtt{R}}$, it seems clear from the outage expression (5.34) that we want the smallest $\mathtt{R}$ possible.

For fair comparison to protocols without feedback, we characterize the outage expression $p_{AFF}^{\text{out}}\left(\mathtt{SNR}, \tilde{h}_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})\right)$ in the large $\mathtt{SNR}$ regime, specifically

$$p_{AFF}^{\text{out}}\left(\mathtt{SNR}, \tilde{h}_{\mathtt{SNR}}^{-1}(\overline{\mathtt{R}})\right) \sim \left(\frac{1}{2\sigma_{s,d(s)}^2} \frac{\sigma_{s,r}^2+\sigma_{r,d(s)}^2}{\sigma_{s,r}^2\sigma_{r,d(s)}^2}\right) \cdot \left(\frac{2^{\overline{\mathtt{R}}}-1}{\mathtt{SNR}}\right)^2 \ , \quad \mathtt{SNR} \text{ large} \ , \tag{5.36}$$

where we have combined the results of Claims 1 and 3 in Appendix B.

Bounds for the simple feedback protocol developed in this section can be obtained by suitably normalizing the results developed Section 5.3.4; however, we stress that treating protocols that exploit more general feedback, along with their associated performance limits, is beyond the scope of this chapter.

5.4 Discussion

In this section, we compare the outage results of Section 5.3. We begin with some observations for statistically asymmetric networks, and then specialize the results to the case of statistically symmetric networks, *e.g.*, $\sigma_{i,j}^2 = 1$, without loss of generality.

5.4.1 Asymmetric Networks

As the results in Section 5.3 indicate, for fixed rates, simple protocols such as fixed amplify-and-forward, adaptive decode-and-forward, and amplify-and-forward with feedback each achieve full (*i.e.*, second-order) diversity: their outage probability performance decays proportional to $1/\mathtt{SNR}^2$ (*cf.* (5.15), (5.23), and (5.36)). We now compare these protocols to the transmit diversity bound, discuss the impacts of spectral efficiency and network geometry on performance, and examine their outage events.

Comparison to Transmit Diversity Bound

In the low spectral efficiency regime, the protocols without feedback are within a factor of

$$\left[\frac{2^{2\mathtt{R}}-1}{2(2^{\mathtt{R}}-1)}\right]\sqrt{1+\left(\frac{\sigma_{r,d(s)}^2}{\sigma_{s,r}^2}\right)} \approx \sqrt{1+\left(\frac{\sigma_{r,d(s)}^2}{\sigma_{s,r}^2}\right)}$$

in SNR from the optimum transmit diversity bound, suggesting that the powerful benefits of multi-antenna systems can indeed be obtained without the need for physical arrays. For statistically symmetric networks, *e.g.*, $\sigma_{i,j}^2 = 1$, the loss is only $\sqrt{2}$ or 1.5 dB; more generally the loss decreases as the path between the source and relay improves relative to the link between the relay and destination.

For larger spectral efficiencies, the fixed and adaptive protocols lose an additional 3 dB per transmitted bit/s/Hz with respect to the transmit diversity bound. This additional loss is due to two factors: the orthogonality constraint imposed at the relay, and the repetition-coded nature of the protocols. As Fig. 5-3 suggests, of the two, repetition-coding appears to be the more significant source of inefficiency in our protocols. In Fig. 5-3, the SNR loss of orthogonal transmit diversity with respect to (unconstrained) transmit diversity is intended to indicate the cost of imposing orthogonality at the relay, while the loss of our cooperative diversity protocols with respect to the transmit diversity bound indicates the

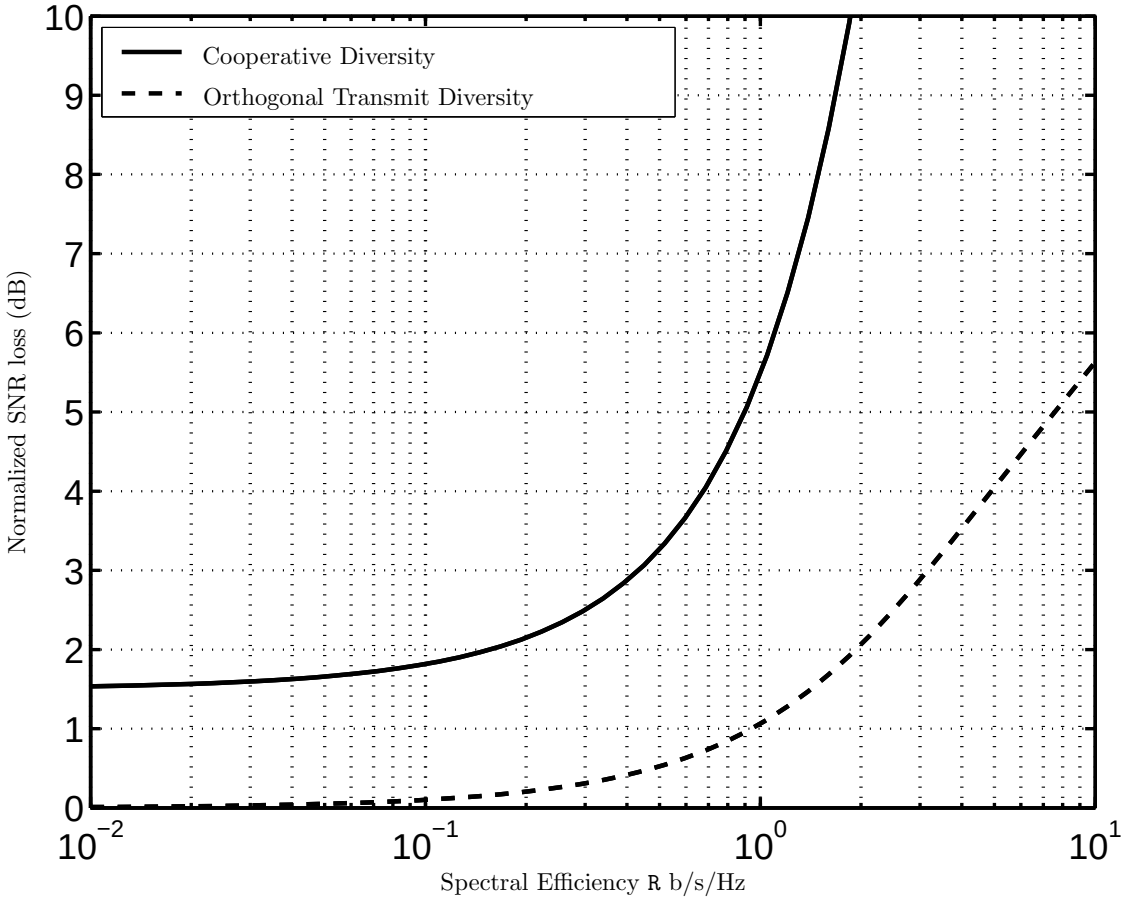


Figure 5-3: SNR loss for cooperative diversity protocols (solid) and orthogonal transmit diversity bounds (dashed) relative to the (unconstrained) transmit diversity bound (0 dB).

cost of both imposing orthogonality at the relay and employing repetition-like codes. The figure suggests that, although the orthogonality constraint contributes, "repetition" in the form of amplifying or repetition codes is the major cause of SNR loss for high rates. By contrast, the amplify-and-forward protocol with feedback overcomes these additional losses by repeating only when necessary.

Outage Events

It is interesting that the amplify-and-forward and adaptive decode-and-forward have the same high SNR performance, especially considering the different shapes of their outage events (*cf.* (5.14), (5.21)), which are shown in the low spectral-efficiency regime in Fig. 5-4. When the relay can fully decode the source message, *i.e.*, $\mathtt{SNR}_{\text{norm}}|a_{s,r}|^2 \geq 2$, and repeat it, the outage event for adaptive decode-and-forward is a strict subset of the outage event of amplify-and-forward, with amplify-and-forward approaching that of decode-and-forward as $|a_{s,r}|^2 \to \infty$. On the other hand, when the relay cannot fully decode the source message, *i.e.*, $\mathtt{SNR}_{\text{norm}}|a_{s,r}|^2 < 2$, and the source repeats, the outage event of amplify-and-forward is neither a subset nor a superset of the outage event for adaptive decode-and-forward. Apparently averaging over the Rayleigh fading coefficients eliminates the differences between amplify-and-forward and adaptive decode-and-forward, at least in the high SNR regime.

Effects of Geometry

To study the effect of network geometry on performance, we compare the high SNR behavior of direct transmission with that of amplify-and-forward with feedback. Using a common model for the path-loss (fading variances), we set $\sigma_{i,j}^2 \propto d_{i,j}^{-\alpha}$, where $d_{i,j}$ is the distance between terminals i and j, and α is the path-loss exponent [61]. Under this model, comparing (5.11) with (5.36), assuming both approximations are good for the $\mathtt{SNR}$ of interest, we prefer amplify-and-forward with feedback whenever

$$\left(\frac{d_{s,r}}{d_{s,d(s)}}\right)^{\alpha} + \left(\frac{d_{r,d(s)}}{d_{s,d(s)}}\right)^{\alpha} < 2\,\mathtt{SNR}_{\text{norm}}\ . \tag{5.37}$$

Thus, amplify-and-forward with feedback is useful whenever the relay lies within a certain normalized ellipse having the source and destination as its foci, with the size of the ellipse increasing in $\mathtt{SNR}_{\text{norm}}$. What is most interesting about the structure of this "utilization

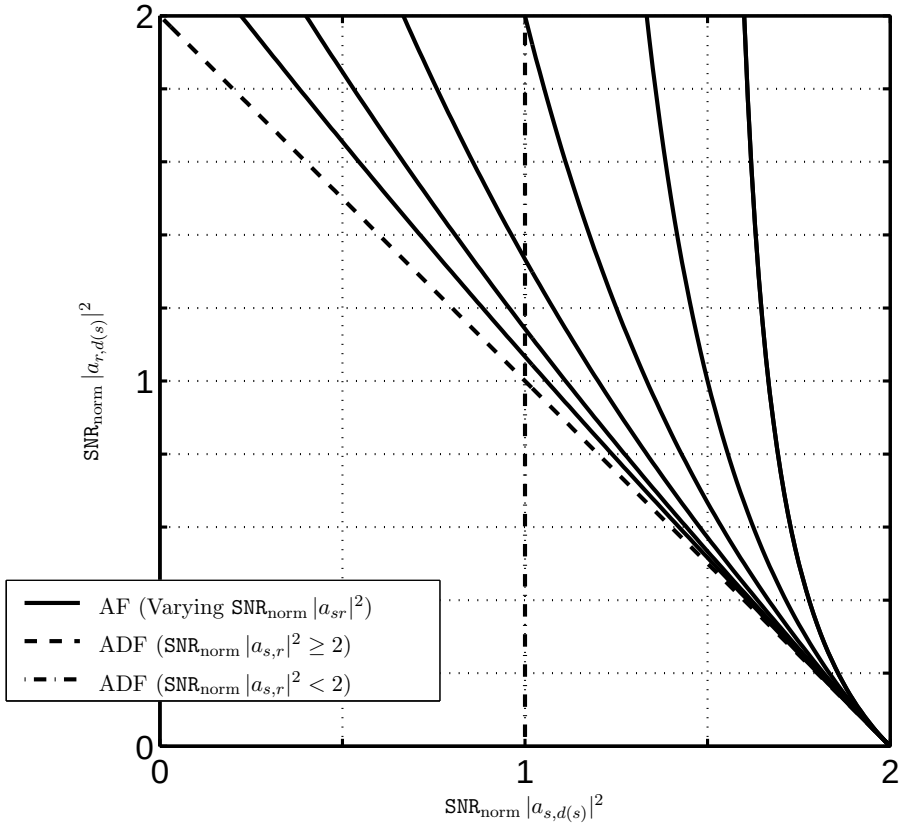


Figure 5-4: Outage event boundaries for amplify-and-forward (solid) and adaptive decode-and-forward (dashed and dash-dotted) transmission as functions of the realized fading coefficient $|a_{s,r}|^2$ between the cooperating terminals. Outage events are to the left and below the respective outage event boundaries. Successively lower solid curves correspond to amplify-and-forward with increasing values of $|a_{s,r}|^2$. The dashed curve corresponds to the outage event for adaptive decode-and-forward when the relay can fully decode, *i.e.*, $\mathtt{SNR}_{\mathrm{norm}}|a_{s,r}|^2 \geq 2$, and the relay repeats, while the dash-dotted curve corresponds to the outage event of adaptive decode-and-forward when the relay cannot fully decode, *i.e.*, $\mathtt{SNR}_{\mathrm{norm}}|a_{s,r}|^2 < 2$, and the source repeats. Note that the dash-dotted curve also corresponds to the outage event for direct transmission.

region" for amplify-and-forward with feedback is that it is *symmetric* with respect to the source and destination. By comparison, the utilization region for fixed decode-and-forward is given by a certain circle about only the source.

Utilization regions of the form (5.37) may be useful in terms of developing higher layer network protocols that select between direct transmission and cooperative diversity transmission using one of a number of potential relays. Such algorithms and their performance represent an interesting area of further research, and a key ingredient for fully incorporating cooperative diversity into wireless networks.

5.4.2 Symmetric Networks

We now specialize all of our results to the case of statistically symmetric networks, *e.g.*, $\sigma_{i,j}^2 = 1$ without loss of generality.

Results under Different Parameterizations

In the sections to follow, we compare performance of the various protocols under the two parameterizations described in Section 5.1.3, namely, $(\mathtt{SNR}_{\rm norm}, \mathtt{R})$ and $(\mathtt{SNR}, \mathtt{R}_{\rm norm})$, respectively. Parameterizing the outage results from Section 5.3 in terms of $(\mathtt{SNR}_{\rm norm}, \mathtt{R})$ is straightforward because $\mathtt{R}$ remains fixed; we simply substitute $\mathtt{SNR} = \mathtt{SNR}_{\rm norm}(2^{\mathtt{R}} - 1)$ to obtain the results listed in the second column of Table 5.1. Parameterizing the outage results from Section 5.3 in terms of $(\mathtt{SNR}, \mathtt{R}_{\rm norm})$ is a bit more involved because $\mathtt{R} = \mathtt{R}_{\rm norm} \log(1 + \mathtt{SNR})$ increases with increasing $\mathtt{SNR}$.

The results in Appendix B are all general enough to allow this latter parameterization. To demonstrate their application, we consider amplify-and-forward transmission. The outage event for amplify-and-forward under this alternative parameterization is given by

$$|a_{s,d(s)}|^2 + \frac{1}{\mathtt{SNR}} f\left(\mathtt{SNR}\,|a_{s,r}|^2,\ \mathtt{SNR}\,|a_{r,d(s)}|^2\right) < \frac{2^{2\mathtt{R}} - 1}{\mathtt{SNR}} = \frac{(1 + \mathtt{SNR})^{2\mathtt{R}_{\rm norm}} - 1}{\mathtt{SNR}} \ .$$

For $\mathtt{R}_{\rm norm} < 1/2$, the outage probability is approximately

$$p_{AF}^{\rm out}\left(\mathtt{SNR}, \mathtt{R}_{\rm norm}\right) \sim \left[\frac{\mathtt{SNR}}{(1 + \mathtt{SNR})^{2\mathtt{R}_{\rm norm}} - 1}\right]^{-2} \sim 1/\mathtt{SNR}^{2(1-2\mathtt{R}_{\rm norm})} \ , \quad \mathtt{SNR} \text{ large} \ ,$$

Protocol	$p^{\text{out}}\left(\mathtt{SNR}_{\text{norm}}, \mathtt{R}\right)$, high $\mathtt{SNR}_{\text{norm}}$	$p^{\text{out}}\left(\mathtt{SNR}, \mathtt{R}_{\text{norm}}\right)$, high $\mathtt{SNR}$
Direct	$1/\mathtt{SNR}_{\text{norm}}$	$1/\mathtt{SNR}^{(1-\mathtt{R}_{\text{norm}})}$
Amplify-and-Forward	$(2^{\mathtt{R}}+1)^2/\mathtt{SNR}_{\text{norm}}^2$	$1/\mathtt{SNR}^{2(1-2\mathtt{R}_{\text{norm}})}$
Decode-and-Forward	$(2^{\mathtt{R}}+1)/\mathtt{SNR}_{\text{norm}}$	$1/\mathtt{SNR}^{(1-2\mathtt{R}_{\text{norm}})}$
Adaptive Decode-and-Forward	$(2^{\mathtt{R}}+1)^2/\mathtt{SNR}_{\text{norm}}^2$	$1/\mathtt{SNR}^{2(1-2\mathtt{R}_{\text{norm}})}$
Amplify-and-Forward with Feedback	$1/\mathtt{SNR}_{\text{norm}}^2$	$1/\mathtt{SNR}^{2(1-\mathtt{R}_{\text{norm}})}$
Transmit Diversity Bound	$2/\mathtt{SNR}_{\text{norm}}^2$	$2/\mathtt{SNR}^{2(1-\mathtt{R}_{\text{norm}})}$
Orthogonal Transmit Diversity Bound	$\left(\frac{2^{2\mathtt{R}}[2\mathtt{R}\ln(2)-1]+1}{(2^{\mathtt{R}}-1)^2}\right)/\mathtt{SNR}_{\text{norm}}^2$	$2[\mathtt{R}_{\text{norm}}\ln(\mathtt{SNR})+1]/\mathtt{SNR}^{2(1-\mathtt{R}_{\text{norm}})}$

Table 5.1: Summary of outage probability approximations. To capture the salient tradeoffs between signal-to-noise ratio $\mathtt{SNR}$, spectral efficiency $\mathtt{R}$ b/s/Hz, and diversity gain of the various protocols, the results are specialized to the case of statistically symmetric networks with fading variances $\sigma_{i,j}^2 = 1$.

where we have utilized the results of Claim 1 in Appendix B with

$$u = |a_{s,d(s)}|^2, \quad v = |a_{s,r}|^2, \quad w = |a_{r,d(s)}|^2, \quad \lambda_u = \lambda_v = \lambda_w = 1$$
$$g(\epsilon) = \epsilon\left[(1+1/\epsilon)^{2\mathtt{R}_{\text{norm}}} - 1\right], \quad t = \mathtt{SNR}, \quad h(t) = 1/t\ .$$

Table 5.1 summarizes the results, which can be obtained in similar fashion using appropriate results from Appendix B.

Fixed $\mathtt{R}$ Systems

Fig. 5-5 shows outage probabilities for the various protocols as functions of $\mathtt{SNR}_{\text{norm}}$ in the small, fixed $\mathtt{R}$ regime. The approximations in the second column Table 5.1 are accurate for moderate to large values of the $\mathtt{SNR}_{\text{norm}}$. The diversity gains of our cooperative diversity protocols appear as a steeper slopes in Fig. 5-5, from a factor of 10 decrease in outage probability for each additional 10 dB of SNR in the case of direct transmission, to a factor of 100 decrease in outage probability for each additional 10 dB of SNR in the case of cooperative diversity transmission. The relative loss of 1.5 dB for our fixed and adaptive cooperative diversity protocols with respective to the transmit diversity bound is also apparent. These curves shift to the right by 3 dB for each additional bit/s/Hz of spectral efficiency in the high $\mathtt{R}$ regime. By contrast, the performance of amplify-and-forward with feedback is unchanged at high SNR for increasing $\mathtt{R}$. Note that, at outage probabilities on the order of 10^{-3}, our cooperative diversity protocols achieve large energy savings over direct transmission—on the order of 12–15 dB.

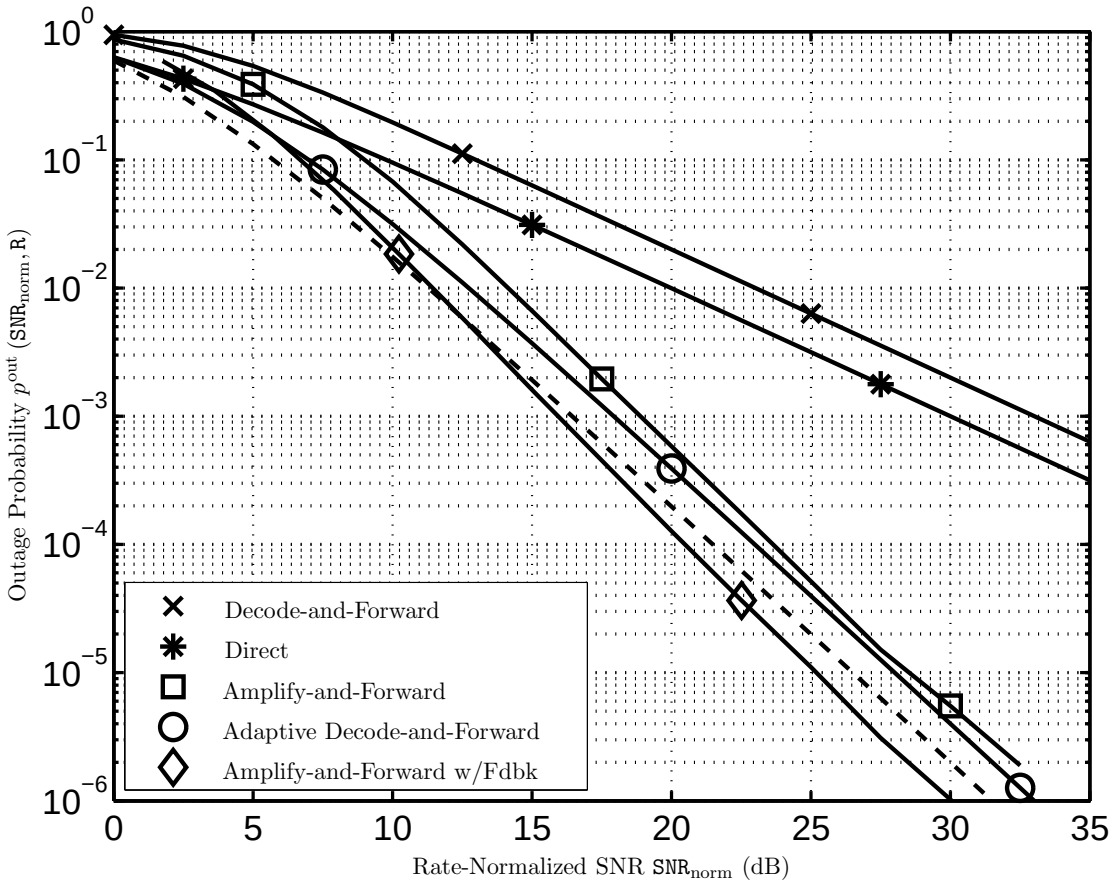


Figure 5-5: Outage probabilities vs. $\texttt{SNR}_{\text{norm}}$, small $\texttt{R}$ regime, for statistically symmetric networks, *i.e.*, $\sigma_{i,j}^2 = 1$. The outage probability curve for amplify-and-forward transmission was obtained via Monte-Carlo simulation, while the other curves are computed from analytical expressions. The dashed curve corresponds to the transmit diversity bounds in this low spectral efficiency regime.

Fixed $\mathtt{R}_{\text{norm}}$ Families of Systems

Another way to examine the high spectral efficiency regime as SNR becomes large is to allow $\mathtt{R}$ to grow with increasing $\mathtt{SNR}$. In particular, the choice of $\mathtt{R} = \mathtt{R}_{\text{norm}} \log(1+\mathtt{SNR})$, with $\mathtt{R}_{\text{norm}}$ fixed, is a natural one: for slower growth, the outage results essentially behave like fixed $\mathtt{R}$ systems for sufficiently large $\mathtt{SNR}$, while for faster growth, the outage probabilities all tend to 1. These observations motivate our parameterization in terms of $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$ in the third column of Table 5.1. As we stressed in Section 5.1.3, while not as familiar to communication engineers, this parameterization can be used to compared systems in different families for a given $\mathtt{SNR}$ and $\mathtt{R}_{\text{norm}}$.

Parameterizing performance in terms of $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$ leads to interesting tradeoffs between the diversity order and normalized spectral efficiency of a protocol. Because these tradeoffs arise naturally in the context of multi-antenna systems [93], it is not surprising that they show up in the context of cooperative diversity. Diversity order can be viewed as the power to which $\mathtt{SNR}^{-1}$ is raised in our outage expressions in the third column of Table 5.1. To be precise, we can define diversity order as

$$\Delta(\mathtt{R}_{\text{norm}}) \stackrel{\triangle}{=} \lim_{\mathtt{SNR}\to\infty} \frac{-\log p^{\text{out}}\left(\mathtt{SNR}, \mathtt{R}_{\text{norm}}\right)}{\log \mathtt{SNR}} \ . \tag{5.38}$$

Larger $\Delta(\mathtt{R}_{\text{norm}})$ implies more robustness to fading (faster decay in the outage probability with increasing SNR), but $\Delta(\mathtt{R}_{\text{norm}})$ generally decreases with increasing $\mathtt{R}_{\text{norm}}$. For example, the diversity order of amplify-and-forward transmission and adaptive decode-and-forward transmission is

$$\Delta_{AF}(\mathtt{R}_{\text{norm}}) = \Delta_{ADF}(\mathtt{R}_{\text{norm}}) = 2(1 - 2\mathtt{R}_{\text{norm}}) \ ; \tag{5.39}$$

thus, their maximum diversity order $\Delta \to 2$ is achieved as $\mathtt{R}_{\text{norm}} \to 0$, and minimal diversity order $\Delta \to 0$ results as $\mathtt{R}_{\text{norm}} \to 1/2$. Fig. 5-6 compares the tradeoffs for direct transmission and cooperative diversity transmission. As we might expect from our previous discussion, amplify-and-forward with feedback yields the highest $\Delta(\mathtt{R}_{\text{norm}})$ for each $\mathtt{R}_{\text{norm}}$; this curve also corresponds to the transmit diversity bound in the high SNR regime. What is most interesting about the results in Fig. 5-6 is the sharp transition at $\mathtt{R}_{\text{norm}} = 1/3$ between our preference for amplify-and-forward transmission (as well as adaptive decode-and-forward) for $\mathtt{R}_{\text{norm}} < 1/3$ and our preference for direct transmission for $\mathtt{R}_{\text{norm}} > 1/3$.

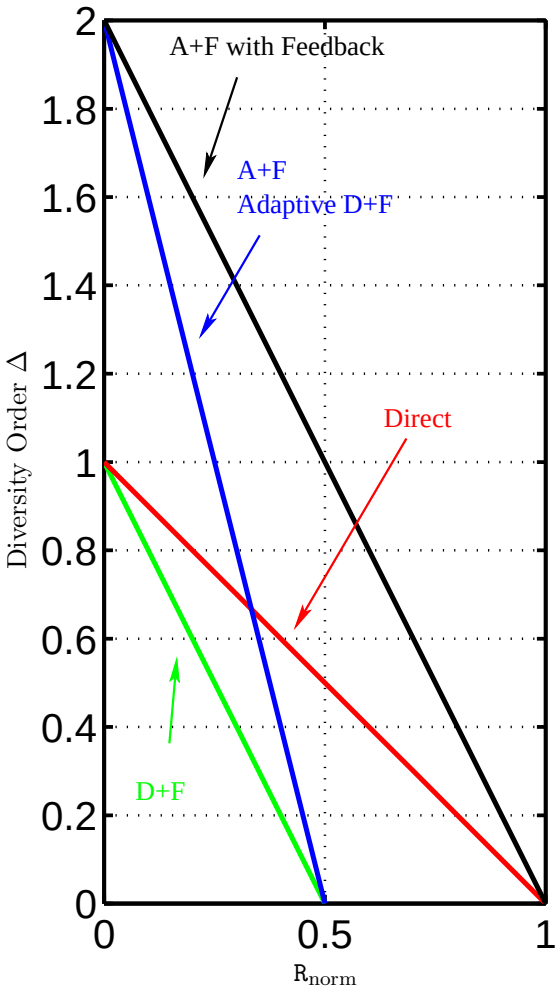


Figure 5-6: Diversity order Δ vs. R_{norm} for direct transmission and cooperative diversity protocols.

Chapter 6

Cooperative Diversity in Networks

This chapter builds upon the results of Chapter 5 and explores several issues related to integrating cooperative diversity into networks with more than two terminals. We begin by considering fully cooperative networks in which all the terminals in the network serve as relays for each other. We generalize the repetition-based algorithms of Chapter 5, and then improve upon their bandwidth efficiency by exploring space-time coded cooperative diversity algorithms. Both classes of algorithms offer full spatial diversity in the number of cooperating terminals, but these diversity gains may not outweigh system losses due to bandwidth inefficiency and path-loss in practical operating regimes. In later sections, we also consider partially cooperative networks in which only certain subsets of terminals serve as relays for each other. We describe various grouping algorithms and relate their importance and implementation to architectural aspects of the networks. We illustrate example scenarios of current interest in infrastructure networks and ad-hoc networks with clusters. We also briefly comment on layering cooperative diversity within and across traditional network protocol layers (*cf.* Fig. 1-2). As we will see, because cooperative diversity is inherently a network approach, it meshes with functionality typically implemented with the network, link, medium-access control, and physical layers

6.1 Fully Cooperative Networks

Chapter 5 developed various cooperative diversity algorithms for a pair of terminals. In this section, we show that these algorithms readily extend to combat multipath fading in larger networks. Full spatial diversity benefits of these *repetition-based cooperative diversity*

algorithms, as we refer to them throughout this section, come at a price of decreasing bandwidth efficiency with the number of cooperating terminals, because each relay requires its own subchannel for repetition. As in Chapter 5, limited feedback from the destination terminal provides one means of overcoming such bandwidth inefficiencies, but we do not repeat the analysis here. Instead, this section develops an alternative approach to improving bandwidth efficiency of the algorithms based upon space-time codes that allow all relays to transmit on the same subchannel. Requiring more computational complexity in the terminals, we will see these *space-time coded cooperative diversity algorithms* also offer full spatial diversity benefits and are more amenable to distributed implementation, without requiring feedback.

We consider a wireless network with a set of transmitting terminals denoted $\mathcal{M} = \{1, 2, \ldots, m\}$. Each transmitting source terminal $s \in \mathcal{M}$ has information to transmit to a single destination terminal, denoted $d(s) \notin \mathcal{M}$, potentially using terminals $\mathcal{M} - \{s\}$ as relays. Thus there are m cooperating terminals communicating to $d(s)$. For algorithms in which we require the relays to fully decode the source message, we define the *decoding set* $\mathcal{D}(s)$ to be the set of relays that can decode the message of source s. In the case of amplify-and-forward cooperative diversity, we take $\mathcal{D}(s) = \mathcal{M} - \{s\}$.

Both classes of algorithms consist of two transmission phases, as in Chapter 5. Fig. 6-1 illustrates these two phases, and allows us to point out the similarities and differences between the algorithms. In the first phase, the source broadcasts to its destination and all potential relays. During the second phase of the algorithms, the other terminals relay to the destination, either on orthogonal subchannels in the case of repetition-based cooperative diversity, or simultaneously on the same subchannel in the case of space-time coded cooperative diversity.

To summarize our results, we show the outage probability performance of repetition-based cooperative diversity decays asymptotically in $\mathtt{SNR}$ proportional to $1/\mathtt{SNR}^{m(1-m\mathtt{R}_{\mathrm{norm}})}$, where $\mathtt{SNR}$ corresponds to the average signal-to-noise ratio (SNR) between terminals, and $0 < \mathtt{R}_{\mathrm{norm}} < 1/m$ corresponds to a suitably-normalized spectral efficiency of the protocol. In this context, full diversity refers to the fact that, as $\mathtt{R}_{\mathrm{norm}} \to 0$, the outage probability decays proportional to $1/\mathtt{SNR}^m$. By contrast, the outage probability performance of non-cooperative transmission decays asymptotically as $1/\mathtt{SNR}^{(1-\mathtt{R}_{\mathrm{norm}})}$, where $0 < \mathtt{R}_{\mathrm{norm}} < 1$ is allowed, and as $1/\mathtt{SNR}$ as $\mathtt{R}_{\mathrm{norm}} \to 0$. Thus, while the outage probability performance

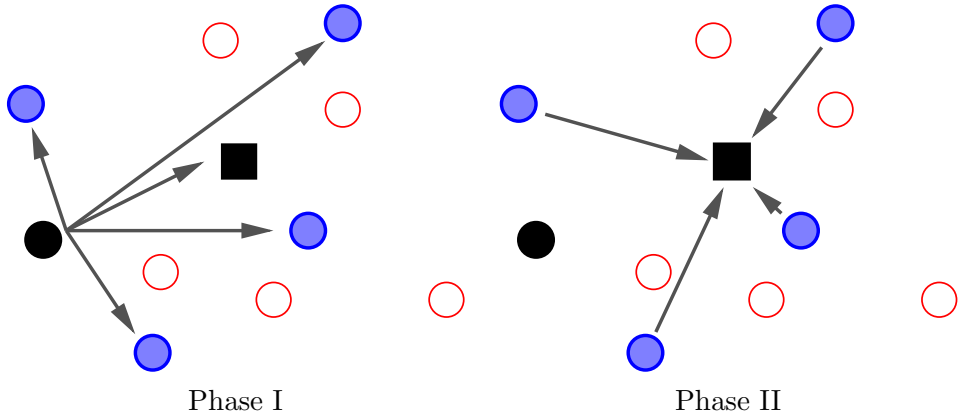


Figure 6-1: Illustration of the two-phases of repetition-based and space-time coded cooperative diversity algorithms. In the first phase, the source broadcasts to the destination as well as potential relays. Decoding relays are shaded. In the second phase, the decoding relays either repeat on orthogonal subchannels or utilize a space-time code to simultaneously transmit to the destination.

of cooperative diversity can decay faster, it does so only for small $\mathtt{R}_{\text{norm}}$, in particular, for $\mathtt{R}_{\text{norm}} < 1/(m+1)$. For $\mathtt{R}_{\text{norm}} > 1/(m+1)$, the inherent bandwidth inefficiency of repetition-based cooperative diversity outweighs the benefits of diversity gain, so that non-cooperative transmission is preferable in this regime. Of course, there are more general forms of decode-and-forward transmission than repetition, just as there are more general forms of space-time codes. Space-time coded cooperative transmission leads to schemes whose outage probability performance decays asymptotically as $1/\mathtt{SNR}^{m(1-2\mathtt{R}_{\text{norm}})}$. Thus, they (a) achieve full spatial diversity order m as $\mathtt{R}_{\text{norm}} \to 0$, (b) have larger diversity order than repetition-based algorithms for all $\mathtt{R}_{\text{norm}}$, and (c) are preferable to non-cooperative transmission if $\mathtt{R}_{\text{norm}} < (m-1)/(2m-1)$. Moreover, we will see that these protocols may be readily implemented in a distributed fashion, because they only require the relays to estimate the SNR of their received signals, decode them if the SNR is sufficiently high, re-encode with the appropriate waveform from a space-time code, and re-transmit in the same subchannel.

In broader context, both classes of cooperative diversity can be viewed as a form of *network coding*, in this case designed to exploit spatial diversity in the network. There is a growing body of work focused on network coding for enhancing performance of wireless and other communication systems and networks [33, 29, 47]. In the area of wireless ad-hoc

networks in particular, many authors have attempted to determine the capacity region, *i.e.*, the set of all reliably achievable rates in the network. A nice summary of this literature can be found in [29].

6.1.1 System Model

This section highlights the system model that we use to develop extensions of the repetition-based algorithms in Chapter 5 as well as the space-time coded cooperative diversity algorithms. Differences between the model employed here and the one employed in Chapter 5 include a larger number of terminals and different medium-access control protocols for repetition-based and space-time coded cooperative diversity. As a result, we repeat only the more fundamental elements of the system model in this section.

As throughout the dissertation, narrowband transmissions suffer the effects of frequency nonselective Rayleigh fading and additive white Gaussian noise. We consider the scenario in which the receivers can accurately measure the realized fading coefficients in their received signals, but the transmitters either do not possess or do not exploit knowledge of the realized fading coefficients. As in Chapter 5, we focus on the case of slow fading and measure performance by outage probability to isolate the benefits of space diversity. As before, we utilize a baseband-equivalent, discrete-time channel model for the continuous-time channel.

Medium-Access Control

For medium-access control, terminals transmit on essentially orthogonal channels as in many current wireless networks. As a baseline for comparison, Fig. 6-2 illustrates example channel allocations for non-cooperative transmission, in which each transmitting terminal utilizes a fraction $1/m$ of the total degrees of freedom in the channel.

For cooperative diversity transmission, the medium-access control protocol also manages orthogonal relaying to ensure that terminals satisfy the half-duplex constraint and do no transmit and receive simultaneously on the same subchannel. Note that these are the same basic restrictions on medium-access control protocols described in Chapter 5. We now describe how the medium-access control protocol differs under repetition-based and space-time coded cooperative diversity.

Fig. 6-3 illustrates example channel and subchannel allocations for repetition-based cooperative diversity, in which relays either amplify what they receive or fully decode and

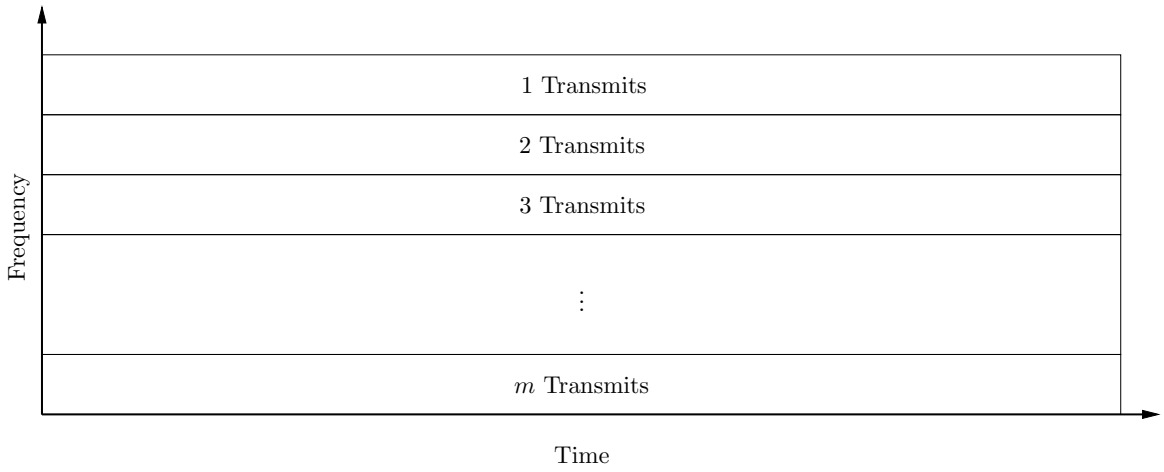


Figure 6-2: Non-cooperative medium-access control. Example source allocations among m transmitting terminals across orthogonal frequency channels.

repeat the source signal, as in Chapter 5. In order for the destination to combine these signals and achieve diversity gains, the repetitions must occur on essentially orthogonal subchannels. For simplicity, Fig. 6-3 shows channel allocations for different source terminals across frequency, and subchannel allocations for different relays across time. More generally, for a given source s and destination $d(s)$, the relays $\mathcal{M} - \{s\}$ can repeat in any pre-determined order. Arbitrary permutations of these allocations in time and frequency do not alter the conclusions to follow, as a long as causality is preserved and each of the subchannels contains a fraction $1/m^2$ of the total degrees of freedom in the channel. As in non-cooperative transmission, transmission between source s and destination $d(s)$ utilizes a fraction $1/m$ of the total degrees of freedom in the channel. Similarly, each cooperating terminal transmits in a fraction $1/m$ of the total degrees of freedom.

Fig. 6-4 illustrates example channel and subchannel allocations for space-time coded cooperative diversity, in which relays utilize a suitable space-time code in the second phase and therefore can transmit simultaneously on the same subchannel. Again, transmission between source s and destination $d(s)$ utilizes $1/m$ of the total degrees of freedom in the channel. However, in contrast to non-cooperative transmission and repetition-based cooperative diversity transmission, each terminal employing space-time coded cooperative diversity transmits in $1/2$ the total degrees of freedom in the channel. It is important to keep track of these fractions when normalizing power and bandwidth in the sequel.

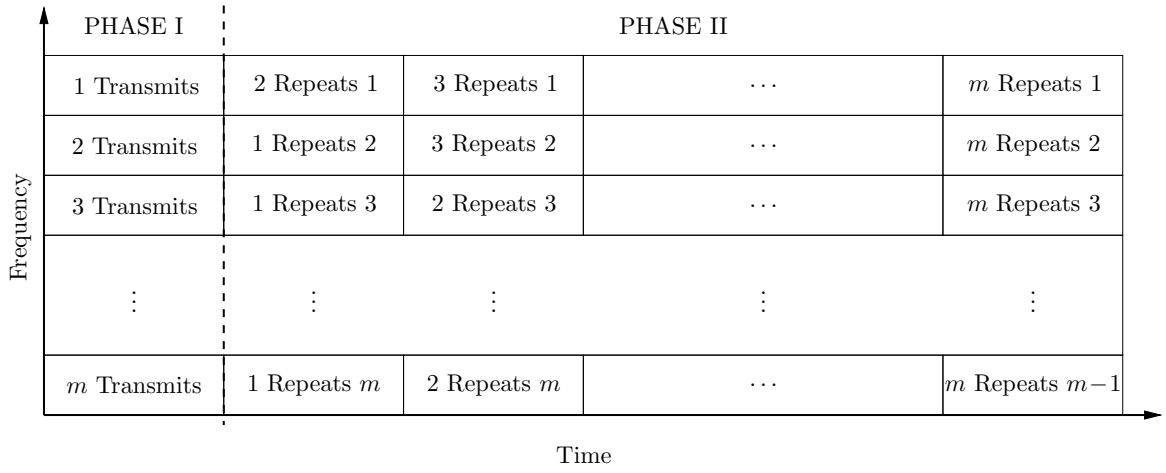


Figure 6-3: Repetition-based medium-access control. Example source channel allocations across frequency and relay subchannel allocations across time for repetition-based cooperative diversity among m terminals.

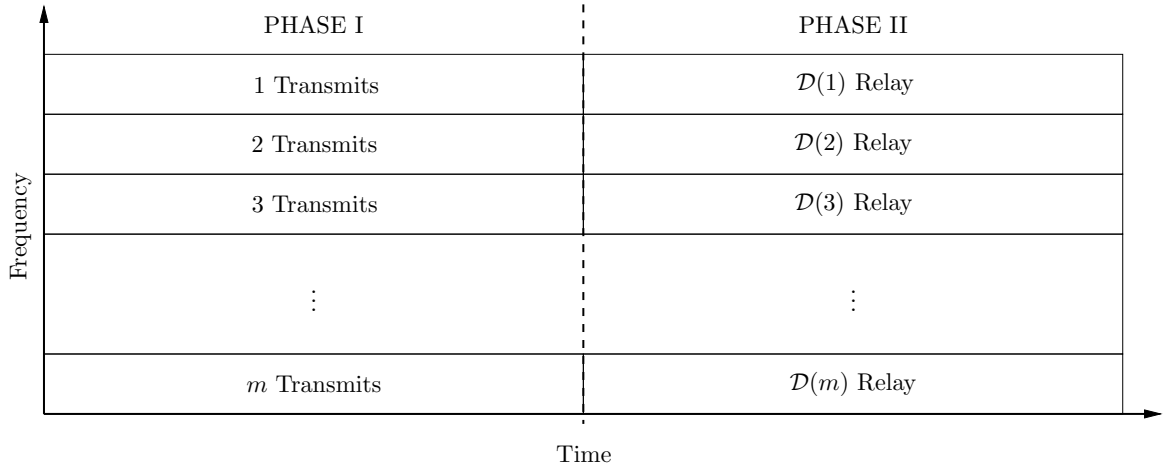


Figure 6-4: Space-time coded medium-access control. Example channel allocations across frequency and time for m transmitting terminals. For source s, $\mathcal{D}(s)$ denotes the set of decoding relays participating in a space-time code during the second phase.

Equivalent Channel Models

Under the above orthogonality constraints, we can now conveniently, and without loss of generality, characterize our channel models. Due to symmetry of the channel allocations, we focus on transmission of a message from source s to its destination $d(s)$ using terminals $\mathcal{M} - \{s\}$ as relays.

During the first phase, each potential relay $r \in \mathcal{M} - \{s\}$ receives

$$y_r[n] = a_{s,r}\, x_s[n] + z_r[n] \ , \tag{6.1}$$

in the appropriate subchannel, where again $x_s[n]$ is the source transmitted signal and $y_r[n]$ is the received signal at r. For decode-and-forward transmission, if the SNR is sufficiently large for r to decode the source transmission, then r serves as a decoding relay for the source s, so that $r \in \mathcal{D}(s)$. Again, for amplify-and-forward transmission, we can think of $\mathcal{D}(s)$ as being the entire set of relays for source s, *i.e.*, $\mathcal{D}(s) = \mathcal{M} - \{s\}$.

The destination receives signals during both phases. During the first phase, we model the received signal at $d(s)$ as

$$y_{d(s)}[n] = a_{s,d(s)}\, x_s[n] + z_{d(s)}[n] \ , \tag{6.2}$$

in the appropriate subchannel. During the second phase, the equivalent channel models are different for repetition-based and space-time coded cooperative diversity. For repetition-based cooperative diversity, the destination receives separate re-transmissions from each of the relays, *i.e.*, for $r \in \mathcal{M} - \{s\}$, we model the received signal at $d(s)$ as

$$y_{d(s)}[n] = a_{r,d(s)}\, x_r[n] + z_{d(s)}[n] \ , \tag{6.3}$$

in the appropriate subchannel, where $x_r[n]$ is the transmitted signal of relay r. For space-time coded cooperative diversity, all of the relay transmissions occur in the same subchannel and superimpose at the destination, so that

$$y_{d(s)}[n] = \sum_{r \in \mathcal{D}(s)} a_{r,d(s)}\, x_r[n] + z_{d(s)} \ , \tag{6.4}$$

in the appropriate subchannel.

As before, in (6.1)–(6.4), $\mathsf{a}_{i,j}$ captures the effects of path-loss, shadowing, and frequency nonselective fading, and $\mathsf{z}_j[n]$ captures the effects of receiver noise and other forms of interference in the system. Note that all the fading coefficients are constant over the example time and frequency axes shown in Figures 6-2–6-4. We again focus on the scenario in which the fading coefficients are known to, *i.e.*, accurately measured by, the appropriate receivers, but not fully known to, or not exploited by, the transmitters. Statistically, we again model $\mathsf{a}_{i,j}$ as zero-mean, independent, circularly-symmetric complex Gaussian random variables with variances $1/\lambda_{i,j}$, so that the magnitudes $|\mathsf{a}_{i,j}|$ are Rayleigh distributed ($|\mathsf{a}_{i,j}|^2$ are exponentially distributed with parameter $\lambda_{i,j}$) and the phases $\angle\mathsf{a}_{i,j}$ are uniformly distributed on $[0, 2\pi)$. Furthermore, we model $\mathsf{z}_j[n]$ as zero-mean mutually independent, circularly-symmetric, complex Gaussian random sequences with variance N_0.

Parameterizations

As we saw in Chapter 5, two important parameters of the system are the transmit signal-to-noise ratio $\mathtt{SNR}$ and the spectral efficiency $\mathtt{R}$. Again, it is natural to define these parameters in terms of standard parameters in the continuous-time channel with non-cooperative transmission (*cf.* Fig. 6-2) as a baseline.

For a continuous-time channel with total bandwidth W Hz available for transmission, the discrete-time model contains W two-dimensional symbols per second (2D/s). If the transmitting terminals have an average power constraint in the continuous-time channel model of P_c Joules/s, we see that this translates into a discrete-time power constraint of $P = mP_c/W$ Joules/2D, since each terminal transmits in a fraction $1/m$ of the available degrees of freedom for non-cooperative transmission (*cf.* Fig. 6-2) and repetition-based cooperative diversity (*cf.* Fig. 6-3). Thus, the channel model is parameterized by the SNR random variables $\mathtt{SNR}\,|\mathsf{a}_{i,j}|^2$, where

$$\mathtt{SNR} \triangleq \frac{mP_c}{N_0 W} = \frac{P}{N_0} \tag{6.5}$$

is the SNR without fading. For space-time coded cooperative diversity (*cf.* Fig. 6-4), the terminals transmit in half the available degrees of freedom, so the discrete-time power constraint becomes $2P/m$.

In addition to SNR, transmission schemes are further parameterized by the spectral

efficiency $\mathtt{R}$ b/s/Hz attempted by the transmitting terminals. Note that throughout this section $\mathtt{R}$ is the transmission rate normalized by the number of degrees of freedom utilized by each terminal under non-cooperative transmission, not by the total number of degrees of freedom in the channel.

As in Chapter 5, our results can be parameterized by either of the pairs $(\mathtt{SNR}_{\text{norm}}, \mathtt{R})$ and $(\mathtt{SNR}, \mathtt{R}_{\text{norm}})$, where $\mathtt{SNR}_{\text{norm}}$ and $\mathtt{R}_{\text{norm}}$ are normalized SNR and spectral efficiency, respectively.

6.1.2 Repetition-Based Cooperative Diversity

We now analyze performance of a repetition decode-and-forward cooperative diversity algorithm for more than two terminals. Such protocols consist of the source broadcasting its transmission to its destination and potential relays. Potential relays that can decode the transmission become decoding relays and participate in the second phase of the protocol by repeating the source message on orthogonal subchannels. Although the set of decoding relays $\mathcal{D}(s)$ is a random set, we will see that protocols of this form offer full spatial diversity in the number of cooperating terminals, not just the number of decoding relays participating in the second phase. Interestingly, potential relays that cannot decode contribute as much to the performance of the protocol as the decoding relays, just as in the adaptive decode-and-forward algorithm developed for two terminals in Chapter 5. We note that similar high SNR results should be obtainable for amplify-and-forward transmission using suitable extensions of the appropriate results in Appendix B.1.

Mutual Information and Outage Probability

Since the channel average mutual information I_{rep} is a function of, *e.g.*, the coding scheme, the rule for including potential relays into the decoding set $\mathcal{D}(s)$, and the fading coefficients of the channel, it too is a random variable. As in Chapter 5, the event $I_{\text{rep}} < \mathtt{R}$ that this mutual information random variable falls below some fixed spectral efficiency $\mathtt{R}$ is referred to as an outage event, and the probability of an outage event, $\Pr[I_{\text{rep}} < \mathtt{R}]$, is referred to as the outage probability of the channel.

Since $\mathcal{D}(s)$ is a random set, we utilize the total probability law and write

$$\Pr\left[I_{\text{rep}} < \mathtt{R}\right] = \sum_{\mathcal{D}(s)} \Pr\left[\mathcal{D}(s)\right] \Pr\left[I_{\text{rep}} < \mathtt{R} | \mathcal{D}(s)\right] \ . \tag{6.6}$$

Outage Conditioned on the Decoding Set For repetition coding, the random codebooks at the source and all potential relays are generated i.i.d. circularly-symmetric, complex Gaussian. Conditioned on $\mathcal{D}(s)$ being the decoding set, the mutual information between s and $d(s)$ is

$$I_{\text{rep}} = \frac{1}{m} \log \left(1 + \mathtt{SNR} \left| a_{s,d(s)} \right|^2 + \mathtt{SNR} \sum_{r \in \mathcal{D}(s)} \left| a_{r,d(s)} \right|^2 \right) \ . \tag{6.7}$$

Thus $\Pr\left[I_{\text{rep}} < \mathtt{R} | \mathcal{D}(s)\right]$ involves[1] $|\mathcal{D}(s)| + 1$ independent fading coefficients, so we expect it to decay asymptotically proportional to $1/\mathtt{SNR}^{|\mathcal{D}(s)|+1}$. Indeed, we develop the following high $\mathtt{SNR}$ approximation[2] in Appendix B.2.2:

$$\Pr\left[I_{\text{rep}} < \mathtt{R} | \mathcal{D}(s)\right] \sim \left[\frac{2^{m\mathtt{R}} - 1}{\mathtt{SNR}} \right]^{|\mathcal{D}(s)|+1} \times \lambda_{s,d(s)} \prod_{r \in \mathcal{D}(s)} \lambda_{r,d(s)} \times \frac{1}{(|\mathcal{D}(s)| + 1)!} \ . \tag{6.8}$$

Note that we have expressed (6.8) in such a way that the first term captures the dependence upon $\mathtt{SNR}$ and the second term captures the dependence upon $\{\lambda_{i,j}\}$.

Decoding Set Probability Next, we consider the term $\Pr\left[\mathcal{D}(s)\right]$, the probability of a particular decoding set. As one rule for selecting from the potential relays, we can require that a potential relay fully decode the source message in order to participate in the second phase. Indeed, full decoding is required in order for the mutual information expression (6.7) to be correct; however, nothing prevents us from imposing additional restrictions on the members of the set $\mathcal{D}(s)$. For example, we might require that a potential relay fully decode *and* see a realized SNR some factor larger than its average, to either the source, the destination, or both.

Since the realized mutual information between s and r for i.i.d. complex Gaussian codebooks is given by

$$\frac{1}{m} \log \left(1 + \mathtt{SNR} \left| a_{s,r} \right|^2 \right) \ ,$$

[1] For a set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality of the set. This should not be confused with the usual notation for absolute value $|x|$ of a variable x.

[2] As before, the approximation $f(\mathtt{SNR}) \sim g(\mathtt{SNR})$ is in the sense of $f(\mathtt{SNR})/g(\mathtt{SNR}) \to 1$ as $\mathtt{SNR} \to \infty$.

under this rule we have

$$\begin{aligned}\Pr\left[r \in \mathcal{D}(s)\right] &= \Pr\left[|\mathsf{a}_{s,r}|^2 > (2^{m\mathtt{R}}-1)/\mathtt{SNR}\right] \\ &= \exp[-\lambda_{s,r}(2^{m\mathtt{R}}-1)/\mathtt{SNR}] \ .\end{aligned}$$

Moreover, since each potential relay makes its decision independently under the above restrictions, and the fading coefficients are independent in our model, we have

$$\begin{aligned}\Pr\left[\mathcal{D}(s)\right] &= \prod_{r\in\mathcal{D}(s)} \exp[-\lambda_{s,r}(2^{m\mathtt{R}}-1)/\mathtt{SNR}] \times \prod_{r\notin\mathcal{D}(s)} (1-\exp[-\lambda_{s,r}(2^{m\mathtt{R}}-1)/\mathtt{SNR}]) \\ &\sim \left[\frac{2^{m\mathtt{R}}-1}{\mathtt{SNR}}\right]^{m-|\mathcal{D}(s)|-1} \times \prod_{r\notin\mathcal{D}(s)} \lambda_{s,r} \ . \end{aligned} \tag{6.9}$$

Note that any selection means by which $\Pr\left[r \in \mathcal{D}(s)\right] \sim 1$ and $(1-\Pr\left[r \in \mathcal{D}(s)\right]) \propto 1/\mathtt{SNR}$, for $\mathtt{SNR}$ large, independently for each r, will result in similar asymptotic behavior for $\Pr\left[\mathcal{D}(s)\right]$.

Combining (6.8) and (6.9) into (6.6), we obtain

$$\Pr\left[I_{\text{rep}} < \mathtt{R}\right] \sim \left[\frac{2^{m\mathtt{R}}-1}{\mathtt{SNR}}\right]^{m} \times \sum_{\mathcal{D}(s)} \lambda_{s,d(s)} \prod_{r\in\mathcal{D}(s)} \lambda_{r,d(s)} \prod_{r\notin\mathcal{D}(s)} \lambda_{s,r} \times \frac{1}{(|\mathcal{D}(s)|+1)!} \ . \tag{6.10}$$

Fig. 6-5 compares the results of numeric integration of the actual outage probability to computing the approximation (6.10), for an increasing number of terminals with $\lambda_{i,j}=1$. As the result (6.10) and Fig. 6-5 indicate, repetition decode-and-forward cooperative diversity achieves full spatial diversity of order m, the number of cooperating terminals, for sufficiently large SNR. However, the SNR loss due to bandwidth inefficiency is exponential in m.

Convenient Bounds

While the approximation given in (6.10) is quite general and can be numerically evaluated to determine performance, it is not very convenient for further analysis. Its complexity results from dependence upon $\{\lambda_{i,j}\}$. In this section, we developed upper and lower bounds for (6.10) that we exploit in the sequel.

Our objective is to simplify the summation in (6.10). To this end, we note that for a given decoding set $\mathcal{D}(s)$, either $r \in \mathcal{D}(s)$, in which case $\lambda_{r,d(s)}$ appears in the corresponding

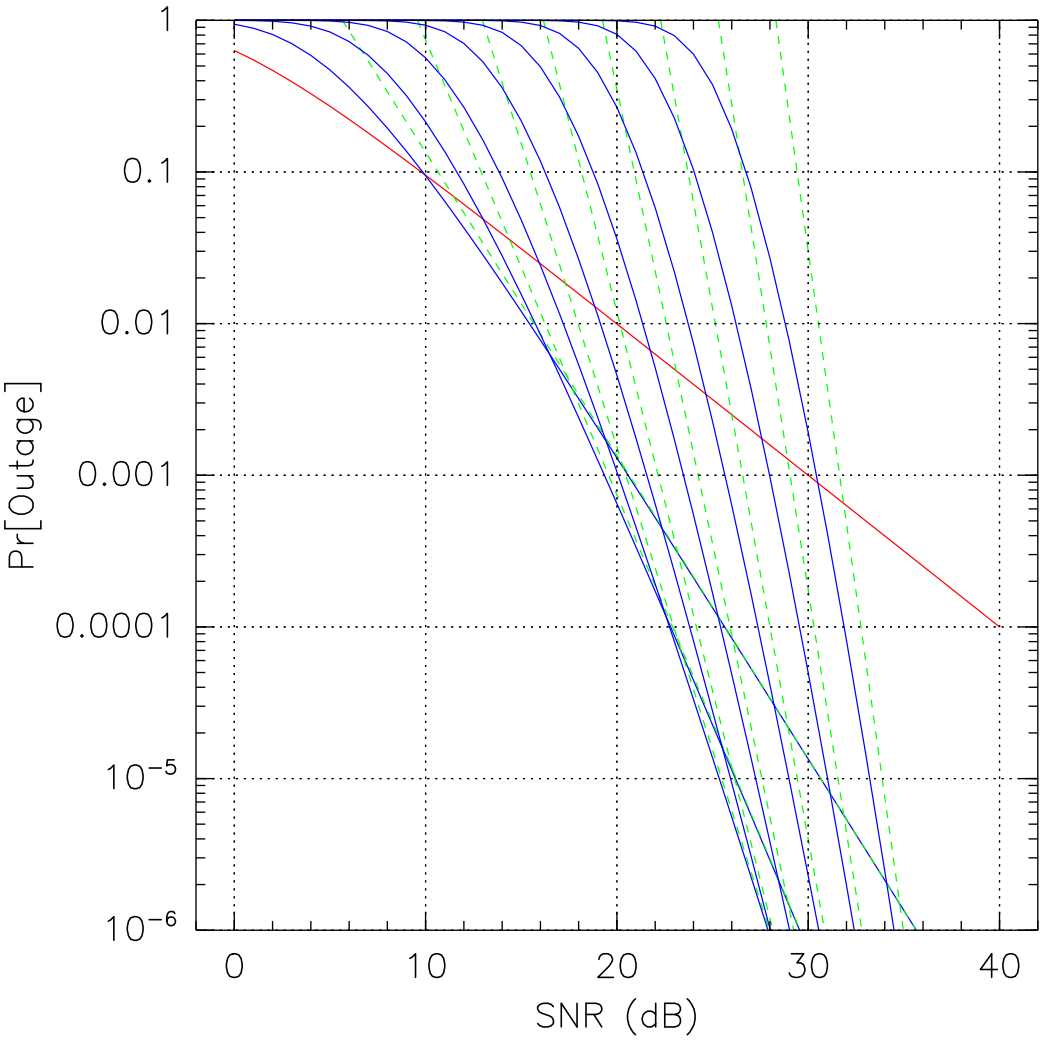


Figure 6-5: Comparison of numeric integration of the outage probability (solid lines) to calculation of the outage probability approximation (6.10) (dashed lines) vs. normalized SNR for different network sizes $m = 1, 2, \ldots, 10$. Successively lower curves at high SNR correspond to larger networks. For simplicity of exposition, we have plotted the case of $\mathtt{R} = 1$ b/s/Hz and $\lambda_{i,j} = 1$; more generally the plot can be readily updated to incorporate a model of the network geometry.

term in (6.10), or $r \notin \mathcal{D}(s)$, in which case $\lambda_{s,d(s)}$ appears in the corresponding term in (6.10). We therefore define

$$\underline{\lambda}_r \stackrel{\triangle}{=} \min\{\lambda_{r,d(s)},\, \lambda_{s,r}\} \ , \quad \overline{\lambda}_r \stackrel{\triangle}{=} \max\{\lambda_{r,d(s)},\, \lambda_{s,r}\} \ , \tag{6.11}$$

and $\overline{\lambda}_s \stackrel{\triangle}{=} \underline{\lambda}_s \stackrel{\triangle}{=} \lambda_{s,d(s)}$. Then the product dependent upon $\{\lambda_{i,j}\}$ is bounded by

$$\underline{\lambda}^m \leq \lambda_{s,d(s)} \prod_{r \in \mathcal{D}(s)} \lambda_{r,d(s)} \prod_{r \notin \mathcal{D}(s)} \lambda_{s,r} \leq \overline{\lambda}^m \ , \tag{6.12}$$

where $\underline{\lambda}$ is the geometric mean of the $\underline{\lambda}_i$ and $\overline{\lambda}$ is the geometric mean of the $\overline{\lambda}_i$, for $i \in \mathcal{M}$. We note that the upper and lower bounds in (6.12) are independent of $\mathcal{D}(s)$. We also note that the bounds in (6.12) coincide, *i.e.*, $\overline{\lambda} = \underline{\lambda}$, if, though not only if, $\overline{\lambda}_i = \underline{\lambda}_i$ for all $i \in \mathcal{M}$. Viewing $\lambda_{i,j}$ as a measure of distance between terminals i and j, the class of planar network geometries that satisfy this condition are those in which all the relays lie with arbitrary spacing along the perpendicular bisector between the source and destination. A complete study of the effects of such network geometry on performance is warranted, but beyond the scope of this dissertation.

Substituting (6.12) into (6.10), we arrive at the following simplified bounds for outage probability

$$\left[\frac{2^{m\mathtt{R}}-1}{\mathtt{SNR}/\underline{\lambda}}\right]^m \sum_{\mathcal{D}(s)} \frac{1}{(|\mathcal{D}(s)|+1)!} \leq \Pr\left[I_{\text{rep}} < \mathtt{R}\right] \leq \left[\frac{2^{m\mathtt{R}}-1}{\mathtt{SNR}/\overline{\lambda}}\right]^m \sum_{\mathcal{D}(s)} \frac{1}{(|\mathcal{D}(s)|+1)!} \ . \tag{6.13}$$

6.1.3 Space-Time Coded Cooperative Diversity

We now analyze performance of a decode-and-forward space-time cooperative diversity algorithm. Such protocols operate in similar fashion to the repetition decode-and-forward cooperative diversity algorithm analyzed in the previous section, except that all the relays transmit simultaneously on the same subchannel using a suitable space-time code. Again, we will see that protocols of this form offer full spatial diversity in the number of cooperating terminals, not just the number of decoding relays participating in the second phase. In addition, these algorithms have superior bandwidth efficiency to repetition-based algorithms.

Mutual Information and Outage Probability

As above, we utilize the total probability law to write

$$\Pr\left[I_{\text{stc}} < \mathtt{R}\right] = \sum_{\mathcal{D}(s)} \Pr\left[\mathcal{D}(s)\right] \Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right] \ , \tag{6.14}$$

and examine each term in the summation.

Outage Conditioned on the Decoding Set Conditioned on $\mathcal{D}(s)$ being the decoding set, the mutual information between s and $d(s)$ for random codebooks generated i.i.d. circularly-symmetric, complex Gaussian at the source and all potential relays can be shown to be

$$I_{\text{stc}} = \frac{1}{2}\log\left(1 + \frac{2}{m}\mathtt{SNR}\left|\mathsf{a}_{s,d(s)}\right|^2\right) + \frac{1}{2}\log\left(1 + \frac{2}{m}\mathtt{SNR}\sum_{r\in\mathcal{D}(s)}\left|\mathsf{a}_{r,d(s)}\right|^2\right) \ , \tag{6.15}$$

the sum of the mutual informations for two "parallel" channels, one from the source to the destination, and one from the set of decoding relays to the destination. Again, $\Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right]$ involves $|\mathcal{D}(s)| + 1$ independent fading coefficients, so we expect it to decay asymptotically proportional to $1/\mathtt{SNR}^{|\mathcal{D}(s)|+1}$. We develop the following high SNR approximation in Appendix B.2.3:

$$\Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right] \sim \left[\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]^{|\mathcal{D}(s)|+1} \times \lambda_{s,d(s)} \prod_{r\in\mathcal{D}(s)} \lambda_{r,d(s)} \times A_{|\mathcal{D}(s)|}(2^{2\mathtt{R}}-1) \ , \tag{6.16}$$

where

$$A_n(t) \triangleq \frac{1}{(n-1)!}\int_0^1 \frac{w^{(n-1)}(1-w)}{(1+tw)}dw \ , \qquad n > 0 \ , \tag{6.17}$$

and $A_0(t) = 1$. Note that we have expressed (6.16) in such a way that the first term captures the dependence upon $\mathtt{SNR}$ and the second term captures the dependence upon $\{\lambda_{i,j}\}$.

Decoding Set Probability Next, we consider the term $\Pr\left[\mathcal{D}(s)\right]$, the probability of a particular decoding set. As before, we require that a potential relay fully decode the source message in order to participate in the second phase, a necessary condition for the mutual information expression (6.15) to be correct.

Since the realized mutual information between s and r for i.i.d. complex Gaussian code-

books is given by

$$\frac{1}{2}\log\left(1+\frac{2}{m}\mathtt{SNR}\,|a_{s,r}|^2\right)\ ,$$

under this rule we have

$$\begin{aligned}\Pr\left[r\in\mathcal{D}(s)\right] &= \Pr\left[|a_{s,r}|^2 > \frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]\\ &= \exp\left[-\lambda_{s,r}\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]\ .\end{aligned}$$

Moreover, since each potential relay makes its decision independently, and the fading coefficients are independent in our model, we have

$$\begin{aligned}\Pr\left[\mathcal{D}(s)\right] &= \prod_{r\in\mathcal{D}(s)}\exp\left[-\lambda_{s,r}\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]\times\prod_{r\notin\mathcal{D}(s)}\left(1-\exp\left[-\lambda_{s,r}\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]\right)\\ &\sim\left[\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]^{m-|\mathcal{D}(s)|-1}\times\prod_{r\notin\mathcal{D}(s)}\lambda_{s,r}\ .\end{aligned}\tag{6.18}$$

Combining (6.16) and (6.18) into (6.14), we obtain

$$\Pr\left[I_{\mathrm{stc}}<\mathtt{R}\right]\sim\left[\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/m}\right]^{m}\times\sum_{\mathcal{D}(s)}\lambda_{s,d(s)}\prod_{r\in\mathcal{D}(s)}\lambda_{r,d(s)}\prod_{r\notin\mathcal{D}(s)}\lambda_{s,r}\times A_{|\mathcal{D}(s)|}(2^{2\mathtt{R}}-1)\ .\tag{6.19}$$

Fig. 6-6 compares the results of numeric integration of the actual outage probability to computing the approximation (6.19), for an increasing number of terminals with $\lambda_{i,j}=1$. As the result (6.19) and Fig. 6-6 indicate, space-time coded cooperative diversity achieves full spatial diversity of order m, the number of cooperating terminals, for sufficiently large SNR. In contrast to repetition-based algorithms, the SNR loss for space-time coded cooperative diversity is only linear in m.

Convenient Bounds

Again, although the approximation given in (6.19) is quite general and can be numerically evaluated to determine performance, it is not very convenient for further analysis. There are two factors contributing to its complexity: dependence upon $\{\lambda_{i,j}\}$, and the involved closed-form expression for $A_n(t)$ as n grows. In this section, we developed upper and lower bounds for (6.19) that we exploit in the sequel.

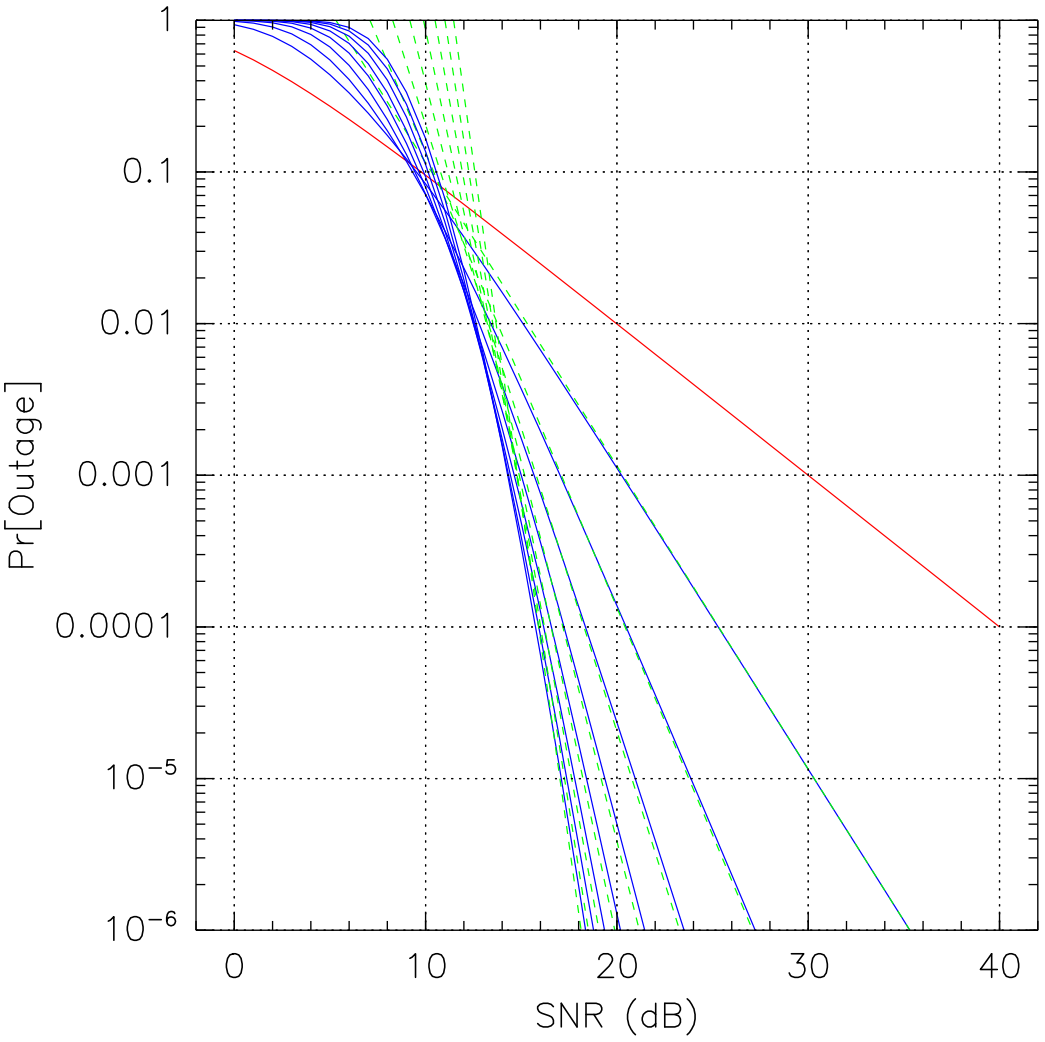


Figure 6-6: Comparison of numeric integration of the outage probability (solid lines) to calculation of the outage probability approximation (6.19) (dashed lines) vs. normalized SNR for different network sizes $m = 1, 2, \ldots, 10$. Successively lower curves at high SNR correspond to larger networks. For simplicity of exposition, we have plotted the case of $\mathtt{R} = 1$ b/s/Hz and $\lambda_{i,j} = 1$; more generally the plot can be readily updated to incorporate a model of the network geometry.

Our objective is to simplify the summation in (6.19). The product dependent upon $\{\lambda_{i,j}\}$ can again be bounded as in (6.12). To avoid dealing with (6.17), we exploit the bounds

$$\frac{1}{(n+1)!(1+t)} \leq A_n(t) \leq \frac{1}{n!} \ . \tag{6.20}$$

Combining (6.12) and (6.20) into (6.19), we arrive at the following simplified bounds for outage probability

$$\Pr\left[I_{\text{stc}} < \mathtt{R}\right] \geq \left[\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/(m\underline{\lambda})}\right]^m 2^{-2\mathtt{R}} \sum_{\mathcal{D}(s)} \frac{1}{(|\mathcal{D}(s)|+1)!} \tag{6.21}$$

$$\Pr\left[I_{\text{stc}} < \mathtt{R}\right] \leq \left[\frac{2^{2\mathtt{R}}-1}{2\mathtt{SNR}/(m\overline{\lambda})}\right]^m \sum_{\mathcal{D}(s)} \frac{1}{|\mathcal{D}(s)|!} \ . \tag{6.22}$$

Practical Issues

Space-Time Code Design The outage analysis in Section 6.1.3 relies on a random coding argument, and demonstrates that full spatial diversity can be achieved using such a rich set of codes. In practice, one may wonder whether or not there exist space-time codes for which the number of participating antennas is not known *a priori* and yet full diversity can be achieved. More specifically, if we design a space-time code for a maximum of N transmit antennas, but only a randomly selected subset of n of those antennas actually transmit, can the space-time code offer diversity n? It turns out that the class of space-time block codes based upon orthogonal designs have this property [56]. Essentially, these codes have orthogonal waveforms emitted from each antenna, corresponding to columns in a code matrix. Absence of an antenna corresponds to deletion of a column in the matrix, analogous to that antenna experiencing a deep fade, but the columns remain orthogonal, allowing the code to maintain its diversity benefits. Thus space-time coded cooperative diversity protocols may be readily deployed in practice using these codes.

Distributed Implementation Given a suitably designed space-time code, space-time coded cooperative diversity reduces to a simple, distributed network protocol. The network must possess a means for distributing columns from the code matrix to the terminals, as well as coordinating the medium-access control. With these elements in place, when each terminal transmits its message, the other terminals receive and potentially decode, requiring only an SNR measurement. If a relay can decode, it transmits the information in the second

phase using its column from the space-time code matrix. Because the destination receiver can measure the fading, it can determine which relays are involved in the second phase and adapt its decoding rule appropriately. Although certainly the terminals could exchange more information in order to adapt power to the network geometry, for example, such overhead is not required in order to obtain full diversity.

One of the key challenges to implementing such protocols could be block and symbol synchronization of the cooperating terminals. Such synchronization might be obtained through periodic transmission of known synchronization prefixes, as proposed in current wireless LAN standards [35]. A detailed study of issues involved with synchronization is beyond the scope of this dissertation.

6.1.4 Diversity-Multiplexing Tradeoff

As we saw in Chapter 5, an interesting tradeoff between diversity and multiplexing arises when we parameterize ours results in terms of $(\mathtt{SNR}, \mathtt{R}_{\mathrm{norm}})$. Specifically, when we approximate $\Pr[I < \mathtt{R}] \doteq \mathtt{SNR}^{-\Delta(\mathtt{R}_{\mathrm{norm}})}$, in the sense of equality to first-order in the exponent, *i.e.*,

$$\Delta(\mathtt{R}_{\mathrm{norm}}) = \lim_{\mathtt{SNR}\to\infty} -\frac{\log(\Pr[I < R])}{\log(\mathtt{SNR})} , \tag{6.23}$$

we find that increasing $\mathtt{R}_{\mathrm{norm}}$ reduces Δ.

Utilizing the lower and upper bounds (6.13) in (6.23) yields diversity order

$$\Delta_{\mathrm{rep}}(\mathtt{R}_{\mathrm{norm}}) = m(1 - m\mathtt{R}_{\mathrm{norm}}) \tag{6.24}$$

for repetition decode-and-forward cooperative diversity. Similarly, utilizing the lower and upper bounds (6.21)-(6.22) in (6.23) yields upper and lower bounds, respectively, on the diversity order

$$m(1 - 2\mathtt{R}_{\mathrm{norm}}) \leq \Delta_{\mathrm{stc}}(\mathtt{R}_{\mathrm{norm}}) \leq m\left(1 - \left\lceil\frac{m-1}{m}\right\rceil 2\mathtt{R}_{\mathrm{norm}}\right) \tag{6.25}$$

for space-time coded cooperative diversity.

Fig. 6-7 compares the diversity exponents, along with the corresponding tradeoff for non-cooperative transmission, $\Delta_{\mathrm{dir}}(\mathtt{R}_{\mathrm{norm}}) = 1 - \mathtt{R}_{\mathrm{norm}}$, from Chapter 5. Both repetition-based and space-time coded cooperative diversity offer full diversity m as $\mathtt{R}_{\mathrm{norm}} \to 0$. Clearly,

space-time coded cooperative diversity offers larger diversity order than repetition-based algorithms and can be effectively utilized for higher spectral efficiencies than repetition-based schemes.

6.2 Partially Cooperative Networks

The results in Section 6.1 suggest that, although the fully cooperative protocols provide full spatial diversity gains (steeper slopes), SNR losses (shifts of the curves to the right) due to bandwidth inefficiency and geometry-dependent path-loss can grow with the number of cooperating terminals, to the point that they outweigh the diversity benefits in practical operating regimes. This observation, coupled with the fact that diversity gains generally exhibit diminishing returns, suggests that it can be beneficial in large networks to partition the terminals into relatively small cooperating groups.

Grouping terminals requires additional network overhead in terms of disseminating information about the network, *e.g.*, path-losses among terminals, executing an algorithm to collect terminals into cooperating groups, and controlling medium access for the cooperating groups as in Figures 6-3 and 6-4. More than anywhere else in the dissertation, the issue of layering becomes particularly important in this context.

In the remainder of this section, we illustrate how, from an architectural perspective, repetition and space-time coded cooperative diversity may each amenable to different settings. Repetition-based schemes require relatively low complexity in the terminals, but require more complexity in the network for deciding which terminals cooperate in order for the algorithms to be effective; thus, these algorithms are well-suited to infrastructure networks, *e.g.*, cellular, satellite, and certain wireless LAN configurations, in which terminals communicate directly to an access point that selects the cooperating groups. To manage complexity in the access point, we use our analytical results from Section 6.1 to describe a variety of potential grouping algorithms in Section 6.2.1 based upon set partitioning and weighting matching in graphs. By contrast, space-time coded cooperative diversity requires more complexity in the terminals, but readily extends to distributed implementation; thus, these algorithms may be well-suited to ad-hoc networks, and especially ad-hoc networks with clusters. As we illustrate in Section 6.2.2, space-time coded cooperative diversity can leverage existing clustering algorithms to perform grouping; thus, our discussion of this case

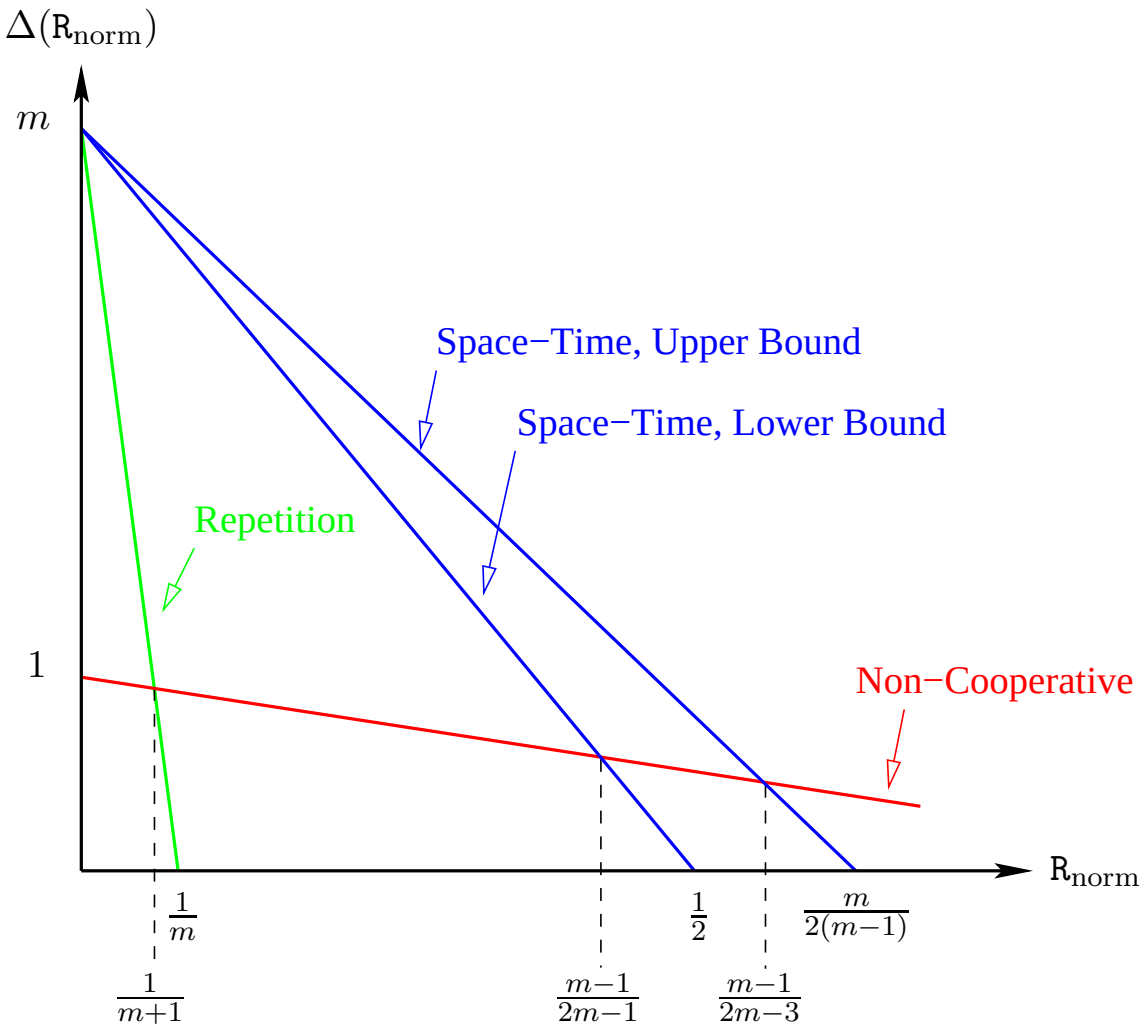


Figure 6-7: Diversity order $\Delta(\mathtt{R}_{\text{norm}})$ for non-cooperative transmission, repetition-based cooperative diversity, and space-time coded cooperative diversity. As $\mathtt{R}_{\text{norm}} \to 0$, all cooperative diversity protocols provide full spatial diversity order m, the number of cooperating terminals. Relative to direct transmission, space-time coded cooperative diversity can be effectively utilized for a much broader range of $\mathtt{R}_{\text{norm}}$ than repetition-coded cooperative diversity, especially as m becomes large.

is brief. Finally, in Section 6.2.3, we briefly comment on layering issues for consideration by protocol designers. In all these sections, our study is by no means complete.

6.2.1 Centralized Partitioning for Infrastructure Networks

Our focus in this section is on infrastructure networks (*cf.* Section 2.3.1), in which all terminals communicate through an access point (AP). In such scenarios, the AP can gather information about the state of the network, *e.g.*, the path-losses among terminals, select a cooperative mode based upon some network performance criterion, and feed back its decision on the appropriate control channels. Here cooperative diversity lives across the medium-access control, and physical layers (*cf.* Fig. 1-2); routing is not considered.

Although centralized grouping can be performed for both repetition-based and space-time coded cooperative diversity, we expect it to be most beneficial for repetition-based algorithms because of their exponentially growing SNR loss in the number of cooperating terminals. Because repetition-based algorithms generally allow for lower computational complexity in the terminals, repetition-based algorithms with centralized grouping seem well-matched to infrastructure networks. This observation biases our discussion through the remainder of this section.

Managing Complexity

In Section 6.1, we developed a particular pair of algorithms in which all the cooperating terminals performed a single cooperating action on behalf of one another. More generally, terminals can utilize any of several different cooperating modes, leading to various combinatorial optimization problems for selection of network-wide policies. Designing a practical grouping algorithm for a partially cooperative network requires balancing the performance of the resulting scheme with the computational complexity required in the access point. For small networks, the access point may be capable of performing brute-force optimization; for growing networks, reasonable approximations must be employed. We discuss some ideas for reducing grouping algorithm complexity in this section.

To develop a sense of the inherent complexity of general cooperation schemes, consider the case of m terminals communicating to the AP using any of n cooperating modes, *e.g.*, direct, amplify-and-forward, decode-and-forward, and superpositions thereof. For a given source, each of the remaining $m-1$ terminals acting as relays has n options for cooperation.

Thus, there are $n^{(m-1)}$ options for the particular source, or $n^{m(m-1)}$ options for the entire network. While optimizing over all these options might be feasible for small m and n, it becomes intractable as m becomes large, even for small n. The combinatorial nature of the general problem prevents us from obtaining structured ways of reducing the search options while maintaining optimal solutions.

Of course, we can impose restrictions on the search in order to reduce the complexity of the problem. For example, we can impose certain symmetries in the communication schemes, so that terminals in a pair operate similarly on each other's behalf. We can also limit the number of cooperating modes and then ask how terminals should be grouped using only those modes. With a suitable cost function, this problem is an instance of the classical *set partitioning problem* [63], an integer programming problem which is, in general, NP-complete. Even with these simplifying constraints, complexity considerations limit the suitability of such algorithm for large m.

Based upon our earlier results, we can develop reasonable heuristics for further reducing the search. For example, because of the increasing bandwidth penalties for exploiting repetition-based cooperative diversity, as well as the diminishing returns of the diversity gains, it seems reasonable to only allow small sets of cooperating terminals, in the range of $2-5$ terminals, for example. Furthermore, using the results of Chapter 5 and Section 6.1, we can relate the performance of a given cooperating set to the set of fading variances $\lambda_{i,j}$ among members of the set. Using path-loss models of the form $\lambda_{i,j} \propto d_{i,j}^{-\alpha}$ for the variances, where $d_{i,j}$ is the distance between a pair of terminals and $2 \leq \alpha \leq 5$ is the path loss exponent, we see that selection of cooperating sets can be related to network geometry. Although the dependence upon the geometry in this case is more involved, recall that many routing algorithms utilize network geometry as a basis for selecting routes, *e.g.*, shortest-path routing algorithms.

Based upon these heuristics, we now examine several algorithms for matching terminals into cooperating *pairs*. We then give an example of their performance. Finally, we comment on hierarchical matching algorithms.

Matching Algorithms

We first consider grouping terminals into cooperating pairs, due to complexity issues and for simplicity of exposition. Matching algorithms seem well-suited to repetition-based algo-

rithms for infrastructure networks. While we comment on more general grouping algorithms later, a detailed study of grouping algorithms is beyond the scope of this dissertation.

As we will see, choosing pairs of cooperating terminals is an instance of a more general set of problems known as *matching* problems on graphs [63, Section 10.2]. To outline the general matching framework, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph, with $\mathcal{V}$ a set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ a set of edges between vertices. A subset $\mathcal{M}$ of $\mathcal{E}$ is called a *matching* if edges in $\mathcal{M}$ are pairwise disjoint, *i.e.*, no two edges in $\mathcal{M}$ are incident on the same vertex. Note that

$$|\mathcal{M}| \leq \lfloor |\mathcal{V}|/2 \rfloor \ , \tag{6.26}$$

where $|\mathcal{M}|$ is again the cardinality of the set $\mathcal{M}$ and $\lfloor x \rfloor$ denotes the usual floor function. When the bound (6.26) is achieved with equality, the matching is called a *perfect matching*. Since we will be working with *complete* graphs, *i.e.*, there is an edge between each pair of vertices, there will always be a perfect matching for $|\mathcal{V}|$ even. As a result, we will not be concerned with so-called *maximal matching* problems.

Instead, we focus on *weighted matching* problems. Given an edge e in $\mathcal{E}$, the *weight* of the edge is some real number $w(e)$. Given a subset $\mathcal{S}$ of $\mathcal{E}$, we denote its sum weight by

$$w(\mathcal{S}) = \sum_{e \in \mathcal{S}} w(e) \ . \tag{6.27}$$

The *minimal weighted matching* problem is to find a matching $\mathcal{M}$ of minimal weight [63]. We also consider two other matching algorithms, both based upon randomization, that approximate minimal weighted matching and offer lower complexity.

Specifically, we consider the following algorithms:

- **Minimal Weighted Matching:** Since algorithms for implementing minimal weighted matching are well-studied and readily available [4, 63], we do not go into their details. We note, however, that more recent algorithms for minimal weighted matching have complexity $O(|\mathcal{V}|^3)$ [63].

- **Greedy Matching:** To reduce complexity and approximate minimal weighted matching, we consider a greedy algorithm in which we randomly select a free vertex v and match it with another free vertex v' such that the edge $e = (v, v')$ has minimal weight. The process continues until all of the vertices have been matched. Since each step of

the algorithm takes at most $|\mathcal{V}|$ comparisons, and there are $|\mathcal{V}|/2$ steps, the complexity of this algorithm is $O(|\mathcal{V}|^2)$. We note that this greedy algorithm need not be optimal for this order of complexity.

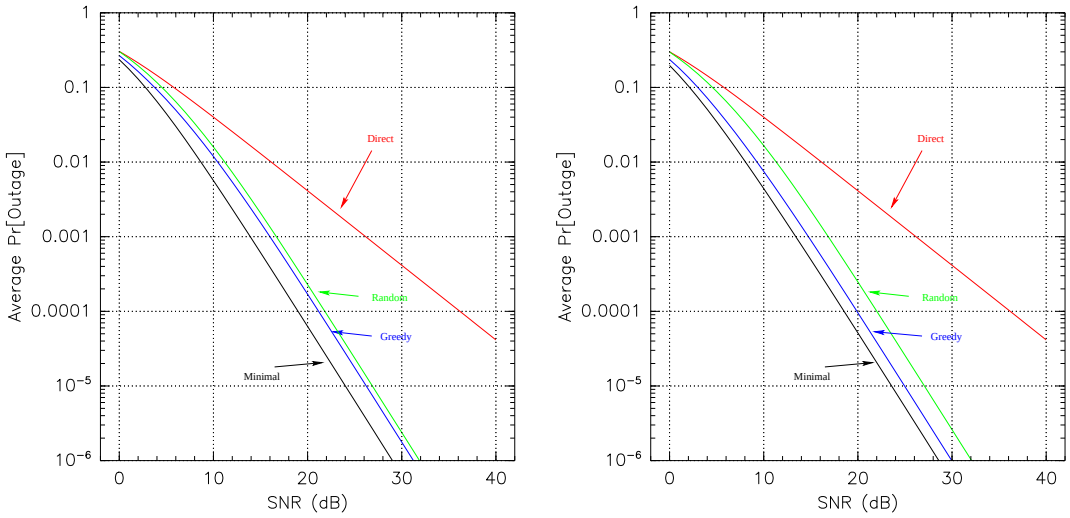
- **Random Matching:** To reduce complexity still further, we consider a random matching algorithm where we vertex pairs randomly. The complexity of this algorithm is $O(|\mathcal{V}|)$.

In addition to the algorithms outlined above for matching cooperating terminals, there are a variety of other possibilities. Instead of the general weighted matching approach, we can randomly partition the terminals into two sets and utilize *bipartite weighted matching* algorithms, which have slightly lower complexity (in terms of their coefficients, not order) and are conceptually simpler to implement than general weighted matching algorithms [63, Section 10.2.2]. Another possibility is to again randomly partition the terminals into two sets and utilize *stable marriage* algorithms with still lower complexity $O(|\mathcal{V}|^2)$. Such algorithms may be suitable for decentralized implementation [4, Section 12.5].

Example Performance

Fig. 6-8 shows a set of example results from the various matching algorithms described above. Terminals are independently and uniformly distributed in a square of side 2000 m, with the basestation/access point located in the center of the square. Fading variances are computed using a $d^{-\alpha}$ path-loss model, with $\alpha = 3$. The weight of an edge $e = (v, v')$ is the average of the outage probabilities for terminal v using v' as a relay, and vice versa. In particular, we utilize the adaptive decoding-and-forward performance result (5.22) from Chapter 5 for this example; more generally, we can employ any of the outage probability expressions for a pair of cooperating terminals as developed in this dissertation. Each set of results is averaged over 100 trial networks with the various matching algorithms applied. The results are normalized so that the direct performance is the same in each trial, *i.e.*, the received SNR for direct transmission averaged over all the terminals in the network is normalized to be the horizontal axis in Fig. 6-8.

We note several features of the results in Fig. 6-8. First, all the matching algorithms exhibit full diversity gain of order two with respect to direct transmission. As we would expect, random, greedy, and minimal matching perform increasingly better, but only in terms



(a) $n = 10$ (b) $n = 50$

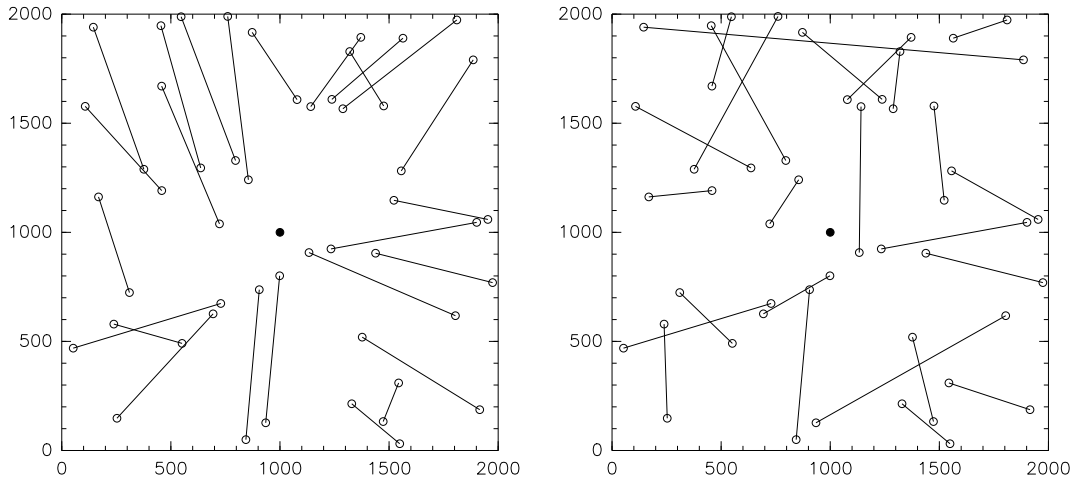
Figure 6-8: Matching algorithm performance in terms of average outage probability vs. received SNR (normalized for direct transmission).

of SNR gain. Although diversity gain remains constant because we only group terminals into cooperating pairs, the relative SNR gain does improve slightly with increasing network size. This effect appears most pronounced in the case of greedy matching. This observation suggests that optimal matching is more crucial to good performance in small networks, because there are fewer choices among a small number of terminals. In general, the SNR gains of the more computationally demanding matching algorithms are most beneficial in low to moderate SNR regimes where the benefits of the diversity gains are smallest. As the diversity gains increase for higher SNR, its becomes less crucial to utilize complex matching algorithms.

Fig. 6-9 compares the results of minimal and greedy matching for a sample network with 50 terminals. We see that the minimal matching tends to have pairs such that one of the terminals is almost on the line connecting the basestation and the other terminal. By comparison, the greedy matching algorithm exhibits much more randomness.

Constrained Partitioning Algorithms and Hierarchical Matching

General weighted matching can be viewed as an instance of the classical set partitioning problem with subsets constrained to be of size two. More generally, grouping into triplets, quartets, and so on, can be viewed as an instance of the set partitioning problem with



(a) Minimal Matching (b) Greedy Matching

Figure 6-9: Matching algorithm results for an example network: (a) minimal matching, (b) greedy matching. Terminals are indicated by circles, and matched terminals are connected with lines.

subsets constrained to be of sizes 3, 4, and so on, respectively. While such partitioning problems can be solved in principle, their complexity grows dramatically with the size of the network.

To reduce this complexity for large networks, one can employ matching algorithms in hierarchical fashion. As an illustration, consider grouping into triplets. We first randomly partition the network into subnetworks of size $2n/3$ and $n/3$, respectively. We execute a matching algorithm on the subnetwork of size $2n/3$ to obtain $n/3$ pairs of terminals. Next we execute another matching algorithm between the $n/3$ pairs and the remaining $n/3$ terminals to form triplets. Alternatively, the greedy matching algorithm described previously can be generalized to first find the best pair of terminals in the network, and then find the best third terminal to join the pair. If n is reasonably large, then the results of Fig. 6-8 suggest that low-complexity algorithms of this form might perform well.

Our point in this discussion is to illustrate some reasonable ways of approximating set partitioning when the size of the network is too large for a brute-force approach. Looking at the case of matching, and hierarchical matching, we obtain relatively simple algorithms that offer full diversity and reasonable SNR gain. A complete examination of the tradeoff between performance and computational complexity is beyond the scope of this dissertation, but we note that the conclusions of any such study will depend upon operating regimes and

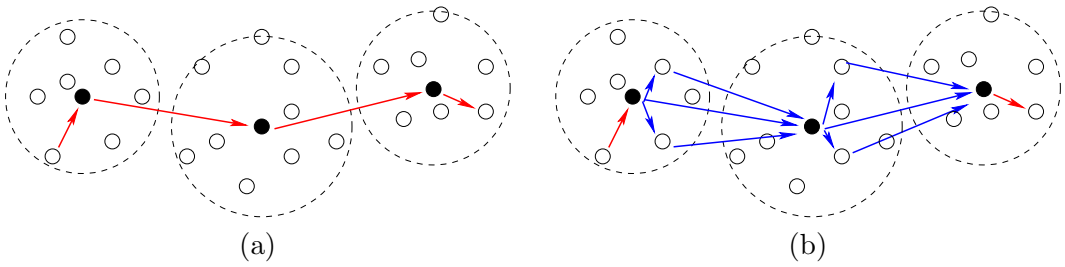


Figure 6-10: Clustering with (a) direct transmission and (b) cooperative diversity transmission.

other system factors.

6.2.2 Clustering in Ad-Hoc Networks

Our focus in this section is on clustering in ad-hoc networks (*cf.* Section 2.3.1), and how cooperative diversity can be integrated into such architectures. As we discussed in Chapter 2, clustering arises in a variety of forms in large, dense ad-hoc networks. Clustering algorithms partition a large ad-hoc network into a set of clusters, each centered around a clusterhead. Terminals communicate directly to their associated clusterhead, and routing is usually performed between clusterheads. In this sense, clustering mimics some of the features of current cellular networks: clusters correspond to cells and clusterheads correspond to basestations. However, in ad-hoc settings the clusters and clusterheads may be varying as the network operates, the clusterheads themselves can have information to transmit, and the clusterhead network must share the wireless bandwidth.

There are many tradeoffs in the design of clustering algorithms, too many to fully address in this dissertation. For example, clustering algorithms can be designed in order to reduce the complexity and overhead of routing through the network [20]; they can be designed in coordination with turning radios on and off in order to reduce power consumption in the network [15]; and they can be designed to facilitate fusion of measurements in sensor networks [36, 37].

Instead our objective in this section is to illustrate how cooperative diversity can be integrated into an existing clustered ad-hoc network. To this end, we consider three clusters as in Fig. 6-10. Fig. 6-10(a) illustrates how direct transmission can be utilized to communicate information between terminals in different clusters. A terminal transmits to

its clusterhead, clusterheads route the transmission to the destination cluster, and finally the destination clusterhead transmits to the destination terminal. Fig. 6-10(b) illustrates how each of inter-cluster direct transmissions can be converted into cooperative diversity transmissions along the lines developed in Chapter 5 and Section 6.1. If the average inter-terminal spacing is $\overline{r}$, we see that inter-cluster transmissions occur roughly over a distance $2\overline{r}$ on average. There are likely to be useful relays between clusterheads, and in order to coordinate the transmissions we utilize relays in the originating cluster.

As we increase the cluster size, the number of clusters in the network decreases. Thus the complexity of routing across the entire network decreases as well; however, the utility and complexity of routing within a cluster increases. From a diversity standpoint, the inter-cluster direct transmissions are over longer and longer distances, but there are more and more potential relays to exploit. The complexity and benefits of cooperative diversity between clusters thus increases as well. Again, at least from a diversity standpoint because of reduced bandwidth efficiency and diminishing returns of diversity gains, there is no reason to grow the cluster size too large. On the other hand, growing the cluster size allows more and more terminals to be asleep when they are not transmitting and still conserve the transport of the network [15]. The point is, there are a variety of issues to explore here, and cooperative diversity should be one of them.

6.2.3 Comments on Layering Issues

As the examples in Sections 6.2.1 and 6.2.2 suggest, cooperative diversity can be exploited in a variety of wireless network architectures. A natural question that arises is how well cooperative diversity fits within traditional layered architectures (*cf.* Fig. 1-2), or, alternatively, how much layered architectures must be modified in order for us to implement cooperative diversity. In this section, we comment on the various components of cooperative diversity and suggest where they might best be implemented. We discuss variations on our main theme depending upon the network and application. It is important to stress that our breakdown into protocol layers is coarse; indeed, new wireless network architectures with more layers and cross-layer functionality are constantly being proposed.

In this chapter and the preceding chapters, we have conceptually separated cooperative diversity into two sets of problems. The first set of problems involves deciding which terminals cooperate using what strategies. The second set of problems involves executing a

particular strategy within a set of cooperating terminals, *i.e.*, cooperatively transmitting, processing, and receiving signals to exploit spatial diversity in the channel. We generally envision the first set of problems being addressed at the medium-access control and perhaps link and network layers, using suitable models developed for the latter set of problems when addressed at the medium-access control and physical layers. Clearly, the medium-access control layer plays a central role in our vision for cooperative diversity in wireless networks.

The extent to which cooperative diversity can be completely implemented at the medium-access control layer can depend upon several considerations, including the connection format, the relaying strategy, and the type of diversity combining. Full maximum-ratio combining of real-valued received signals, or fine quantizations thereof, is difficult to implement at the medium-access control layer in datagram networks, because real valued signals must be passed up the protocol stack. Similarly, amplify-and-forward relaying can be difficult to implement at the MAC layer. On the other hand, selection combining at the destination and decode-and-forward transmission at the relay can be implemented in the medium-access control layer in such networks without substantial modifications to the data paths. Because selection combining performs 3 dB worse than maximum-ratio combining, which we have focused on throughout the dissertation, again the system designer must balance performance with computational and architectural complexity.

Chapter 7

Conclusions

This final chapter summarizes the contributions of the dissertation and highlights numerous areas for further research.

7.1 Contributions

In this dissertation, we have created a framework for designing and evaluating wireless network algorithms that take advantage of certain kinds of cooperation among terminals. Specifically, in these cooperative diversity techniques, as we have referred to them throughout, sets of terminals relay signals for each other to create a virtual antenna array. In so doing, the terminals trade off the costs, in power, bandwidth, and computational complexity, for the greater benefits gained by exploiting spatial diversity in the channel to combat multipath fading.

Tradeoffs among multiple cooperating terminals, each with its own information to communicate, seem natural, and were a key ingredient of our treatment. By contrast, in the classical relay channel model and its extensions, there is a single source terminal with information to communicate, and additional relay terminals without information to communicate. At a high-level, the relays represent additional resources, *e.g.*, power and computation, that can be freely utilized by the source terminal; thus, there is no tradeoff. Not surprisingly from this perspective, work on the relay channel often demonstrates that cooperative diversity is beneficial in certain regimes.

Our examination of a multiple-access channel with cooperative diversity in Chapter 4 led to refined or alternative conclusions. Specifically, we showed in the Gaussian noise

case, with and without ergodic fading, that the performance advantages in terms of ergodic capacity must be *shared* by the cooperating terminals; thus, there is an essential tradeoff in terms of which terminals receive the benefits of cooperative diversity. Moreover, we also demonstrated that, at least in terms of sum-rate in ergodic Rayleigh fading, there are no advantages to cooperative diversity if the cooperating terminals do not have channel state information available to them. These observations suggest that conclusions about the utility of cooperative diversity can vary with the channel model and system characteristics as well the target performance objective.

The next key ingredient of our framework was a focus on non-ergodic Rayleigh fading settings using outage probability as performance measures. We focused on scenarios in which the appropriate receivers obtain channel state information, but the transmitters do not obtain this information, or otherwise do not exploit it. Although certainly not the most general setting, we surveyed in Chapters 5 and 6 numerous issues that arose in the context of designing practical cooperative diversity algorithms, evaluating their performance in terms of outage probability, and formulating approaches to system design and architecture with these algorithms. These issues included:

- Practical System Constraints
  - Orthogonal transmission, for integration into many current network architectures
  - Half-duplex constraints, which led to performance losses due to inefficient use of channel bandwidth
- Channel Model Features
  - Geometry-dependent path-losses, which led to SNR losses or gains
  - Limited feedback from the destination, which allowed for improved bandwidth efficiency by re-transmitting only when necessary
- Terminal Coding and Processing
  - Source terminal channel coding and modulation
  - Relay processing, including amplify-and-forward and decode-and-forward
  - Destination processing, largely focusing on maximum-ratio combining, though the results can be carried through for selection or other forms of combining if desired.

- Grouping Algorithms, for grouping terminals into cooperating sets
- Layering and Cross-Layer Architectures

Many of these issues will arise under other channel models and system characteristics, so these chapters can serve as a blueprint for further design and analysis.

Chapter 5 developed energy-efficient algorithms for two cooperating terminals based upon the relays amplifying what they receive or fully decoding, re-encoding, and repeating the source messages. We designed simple protocols that achieved full spatial diversity and led to improved robustness to fading or substantial power savings, up to an order of magnitude or more in certain regimes. Chapter 6 extended these repetition-based cooperative diversity algorithms to fully cooperative networks, and also improved upon their bandwidth efficiency by developing space-time coded cooperative diversity. These developments also demonstrated that performance losses due to network geometry and bandwidth inefficiency can grow with the number of cooperating terminals, and even outweigh the benefits of additional diversity gain in practical operating regimes. For such circumstances, Chapter 6 also considered various grouping algorithms for collecting terminals into smaller cooperating sets. For infrastructure networks in which a centralized terminal can gather network information and implement a centralized algorithm, we proposed set partitioning as well as low-complexity approximations to it. For ad-hoc networks with clusters, we illustrated how grouping might occur in a more distributed fashion and take hints from other architectural considerations in the network, such as cluster formulation algorithms. These examples contribute reasonable guidelines for system design and architecture with cooperative diversity algorithms.

7.2 Future Research

There are a variety of fruitful areas for future research on cooperative diversity and related topics. We have mentioned many issues in earlier chapters, but we repeat some of the larger and more important ones here.

- **Radio and Network Implementation:** There are a number of modeling and analytical considerations that can be further explored in the context of cooperative diversity, as we point out below. We feel it is just as important, if not more impor-

tant, to prototype an implementation of cooperative diversity with practical hardware for some wireless network application. Our analysis has assumed a reasonable amount of block and symbol synchronization among the radios belonging to distributed terminals, but precluded the possibility of perfect phase synchronization and beamforming; a test-bed implementation would indicate the degree to which these assumptions are reasonable. Moreover, by fleshing out what it is and is not possible with distributed radio hardware in various regimes, a test-bed implementation would help prioritize the analytical issues to follow, as well as potentially reveal many other issues.

From a networking perspective, we suspect there are a whole host of issues related to layering and cross-layer design that we have not explored because we did not set out to fully specify protocols. A protocol designer trying to implement cooperative diversity within a given network stack could again help prioritize the analytical issues to follow as well as reveal many other important issues. For example, while we have briefly addressed the issue of limited feedback from the destination, it would be interesting to determine how general cooperative diversity transmissions and ARQ should interact, what the right abstractions are, and so forth.

- **Relative Value of Cooperative Diversity with Temporal or Spectral Diversity:** We saw that cooperative diversity may not be very beneficial when temporal diversity can be fully exploited in ergodic settings. We focused in the latter parts of the dissertation on scenarios in which only spatial diversity can be exploited through cooperative diversity. In more general channel environments, the channel may exhibit variations across time and frequency as well as space. Since there are many simple and effective methods for exploiting both temporal and spectral diversity, it is important to address whether, and to what extent, it is useful to additionally exploit spatial diversity through cooperative diversity. We expect spatial diversity gains obtained through cooperative diversity to be especially attractive in indoor settings where large amounts of bandwidth are required to experience frequency selectivity, and fixed wireless scenarios where there is little or no temporal selectivity.

- **General Channel Models and System Characteristics:** We focused through the dissertation on Rayleigh fading channels using capacity regions and outage probability as a performance measure. Moreover, we have considered scenarios in which channel

state information is available only to the appropriate receivers. We expect our results will readily extend to more general fading distributions, and in fact depend only upon the characteristics of those distributions near the origin. More general communication strategies than the ones considered in preceding chapters arise under different system characteristics and using alternative performance measures.

For example, if the transmitters obtain channel state information, power control becomes possible. For Gaussian multiple-access channels with ergodic fading, allocating the channel only to the terminal with the largest instantaneous SNR at each given instant achieves the maximum sum-rate for reliable communication in Shannon's sense [83]. Such operation can be viewed as a form of cooperation, in the sense that the terminals do not interfere with one another; however, it is not cooperation in the sense of cooperative diversity. It would be interesting to determine whether cooperative diversity with power control improves substantially over the case of power control alone, what the salient characteristics of effective algorithms are, and how they relate to the basic algorithms developed in this dissertation.

As another example, instead of using outage probability as a measure, in which a given fixed rate is either achieved or not, using channel superposition coding as in the classical broadcast channel allows for a range of rates to be achieved, and expected capacity becomes a relevant performance measure [16], assuming successive-refinement source coding techniques [62] are employed. This "broadcast approach" was first studied in the context of fading channels by Shamai [71]. In the context of cooperative diversity, superposition coding might lead to ways of having relays partially decode information from the source. For example, suppose the source encoded information into two-level superposition code. A relay not able to decode both levels might be able to decode the first level, strip it out, and amplify the second level. After suitable combining, the destination might receive both levels or only the first level. Superposition coding generally allows for larger expected capacity, and it would be interesting to see how cooperative diversity gains, if any, would reveal themselves in this context.

- **General Networks, Feed-Forward Graphs, and Network Coding:** Throughout the dissertation we have focused on multiple-access and to some extent interference channels with cooperative diversity because there is a natural way to tradeoff power

between the cooperating transmitters in these channels. As we pointed out in Chapter 2, cooperative diversity can arise in many other simple networks, such as broadcast networks. The broadcast network can be viewed as a generalization of the relay channel in which the relays also have information to receive. It would be useful to develop a framework for trading off system resources and performance in such networks.

More generally, in this dissertation we evaluated our algorithms for cooperative diversity in terms of essentially local transmissions in which the source and destination are determined by higher layer algorithms, such as a multihop routing algorithm. Our perspective has been to improve the performance on each of these links by having nearby terminals cooperate with each other. At a high level, grouping into cooperating sets and determining one or several routes through a wireless network could be performed jointly. An example in this area is the generalization of the relay channel decode-and-forward to communication of feed-forward graphs by Gupta and Kumar [33]. Another potentially useful framework for further study of cooperative diversity is the network coding approach of Koetter and Medard [47]. It seems clear that the basic building block of such algorithms should be something more general than a point-to-point link, but should not be something so complicated that the resulting algorithms become intractable. Much more research is necessary to obtain a better understanding of these issues.

- **Practical Coding and Decoding Algorithms:** Throughout the dissertation we have employed random coding arguments to evaluate performance of coded cooperative diversity schemes. In both uncoded and coded settings, different practical decoding algorithms at the relays as well as combining and decoding algorithms at the destination induce many possible communication strategies with related, but possible different code design criteria. We have performed some analysis of detectors in uncoded settings [49], and obtained similar performance advantages in terms of bit-error rate to those obtained here in terms of outage probability. Recent work on coded settings has also appeared [41]. More generally, designing effective algorithms, evaluating performance, and selecting codes are necessary for practical implementation of cooperative diversity.

Appendix A

Coding Theorem for Cooperative Diversity

This appendix proves a coding theorem for an achievable rate region for the multiple-access channel with cooperative diversity, the model described in Section 4.1. In particular, we focus on decode-and-forward transmission, in which the encoders decode each other's messages and cooperatively send refinement information to the ultimate decoder. We treat the discrete memoryless channel, with the usual extension to the Gaussian case following in the standard way by incorporating the power constraints into the conditions on the joint typicality at the decoder.

The proof utilizes a random coding argument with superposition block-Markov coding [18] and backward decoding [90]. It can be viewed as a two-user generalization of the cooperation strategy introduced by Cover and El Gamal [17] for the relay channel, or a simplified version of the strategy developed by Willems, van der Meulen, and Schalkwijk [91] for the multiple access channel with generalized feedback. Although backwards decoding incurs longer decoding delay than the list decoding techniques employed in, *e.g.*, [17], the proofs are conceptually simpler.

A.1 Definitions

For convenience, we repeat here the definitions of Section 4.1.4.

A (two-user) *discrete memoryless multiple-access channel with cooperative diversity* consists of the following:

- Channel input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, and channel output alphabets $\mathcal{Y}_0$, $\mathcal{Y}_1$, and $\mathcal{Y}_2$.

- Channel probability law mapping inputs to outputs

$$p_{y_0,y_1,y_2|x_1,x_2}(y_0,y_1,y_2|x_1,x_2) = \prod_{i=0}^{2} p_{y_i|x_1,x_2}(y_i|x_1,x_2) \ . \tag{A.1}$$

  By memoryless, we mean that the probability law for n consecutive uses of the channel is the product

$$p_{\mathbf{y}_0,\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}_1,\mathbf{x}_2}(\mathbf{y}_0,\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}_1,\mathbf{x}_2) = \prod_{k=1}^{n} p_{y_0[k],y_1[k],y_2[k]|x_1[k],x_2[k]}(y_0[k],y_1[k],y_2[k]|x_1[k],x_2[k]) \ . \tag{A.2}$$

A *communication strategy* for the discrete memoryless multiple-access channel with cooperative diversity consists of the following:

- Messages $w_i \in \mathcal{M}_i = \{1,2,\ldots,M_i\}$, $i = 1,2$, distributed uniformly and independently. The rates in bits per channel use are then

$$\frac{1}{n}\log_2 M_i \ , \qquad i = 1,2 \ . \tag{A.3}$$

- Encoding functions, one for each encoder for each time k, that map the encoder's message w_i and past channel observations $y_i[1],\ldots,y_i[k-1]$ into its transmitted signal. That is

$$x_i[k] = f_{i,k}(w_i, y_i[1], y_i[2], \ldots, y_i[k-1]) \ , \qquad i = 1,2 \ . \tag{A.4}$$

  Note that these encoders are *causal* functions of the respective channel outputs.

- A decoding function mapping the channel output vector $\mathbf{y}_0$ into $\mathcal{M}_1 \times \mathcal{M}_2$, *i.e.*,

$$(\hat{w}_1, \hat{w}_2) = g(\mathbf{y}_0) \ . \tag{A.5}$$

Definition 4 *The average probability of error for a communication strategy operating over a discrete memoryless multiple-access channel with cooperative diversity is*

$$P_e^{(n)} = \Pr\left[g(\mathbf{y}_0) \neq (w_1, w_2)\right] \ . \tag{A.6}$$

Definition 5 *A pair of transmission rates (R_1, R_2) is said to be achievable on a discrete memoryless multiple-access channel with cooperative diversity if there exists a sequence of communication strategies operating over the channel with*

$$M_i^{(n)} = 2^{\lceil nR_i \rceil} \ , \qquad i = 1, 2$$

and

$$P_e^{(n)} \to 0 \ , \qquad as\ n \to \infty$$

Definition 6 *The capacity region of a discrete memoryless multiple-access channel with cooperative diversity is the closure of the set of achievable rates.*

A.2 Preliminaries

The random coding argument to follow relies on properties of *jointly typical sequences* for a probability distribution. We summarize the relevant properties in this section. These definitions and properties are taken directly from [19, Section 14.2], where more details are also available.

Let $(\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_k)$ denote a finite collection of discrete random variables with joint probability distribution function

$$p_{\mathsf{x}_1,\mathsf{x}_2,\ldots,\mathsf{x}_k}(x_1, x_2, \ldots, x_k) \ .$$

Let s denote any ordered subset of these random variables, and consider n independent copies $\mathsf{s}_1, \mathsf{s}_2, \ldots, \mathsf{s}_n$ of s. Then

$$p_{\mathsf{s}_1,\mathsf{s}_2,\ldots,\mathsf{s}_n}(s_1, s_2, \ldots, s_n) = \prod_{i=1}^{n} p_{\mathsf{s}}(s_i) \ ,$$

and, by the weak law of large numbers, we have

$$-\frac{1}{n} \log p_{\mathsf{s}_1,\mathsf{s}_2,\ldots,\mathsf{s}_n}(\mathsf{s}_1, \mathsf{s}_2, \ldots, \mathsf{s}_n) = -\frac{1}{n} \sum_{i=1}^{n} \log p_{\mathsf{s}}(\mathsf{s}_i) \to H(\mathsf{s}) \ , \quad \text{w.p. } 1 \ .$$

Definition 7 *The set* $\mathcal{A}_\epsilon^{(n)}(x_1, \ldots, x_k)$ *of jointly* ϵ*-typical* n*-sequences* $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ *is defined by*

$$\mathcal{A}_\epsilon^{(n)}(x_1, \ldots, x_k) = \{(\mathbf{x}_1, \ldots, \mathbf{x}_k) : \left| \frac{1}{n} \log p_{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n}(s_1, s_2, \ldots, s_n) - H(\mathbf{s}) \right| < \epsilon, \forall \mathbf{s} \subseteq (x_1, x_2, \ldots, x_k)\} \ . \tag{A.7}$$

The following Lemma, often called the *joint asymptotic equipartition property (AEP)*, is proven in [19, Theorem 14.2.1], and captures all the properties we need for the achievability proof to follow.

Lemma 1 *For any* $\epsilon > 0$*, and for sufficiently large* n*,*

1. $\Pr\left[\mathcal{A}_\epsilon^{(n)}(\mathbf{s})\right] \geq 1 - \epsilon \ , \forall \mathbf{s} \subseteq (x_1, x_2, \ldots, x_k)$.
2. *If* $\mathbf{s} \in \mathcal{A}_\epsilon^{(n)}(\mathbf{s})$*, then*
$$2^{-n[H(\mathbf{s})+\epsilon]} \leq p(\mathbf{s}) \leq 2^{-n[H(\mathbf{s})-\epsilon]} \ .$$
3. $2^{n[H(\mathbf{s})-2\epsilon]} \leq |\mathcal{A}_\epsilon^{(n)}(\mathbf{s})| \leq 2^{n[H(\mathbf{s})+2\epsilon]}$.
4. *Let* $\mathbf{s}_1, \mathbf{s}_2 \subseteq (x_1, x_2, \ldots, x_k)$*. If* $(\mathbf{s}_1, \mathbf{s}_2) \in \mathcal{A}_\epsilon^{(n)}(\mathbf{s}_1, \mathbf{s}_2)$*, then*
$$2^{-n[H(\mathbf{s}_1|\mathbf{s}_2)+2\epsilon]} \leq p(\mathbf{s}_1|\mathbf{s}_2) \leq 2^{-n[H(\mathbf{s}_1|\mathbf{s}_2)-2\epsilon]} \ .$$

A.3 Decode-and-Forward Transmission

We now prove a coding theorem for decode-and-forward transmission. In this communication strategy, the cooperating users decode each other's messages and jointly transmit "refinement" information for the destination's benefit. Specifically, we prove the following theorem.

Theorem 4 *The set of achievable rates for decode-and-forward transmission over a discrete-memoryless multiple-access channel with cooperative diversity is given by the closure of the*

convex hull of all (R_1, R_2) *satisfying*

$$R_1 < I(x_1; y_2|x_2, u) \ , \tag{A.8}$$

$$R_2 < I(x_2; y_1|x_1, u) \ , \tag{A.9}$$

$$R_1 + R_2 < I(x_1, x_2; y) \ , \tag{A.10}$$

for some distribution $p_u(u)p_{x_1|u}(x_1|u)p_{x_2|u}(x_2|u)$ *on* $\mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2$.

As is frequently done in random coding arguments, we break the proof into the following steps: generation of random codebooks, encoding and decoding, and analysis of the probability of error.

A.3.1 Codebook Generation

Fix a blocklength n. Let $\mathcal{M}_i = \{1, 2, \ldots, M_i\}$, with $M_i = 2^{\lceil nR_i \rceil}$, $i = 1, 2$. Suppose $w_i \in \mathcal{M}_i$, $i = 1, 2$, and $w_0 \in \mathcal{M}_1 \times \mathcal{M}_2$. Throughout the proof, w_1 (resp. w_2) indexes "fresh" information at encoder 1 (resp. encoder 2), and w_0 indexes "refinement" information at both encoders.

We fix a distribution $p_u(u)p_{x_1|u}(x_1|u)p_{x_2|u}(x_2|u)$. Note that x_1 and x_2 are conditionally independent given u under this distribution, or, equivalently, $x_1 \leftrightarrow u \leftrightarrow x_2$ forms a Markov chain. Also note that $u \leftrightarrow (x_1, x_2) \leftrightarrow (y, y_1, y_2)$ forms a Markov chain.

The random codebooks are generated according to the following steps:

- Generate $M_1 \cdot M_2$ codewords $\mathbf{u}(w_0)$ i.i.d.according to $p_{\mathbf{u}}(\mathbf{u}) = \prod_{i=1}^{n} p_u(u_i)$. As we will see, $\mathbf{u}(w_0)$, $w_0 \in \mathcal{M}_1 \times \mathcal{M}_2$, communicates refinement information in each block.

- For each $\mathbf{u}(w_0)$, generate M_1 codewords $\mathbf{x}_1(w_1, w_0)$ i.i.d.according to

$$p_{\mathbf{x}_1|\mathbf{u}}(\mathbf{x}_1|\mathbf{u}(w_0)) = \prod_{i=1}^{n} p_{x_1|u}(x_{1,i}|u_i(w_0)) \ .$$

  As we will see, $\mathbf{x}_1(w_1, w_0)$, with $w_1 \in \mathcal{M}_1$ and w_0 fixed, communicates fresh information from encoder 1 in each block.

- Similarly, for each $\mathbf{u}(w_0)$, generate M_2 codewords $\mathbf{x}_2(w_2, w_0)$ i.i.d.according to

$$p_{\mathbf{x}_2|\mathbf{u}}(\mathbf{x}_2|\mathbf{u}(w_0)) = \prod_{i=1}^{n} p_{x_2|u}(x_{2,i}|u_i(w_0)) \ .$$

Block 1	Block 2	$\cdots$	Block $B-1$	Block B
$\mathbf{x}_1(w_{1,1},(1,1))$	$\mathbf{x}_1(w_{1,2},(w_{1,1},\hat{\hat{w}}_{2,1}))$	$\cdots$	$\mathbf{x}_1(w_{1,B-1},(w_{1,B-2},\hat{\hat{w}}_{2,B-2}))$	$\mathbf{x}_1(1,(w_{1,B-1},\hat{\hat{w}}_{2,B-1}))$
$\mathbf{x}_2(w_{2,1},(1,1))$	$\mathbf{x}_2(w_{2,2},(\hat{\hat{w}}_{1,1},w_{2,1}))$	$\cdots$	$\mathbf{x}_2(w_{2,B-1},(\hat{\hat{w}}_{1,B-2},w_{2,B-2}))$	$\mathbf{x}_2(1,(\hat{\hat{w}}_{1,B-1},w_{2,B-1}))$

Figure A-1: Block-Markov encoding structure for decode-and-forward transmission.

As we will see, $\mathbf{x}_2(w_2, w_0)$, with $w_2 \in \mathcal{M}_2$ and w_0 fixed, communicates fresh information from encoder 2 in each block.

A.3.2 Encoding

As in many superposition block-Markov encoding strategies, we encode information into B blocks each of length n channel uses. The same codebooks of length n are used in each of the B blocks. Fig. A-1 shows what the encoders send in the respective blocks. For block $b = 2, \ldots, B-1$, the encoders send fresh information from block b superimposed on refinement information from block $b-1$. In block $b = 1$, the encoders only send fresh information, while in block $b = B$, the encoders only send refinement information. Note that the actual transmission rates $\tilde{R}_i = R_i(B-1)/B$ approach R_i as $B \to \infty$.

To determine the refinement information for block b, each encoder must estimate the other's fresh information from block $b-1$. These estimates are denoted in Fig. A-1 by $\hat{\hat{w}}_{2,b}$ and $\hat{\hat{w}}_{1,b}$ at encoders 1 and 2, respectively. If these estimates contain no errors, *i.e.*, $\hat{\hat{w}}_{1,b} = w_{1,b}$ and $\hat{\hat{w}}_{2,b} = w_{2,b}$, then the separate encoders each superimpose their fresh information onto the same refinement information $\mathbf{u}((w_{1,b}, w_{2,b}))$. It is in this sense that the two encoders cooperate to refine the decoder's estimate of the fresh information in the previous block.

To determine these estimates, each encoder looks for a sequence that is jointly typical with its emission and channel observations. For example, in block 1, knowing $w_{1,1}$ and that $w_{1,0} = w_{2,0} = 1$, encoder 1 looks for a unique $w_2 \in \mathcal{M}_2$ such that the event

$$(\mathbf{u}((1,1)), \mathbf{x}_1(w_{1,1},(1,1)), \mathbf{x}_2(w_2,(1,1)), \mathbf{y}_1) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y_1) \tag{A.11}$$

occurs. If no such w_2 exists, or more than one exists, then the encoder 1 declares an error. Otherwise, it sets its estimate $\hat{\hat{w}}_{2,1}$ to the unique w_2 satisfying (A.11). Similarly, knowing

$w_{2,1}$ and that $w_{1,0} = w_{2,0} = 1$, encoder 2 looks for a unique $w_1 \in \mathcal{M}_1$ such that the event

$$(\mathbf{u}((1,1)), \mathbf{x}_1(w_1,(1,1)), \mathbf{x}_2(w_{2,1},(1,1)), \mathbf{y}_2) \in A_\epsilon^{(n)}(u, x_1, x_2, y_2) \tag{A.12}$$

occurs. If no such w_1 exists, or more than one exists, then the encoder 2 declares an error. Otherwise, it sets its estimate $\hat{\hat{w}}_{1,1}$ to the unique w_1 satisfying (A.12).

In each of blocks $b = 2, 3, \ldots, B-1$, knowing $w_{1,b}$, $w_{1,b-1}$, encoder 1 looks for a unique $w_2 \in \mathcal{M}_2$ such that the event

$$(\mathbf{u}((w_{1,b-1}, \hat{\hat{w}}_{2,b-1})), \mathbf{x}_1(w_{1,b}, (w_{1,b-1}, \hat{\hat{w}}_{2,b-1})), \mathbf{x}_2(w_2, (w_{1,b-1}, \hat{\hat{w}}_{2,b-1})), \mathbf{y}_1) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y_1) \tag{A.13}$$

occurs. If no such w_2 exists, or more than one exists, then encoder 1 declares an error. Otherwise, it sets its estimate $\hat{\hat{w}}_{2,b}$ to the unique w_2 satisfying (A.13). Similarly, knowing $w_{2,b}$, $w_{2,b-1}$, encoder 2 looks for a unique $w_1 \in \mathcal{M}_1$ such that the event

$$(\mathbf{u}((\hat{\hat{w}}_{1,b-1}, w_{2,b-1})), \mathbf{x}_1(w_1, (\hat{\hat{w}}_{1,b-1}, w_{2,b-1})), \mathbf{x}_2(w_{2,b}, (\hat{\hat{w}}_{1,b-1}, w_{2,b-1})), \mathbf{y}_2) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y_2) \tag{A.14}$$

occurs. If no such w_1 exists, or more than one exists, then encoder 2 declares an error. Otherwise, it sets its estimate $\hat{\hat{w}}_{1,b}$ to the unique w_1 satisfying (A.14).

A.3.3 Backwards Decoding

Backwards decoding at the decoder operates as follows. The decoder waits for all B blocks to be received, and then begins decoding with block B. It decodes the refinement information in block b given knowledge of the fresh information in block b. The refinement information in block b determines the fresh information in block $b-1$, and the process continues.

In block B, knowing that $w_{1,B} = w_{2,B} = 1$, the decoder looks for a unique $w_0 \in \mathcal{M}_1 \times \mathcal{M}_2$ such that the event

$$(\mathbf{u}(w_0), \mathbf{x}_1(1, w_0), \mathbf{x}_2(1, w_0), \mathbf{y}) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y) \tag{A.15}$$

occurs. If no such w_0 exists, or more than one exists, then the decoder declares an error. If such a w_0 exists, then its components are the decoder estimates of the messages in block $B-1$, *i.e.*, $(\hat{w}_{1,B-1}, \hat{w}_{2,B-1}) = w_0$.

Similarly, in block $b = B-1, B-2, \ldots, 2$, with estimates $\hat{w}_{1,b}$ and $\hat{w}_{2,b}$ obtained from decoding refinement information in block $b+1$, the decoder looks for a unique $w_0 \in \mathcal{M}_1 \times \mathcal{M}_2$ such that the event

$$(\mathbf{u}(w_0), \mathbf{x}_1(\hat{w}_{1,b}, w_0), \mathbf{x}_2(\hat{w}_{2,b}, w_0), \mathbf{y}) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y) \tag{A.16}$$

occurs. If no such w_0 exists, or more than one exists, then the decoder declares an error. If such a w_0 exists, then its components are the estimates of the messages in block $b-1$, *i.e.*, $(\hat{w}_{1,b-1}, \hat{w}_{2,b-1}) = w_0$.

We note that, under backwards decoding, the fresh signals $\mathbf{x}_i$, $i = 1, 2$, are mainly used to share information among the encoders. The decoder recovers the message exclusively from $\mathbf{u}$. By contrast, under list decoding as employed in [17], both $\mathbf{x}_i$, $i = 1, 2$, and $\mathbf{u}$ are utilized to recover the message at the decoder.

A.3.4 Probability of Error and Achievable Rates

In this section, we examine the probability of error in the sequence of B blocks. The analysis is somewhat complicated by the fact that the encoders operate in the forward direction, while the decoder operates in the backward direction. Nevertheless, backward decoding allows us to avoid the more involved list decoding and its associated analysis.

An outline of the proof is as follows. We show that the error probability can be made arbitrarily small given certain conditions on the transmitted rates. To bound the overall probability of error at the decoder, we first separate the event that the encoders make errors from the event that the decoder makes error given perfect encoders. We then break the error events for all B blocks into a union of simpler, conditional error events in the individual blocks. Finally, fixing the transmission rates so that they fall below the appropriate mutual informations, we show that the error events in the individual blocks have arbitrarily small probability of error.

Sequence Errors to Block Errors

Let $\mathcal{E} = \{(\hat{\hat{\mathbf{w}}}_1, \hat{\hat{\mathbf{w}}}_2) \neq (\mathbf{w}_1, \mathbf{w}_2)\}$, the event that either encoder makes an error in estimating the other's fresh information during any of the B blocks. Similarly, let $\mathcal{D} = \{(\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2) \neq (\mathbf{w}_1, \mathbf{w}_2)\}$, the event that the decoder makes an error in estimating the encoder messages

during any of the B blocks. Note that these error events are defined not only over the uniform distribution for the messages, but also the random codebooks.

Using the total probability law, and the fact that probabilities are bounded by 1, we have

$$\begin{aligned}\Pr[\mathcal{D}] &= \Pr[\mathcal{D}|\mathcal{E}]\Pr[\mathcal{E}] + \Pr\left[\mathcal{D}|\overline{\mathcal{E}}\right]\Pr\left[\overline{\mathcal{E}}\right] \ , \\ &\leq \Pr[\mathcal{E}] + \Pr\left[\mathcal{D}|\overline{\mathcal{E}}\right] \ . \end{aligned} \tag{A.17}$$

Conveniently, (A.17) allows us to separately examine the error probability of the encoders and the error probability of the decoders assuming no encoder errors. To show that, for any $\epsilon > 0$, $\Pr[\mathcal{D}] \leq \epsilon$, we will show that $\Pr[\mathcal{E}] \leq \epsilon/2$ and $\Pr\left[\mathcal{D}|\overline{\mathcal{E}}\right] \leq \epsilon/2$ for n sufficiently large.

Let $\mathcal{E}_b = \{(\hat{\hat{w}}_{1,b}, \hat{\hat{w}}_{2,b}) \neq (w_{1,b}, w_{2,b})\}$, $b = 1, 2, \ldots, B-1$. Then $\mathcal{E}_b$ is the event that either encoder makes a decoding error in block b. Using basic properties of sets and probabilities, we have

$$\begin{aligned}\Pr[\mathcal{E}] &= \Pr\left[\bigcup_{b=1}^{B-1} \mathcal{E}_b\right] \\ &= \Pr\left[\bigcup_{b=1}^{B-1} \left\{\mathcal{E}_b - \bigcup_{b'=1}^{b-1} \mathcal{E}_{b'}\right\}\right] \\ &= \Pr\left[\bigcup_{b=1}^{B-1} \mathcal{E}_b \cap \overline{\mathcal{E}}_1 \cap \ldots \cap \overline{\mathcal{E}}_{b-1}\right] \\ &= \sum_{b=1}^{B-1} \Pr\left[\mathcal{E}_b \cap \overline{\mathcal{E}}_1 \cap \ldots \cap \overline{\mathcal{E}}_{b-1}\right] \\ &\leq \sum_{b=1}^{B-1} \Pr\left[\mathcal{E}_b \cap \overline{\mathcal{E}}_{b-1}\right] \\ &\leq \sum_{b=1}^{B-1} \Pr\left[\mathcal{E}_b \,\middle|\, \overline{\mathcal{E}}_{b-1}\right] \ . \end{aligned} \tag{A.18}$$

Thus, to show that $\Pr[\mathcal{E}] \leq \epsilon/2$, it is sufficient to show that $\Pr\left[\mathcal{E}_b \,\middle|\, \overline{\mathcal{E}}_{b-1}\right] \leq \epsilon/(2B)$, as we develop in Section A.3.4

Let $\mathcal{D}_b = \{(\hat{w}_{1,b-1}, \hat{w}_{2,b-1}) \neq (w_{1,b-1}, w_{2,b-1})\}$ be the event that the decoder makes an error in block $b = 2, 3, \ldots, B$. Because of the backwards decoding, we define the reversed

block index $b' = -b + B + 1$. As in (A.18), we have

$$\begin{aligned}\Pr\left[\mathcal{D}\left|\overline{\mathcal{E}}\right.\right] &= \Pr\left[\bigcup_{b'=1}^{B-1}\mathcal{D}_{b'}\middle|\overline{\mathcal{E}}\right] \\ &\leq \sum_{b'=1}^{B-1}\Pr\left[\mathcal{D}_{b'}\left|\overline{\mathcal{D}}_{b'-1},\overline{\mathcal{E}}\right.\right] \\ &= \sum_{b=2}^{B-1}\Pr\left[\mathcal{D}_b\left|\overline{\mathcal{D}}_{b+1},\overline{\mathcal{E}}\right.\right] + \Pr\left[\mathcal{D}_B\left|\overline{\mathcal{E}}\right.\right] \ .\end{aligned}$$

Thus, to show that $\Pr\left[\mathcal{D}\left|\overline{\mathcal{E}}\right.\right] \leq \epsilon/2$, it is sufficient to show that $\Pr\left[\mathcal{D}_b\left|\overline{\mathcal{D}}_{b+1},\overline{\mathcal{E}}\right.\right] \leq \epsilon/(2B)$, as we develop in Section A.3.4.

Encoder Block Error Probability

Let $\mathcal{E}_{b,i} = \{\hat{\hat{w}}_{i,b} \neq w_{i,b}\}$, $i = 1, 2$. Then

$$\begin{aligned}\Pr\left[\mathcal{E}_b\left|\overline{\mathcal{E}}_{b-1}\right.\right] &= \Pr\left[\mathcal{E}_{b,1}\cup\mathcal{E}_{b,2}\left|\overline{\mathcal{E}}_{b-1}\right.\right] \\ &\leq \Pr\left[\mathcal{E}_{b,1}\left|\overline{\mathcal{E}}_{b-1}\right.\right] + \Pr\left[\mathcal{E}_{b,2}\left|\overline{\mathcal{E}}_{b-1}\right.\right]\end{aligned}$$

Our objective in this section is to show that, for any $\epsilon > 0$, $\Pr\left[\mathcal{E}_{b,i}\left|\overline{E}_{b-1}\right.\right] \leq \epsilon/(4B)$ for n sufficiently large given certain conditions on the transmitted rates.

We now upper bound $\Pr\left[\mathcal{E}_{b,2}\left|\overline{\mathcal{E}}_{b-1}\right.\right]$, for $b = 1, 2, \ldots, B-1$, and note that a similar argument applies to $\Pr\left[\mathcal{E}_{b,1}\left|\overline{\mathcal{E}}_{b-1}\right.\right]$. Given event $\overline{\mathcal{E}}_{b-1}$, the pair of encoders have correct estimates, *i.e.*, $\hat{\hat{w}}_{i,b-1} = w_{i,b-1}$, $i = 1, 2$, allowing them to select identical refinement information $\mathbf{u}(w_{1,b-1}, w_{2,b-1})$ in block b (*cf.* Fig. A-1).

As is often done in random coding arguments [19], we exploit the symmetry of the random codebook to obtain

$$\Pr\left[\mathcal{E}_{b,2}\left|\overline{\mathcal{E}}_{b-1}\right.\right] = \Pr\left[\mathcal{E}_{b,2}\left|\overline{\mathcal{E}}_{b-1}, w_{1,b-1} = w_{1,b} = w_{2,b-1} = w_{2,b} = 1\right.\right] \ .$$

That is, we may, without loss of generality, consider the conditional probability space in which both current and previous messages take the value 1. With this in mind, for $w_2 \in \mathcal{M}_2$, we define the events

$$\mathcal{T}_1(w_2) = \left\{(\mathbf{u}((1,1)), \mathbf{x}_1(1,(1,1)), \mathbf{x}_2(w_2,(1,1)), \mathbf{y}) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y_1)\right\} \ ,$$

so that the joint typicality decoder (A.11) and (A.13) has conditional error event

$$\overline{\mathcal{T}}_1(1) \cup \bigcup_{w_2=2}^{M_2} \mathcal{T}_1(w_2) \ ,$$

i.e., either the codewords corresponding to message 1 are not jointly typical with the channel output, or the codewords corresponding to some other message are jointly typical with the channel output. The conditional error probability then becomes

$$\begin{aligned} \Pr\left[\mathcal{E}_{b,1} \middle| \overline{\mathcal{E}}_{b-1}\right] &= \Pr\left[\overline{\mathcal{T}}_1(1) \cup \bigcup_{w_2=2}^{M_2} \mathcal{T}_1(w_2)\right] \\ &\leq \Pr\left[\overline{\mathcal{T}}_1(1)\right] + \sum_{w_2=2}^{M_2} \Pr\left[\mathcal{T}_1(w_2)\right] \ . \end{aligned} \tag{A.19}$$

By the joint AEP, $\Pr\left[\overline{\mathcal{T}}_1(1)\right] \to 0$ as $n \to \infty$. For each of the terms in the summation, with $w_2 \neq 1$, we have

$$\begin{aligned} \Pr\left[\mathcal{T}_1(w_2)\right] &= \Pr\left[(\mathbf{u}((1,1)), \mathbf{x}_1(1,(1,1)), \mathbf{x}_2(w_2,(1,1)), \mathbf{y}_1) \in \mathcal{A}_\epsilon^{(n)}(u, x_1, x_2, y_1)\right] \\ &= \sum_{(\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1) \in \mathcal{A}_\epsilon^{(n)}(u,x_1,x_2,y_1)} p_{\mathbf{u}}(\mathbf{u}) p_{\mathbf{x}_1|\mathbf{u}}(\mathbf{x}_1|\mathbf{u}) p_{\mathbf{x}_2|\mathbf{u}}(\mathbf{x}_2|\mathbf{u}) p_{\mathbf{y}_1|\mathbf{u},\mathbf{x}_1}(\mathbf{y}_1|\mathbf{u},\mathbf{x}_1) \\ &\leq \sum_{(\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1) \in \mathcal{A}_\epsilon^{(n)}(u,x_1,x_2,y_1)} 2^{-n[H(u)-\epsilon]} \cdot 2^{-n[H(x_1|u)-2\epsilon]} \cdot 2^{-n[H(x_2|u)-2\epsilon]} \cdot 2^{-n[H(y_1|u,x_1)-2\epsilon]} \\ &\leq 2^{n[H(u,x_1,x_2,y_1)+2\epsilon]} \cdot 2^{-n[H(u)-\epsilon]} \cdot 2^{-n[H(x_1|u)-2\epsilon]} \cdot 2^{-n[H(x_2|u)-2\epsilon]} \cdot 2^{-n[H(y_1|u,x_1)-2\epsilon]} \\ &\leq 2^{-n[I(x_2;y_1|u,x_1)-9\epsilon]} \ . \end{aligned} \tag{A.20}$$

Substituting (A.20) into (A.19), we obtain

$$\Pr\left[\mathcal{E}_{b,2} \middle| \overline{\mathcal{E}}_{b-1}\right] \leq \Pr\left[\overline{\mathcal{T}}_1(1)\right] + M_2 \cdot 2^{-n(I(x_2;y_1|u,x_1)-9\epsilon)} \tag{A.21}$$

$$\leq \Pr\left[\overline{\mathcal{T}}_1(1)\right] + 2 \cdot 2^{nR_2} 2^{-n[I(x_2;y_1|u,x_1)-9\epsilon]} \tag{A.22}$$

$$= \Pr\left[\overline{\mathcal{T}}_1(1)\right] + 2 \cdot 2^{-n[I(x_2;y_1|u,x_1)-9\epsilon-R_2]} \tag{A.23}$$

$$\to 0 \tag{A.24}$$

as $n \to \infty$ if $R_2 + 9\epsilon < I(x_2; y_1|u, x_1)$.

Similarly, $\Pr\left[\mathcal{E}_{b,1} \middle| \overline{\mathcal{E}}_{b-1}\right] \to 0$ as $n \to \infty$ if $R_1 + 9\epsilon < I(x_1; y_2|u, x_2)$.

Decoder Block Error Probability

Our objective in this section is to show that, for any $\epsilon > 0$, $\Pr\left[\mathcal{D}_b \left| \overline{\mathcal{D}}_{b+1}, \overline{\mathcal{E}} \right.\right] \leq \epsilon/(2B)$ for n sufficiently large given certain conditions on the transmitted rates.

Again, by symmetry of the random code construction,

$$\Pr\left[\mathcal{D}_b \left| \overline{\mathcal{D}}_{b+1}, \overline{\mathcal{E}} \right.\right] = \Pr\left[\mathcal{D}_b \left| \overline{\mathcal{D}}_{b+1}, \overline{\mathcal{E}}, \mathsf{w}_{1,b-1} = \mathsf{w}_{1,b} = \mathsf{w}_{2,b-1} = \mathsf{w}_{2,b} = 1 \right.\right] . \tag{A.25}$$

As we did with the encoders, for $w_1 \in \mathcal{M}_1$, $w_2 \in \mathcal{M}_2$, we define the events

$$\mathcal{T}(w_1, w_2) = \left\{ (\mathbf{u}((w_1, w_2)), \mathbf{x}_1(1, (w_1, w_2)), \mathbf{x}_2(1, (w_1, w_2)), \mathbf{y}) \in \mathcal{A}_\epsilon^{(n)}(\mathsf{u}, \mathsf{x}_1, \mathsf{x}_2, \mathsf{y}) \right\} , \tag{A.26}$$

so that the joint typicality decoder (A.15) and (A.16) has conditional error event

$$\overline{\mathcal{T}}(1,1) \cup \bigcup_{w_1=2}^{M_1} \mathcal{T}(w_1, 1) \cup \bigcup_{w_2=2}^{M_2} \mathcal{T}(1, w_2) \cup \bigcup_{w_1=2}^{M_1} \bigcup_{w_2=2}^{M_2} \mathcal{T}(w_1, w_2) ,$$

i.e., either the codewords corresponding to the message $(1, 1)$ are not jointly typical with the channel output, or the codewords corresponding to some other message are jointly typical with the channel output. Then by the union of events bound for probabilities,

$$\begin{aligned} \Pr\left[\mathcal{D}_b \left| \overline{\mathcal{D}}_{b+1}, \overline{\mathcal{E}} \right.\right] &= \Pr\left[\overline{\mathcal{T}}(1,1) \cup \bigcup_{w_1=2}^{M_1} \mathcal{T}(w_1, 1) \cup \bigcup_{w_2=2}^{M_2} \mathcal{T}(1, w_2) \cup \bigcup_{w_1=2}^{M_1} \bigcup_{w_2=2}^{M_2} \mathcal{T}(w_1, w_2) \right] \\ &\leq \Pr\left[\overline{\mathcal{T}}(1,1)\right] + \sum_{w_1=2}^{M_1} \Pr\left[\mathcal{T}(w_1, 1)\right] + \sum_{w_2=2}^{M_2} \Pr\left[\mathcal{T}(1, w_2)\right] \\ &\qquad + \sum_{w_1=2}^{M_1} \sum_{w_2=2}^{M_2} \Pr\left[\mathcal{T}(w_1, w_2)\right] . \end{aligned} \tag{A.27}$$

By the joint AEP, $\Pr\left[\overline{\mathcal{T}}(1,1)\right] \to 0$ as $n \to \infty$. For the terms in each of the sums, with

$w_1 \neq 1$ or $w_2 \neq 1$, we have

$$\begin{aligned}
\Pr\left[\mathcal{T}(w_1,w_2)\right] &= \Pr\left[(\mathbf{u}((w_1,w_2)),\mathbf{x}_1(1,(w_1,w_2)),\mathbf{x}_2(1,(w_1,w_2)),\mathbf{y}) \in \mathcal{A}_\epsilon^{(n)}(u,x_1,x_2,y)\right] \\
&= \sum_{(\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{y})\in\mathcal{A}_\epsilon^{(n)}(u,x_1,x_2,y)} p_{\mathbf{u}}(\mathbf{u})p_{\mathbf{x}_1|\mathbf{u}}(\mathbf{x}_1|\mathbf{u})p_{\mathbf{x}_2|\mathbf{u}}(\mathbf{x}_2|\mathbf{u})p_{\mathbf{y}}(\mathbf{y}) \\
&\leq \sum_{(\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{y})\in\mathcal{A}_\epsilon^{(n)}(u,x_1,x_2,y)} 2^{-n[H(u)-\epsilon]}.2^{-n[H(x_1|u)-2\epsilon]}.2^{-n[H(x_2|u)-2\epsilon]}.2^{-n[H(y)-\epsilon]} \\
&\leq 2^{n[H(u,x_1,x_2,y)+2\epsilon]}.2^{-n[H(u)-\epsilon]}.2^{-n[H(x_1|u)-2\epsilon]}.2^{-n[H(x_2|u)-2\epsilon]}.2^{-n[H(y)-\epsilon]} \qquad \text{(A.28)} \\
&= 2^{-n[I(x_1,x_2;y)-8\epsilon]} \qquad \text{(A.29)}
\end{aligned}$$

Substituting (A.29) into (A.27) and collecting terms, we have

$$\begin{aligned}
\Pr\left[\mathcal{D}_b \,\middle|\, \overline{\mathcal{D}}_{b+1}, \overline{\mathcal{E}}\right] &\leq \Pr\left[\overline{T}(1,1)\right] + 2\cdot M_1 \cdot M_2 \cdot 2^{-n[I(x_1,x_2;y)-8\epsilon]} \\
&\leq \Pr\left[\overline{T}(1,1)\right] + 8\cdot 2^{-n[R_1+R_2]}2^{-n[I(x_1,x_2;y)-8\epsilon]} \\
&\leq \Pr\left[\overline{T}(1,1)\right] + 8\cdot 2^{-n[I(x_1,x_2;y)-8\epsilon-(R_1+R_2)]} \\
&\to 0 \qquad \text{(A.30)}
\end{aligned}$$

as $n \to \infty$ if $R_1 + R_2 + 8\epsilon < I(x_1, x_2; y)$.

Appendix B

Asymptotic CDF Approximations

B.1 Results for Chapter 5

To keep the presentation in the Chapter 5 concise, in this appendix we collect several results for the limiting behavior of the cumulative distribution function (CDF) of certain combinations of exponential random variables. All our results are of the form

$$\lim_{t \to t_0} \frac{P_{\mathsf{u}(t)}(g_1(t))}{g_2(t)} = c \tag{B.1}$$

where: t is a parameter of interest; $P_{\mathsf{u}(t)}(g_1(t))$ is the CDF[1] of a certain random variable $\mathsf{u}(t)$ that can, in general, depend upon t; $g_1(t)$ and $g_2(t)$ are two (continuous) functions; and t_0 and c are constants. Among other things, for example, (B.1) implies the approximation $P_{\mathsf{u}(t)}(g_1(t)) \sim c g_2(t)$ is accurate for t close to t_0. For example, in the case of approximations for large channel SNR, *i.e.*, $\mathtt{SNR} \to \infty$, we identify t with $\mathtt{SNR}$ and t_0 with ∞; similarly, in the case of approximations for low spectral efficiency, *i.e.*, $\mathtt{R} \to 0$, we identify t with $\mathtt{R}$ and t_0 with 0.

Recall that an exponential random variable u with parameter λ_{u} has probability density function (PDF)

$$p_{\mathsf{u}}(u) = \begin{cases} \lambda_{\mathsf{u}}\, e^{-\lambda_{\mathsf{u}} u} & u > 0 \\ 0 & u \leq 0 \end{cases}, \tag{B.2}$$

[1] Given random variable u, we denote its cumulative distribution function (CDF) by the function $P_{\mathsf{u}}(\cdot)$ and its probability density function (PDF) by $p_{\mathsf{u}}(\cdot)$. The two are related by $P_{\mathsf{u}}(u) = \Pr[\mathsf{u} \leq u] = \int_{-\infty}^{u^+} p_{\mathsf{u}}(v) dv$.

expectation $\mathrm{E}[u] = 1/\lambda_u$.

Fact 1 *Let u be an exponential random variable with parameter λ_u. Then the CDF $P_u(u) = 1 - e^{-\lambda_u u}$ satisfies*

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} P_u(\epsilon) = \lambda_u \ . \tag{B.3}$$

Moreover, if a function $g(t)$ is continuous about $t = t_0$ and satisfies $g(t) \to 0$ as $t \to t_0$, then

$$\lim_{t \to t_0} \frac{1}{g(t)} P_u(g(t)) = \lambda_u \ . \tag{B.4}$$

Fact 2 *Let $w = u + v$, where u and v are independent exponential random variables with parameters λ_u and λ_v, respectively. Then the CDF*

$$P_w(w) = \begin{cases} 1 - \left[\left(\frac{\lambda_v}{\lambda_v - \lambda_u} \right) e^{-\lambda_u w} + \left(\frac{\lambda_u}{\lambda_u - \lambda_v} \right) e^{-\lambda_v w} \right] & \lambda_u \neq \lambda_v \\ 1 - (1 + \lambda w) e^{-\lambda w} & \lambda_u = \lambda_v = \lambda \end{cases} \tag{B.5}$$

satisfies

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^2} P_w(\epsilon) = \frac{\lambda_u \lambda_v}{2} \ . \tag{B.6}$$

Moreover, if a function $g(t)$ is continuous about $t = t_0$ and satisfies $g(t) \to 0$ as $t \to t_0$, then

$$\lim_{t \to t_0} \frac{1}{g^2(t)} P_w(g(t)) = \frac{\lambda_u \lambda_v}{2} \ . \tag{B.7}$$

Claim 1 *Let u, v, and w be independent exponential random variables with parameters λ_u, λ_v, and λ_w, respectively. Let $f(x, y) = (xy)/(x + y + 1)$ as in (5.13). Let ϵ be positive, and let $g(\epsilon) > 0$ be continuous with $g(\epsilon) \to 0$ and $\epsilon / g(\epsilon) \to c < \infty$ as $\epsilon \to 0$. Then*

$$\lim_{\epsilon \to 0} \frac{1}{g^2(\epsilon)} \Pr\left[u + \epsilon \, f\left(v/\epsilon, \, w/\epsilon \right) < g(\epsilon) \right] = \frac{\lambda_u \left(\lambda_v + \lambda_w \right)}{2} \ . \tag{B.8}$$

Moreover, if a function $h(t)$ is continuous about $t = t_0$ and satisfies $h(t) \to 0$ as $t \to t_0$, then

$$\lim_{t \to t_0} \frac{1}{g^2(h(t))} \Pr\left[u + h(t) \, f\left(v/h(t), \, w/h(t) \right) < g(h(t)) \right] = \frac{\lambda_u (\lambda_v + \lambda_w)}{2} \ . \tag{B.9}$$

The following lemma will be useful in the proof of Claim 1.

Lemma 2 *Let δ be positive, and let $r_\delta \stackrel{\triangle}{=} \delta\, f(v/\delta, w/\delta)$, where v and w are independent exponential random variables with parameters λ_v and λ_w, respectively. Let $h(\delta) > 0$ be continuous with $h(\delta) \to 0$ and $\delta/h(\delta) \to d < \infty$ as $\delta \to 0$. Then the probability $\Pr[r_\delta < \delta]$ satisfies*

$$\lim_{\delta \to 0} \frac{1}{h(\delta)} \Pr[r_\delta < h(\delta)] = \lambda_v + \lambda_w. \tag{B.10}$$

Proof: (Of Lemma 2) The proof of this result is not as straightforward as that of the previous facts. Without assuming the limit exists, we upper bound the lim sup and lower bound the lim inf. When the bounds are the same, we can immediately conclude that the lim inf and lim sup are equal (because, in general, lim inf $\leq$ lim sup); hence, the limit exists and corresponds to the value of the identical bounds.

Starting with the lower bound,

$$\begin{aligned}
&\Pr[r_\delta < h(\delta)] \\
&\quad = \Pr[1/v + 1/w + \delta/(vw) \geq 1/h(\delta)] \\
&\quad \geq \Pr[1/v + 1/w \geq 1/h(\delta)] \\
&\quad \geq \Pr[\max(1/v, 1/w) \geq 1/h(\delta)] \\
&\quad = 1 - \Pr[v \geq h(\delta)] \Pr[w \geq h(\delta)] \\
&\quad = 1 - \exp[-\lambda_v h(\delta)] \exp[-\lambda_w h(\delta)] \\
&\quad = 1 - \exp[-(\lambda_v + \lambda_w) h(\delta)]
\end{aligned} \tag{B.11}$$

so, utilizing Fact 1,

$$\liminf_{\delta \to 0} \frac{1}{h(\delta)} \Pr[r_\delta < h(\delta)] \geq \lambda_v + \lambda_w\ . \tag{B.12}$$

To prove the other direction, let $l > 1$ be a fixed constant.

$$\begin{aligned}
&\Pr\left[r_\delta < h(\delta)\right] \\
&\quad = \Pr\left[1/v + 1/w + \delta/(vw) \geq 1/h(\delta)\right] \\
&\quad = \int_0^\infty \Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] p_w(w)dw \\
&\quad \leq \Pr\left[w < lh(\delta)\right] + \int_{lh(\delta)}^\infty \Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] p_w(w)dw \quad \text{(B.13)}
\end{aligned}$$

But

$$\Pr\left[w < lh(\delta)\right]/h(\delta) \leq \lambda_w l \ , \quad \text{(B.14)}$$

which takes care of the first term of (B.13). To bound the second term of (B.13), let $k > l$ be another fixed constant, and note that

$$\begin{aligned}
&\int_{lh(\delta)}^\infty \Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] p_w(w)dw \\
&\quad = \int_{kh(\delta)}^\infty \Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] p_w(w)dw \\
&\qquad + \int_{lh(\delta)}^{kh(\delta)} \Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] p_w(w)dw \\
&\quad \leq \Pr\left[1/v \geq (1-1/k)/(h(\delta)+\delta/k)\right] \\
&\qquad + \lambda_w \int_{lh(\delta)}^{kh(\delta)} \Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] dw \ , \quad \text{(B.15)}
\end{aligned}$$

where the first term in the bound of (B.15) follows from the fact that

$$\Pr\left[1/v \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right]$$

is non-increasing in w, and the second term in the bound of (B.15) follows from the fact that $p_w(w) = \lambda_w \exp(-\lambda_w w) \leq \lambda_w$.

Now, the first term of (B.15) satisfies

$$\Pr\left[1/v \geq (1-1/k)/(h(\delta)+\delta/k)\right]/h(\delta) \leq \lambda_v(1+\delta/(kh(\delta)))/(1-1/k) \quad \text{(B.16)}$$

and, by a change of variable $w' = w/h(\delta)$, the second term of (B.15) satisfies

$$\begin{aligned}
&\frac{1}{h(\delta)} \int_{lh(\delta)}^{kh(\delta)} \Pr\left[1/\mathsf{v} \geq (1/h(\delta) - 1/w)/(1+\delta/w)\right] dw \\
&= h(\delta) \int_l^k \frac{1}{h(\delta)} \left(1 - \exp\left[-\frac{\lambda_{\mathsf{v}}(h(\delta)+\delta/w')}{(1-1/w')}\right]\right) dw' \\
&\leq h(\delta) \underbrace{\int_l^k \lambda_{\mathsf{v}} \left(\frac{1+\delta/(w'h(\delta))}{1-1/w'}\right) dw'}_{B(\delta,h(\delta),k,l)} ,
\end{aligned} \tag{B.17}$$

where $B(\delta, h(\delta), k, l)$ remains finite for any $k > l > 1$ as $\delta \to 0$.

Combining (B.14), (B.16), and (B.17), we have

$$\frac{1}{h(\delta)} \Pr\left[\mathsf{r}_\delta < h(\delta)\right] \leq \lambda_{\mathsf{w}} l + \lambda_{\mathsf{v}} \left(\frac{1+\delta/(kh(\delta))}{1-1/k}\right) + h(\delta) B(\delta, h(\delta), k, l) \ , \tag{B.18}$$

and furthermore

$$\limsup_{\delta \to 0} \frac{1}{h(\delta)} \Pr\left[\mathsf{r}_\delta < h(\delta)\right] \leq \lambda_{\mathsf{w}} l + \lambda_{\mathsf{v}} \left(\frac{1+d/k}{1-1/k}\right) \ ,$$

since $\lim_{\delta\to 0} B(\delta, h(\delta), k, l) < \infty$ and, by assumption, $h(\delta) \to 0$ and $\delta/h(\delta) \to d$ as $\delta \to 0$.

The constants $k > l > 1$ are arbitrary. In particular, k can be chosen arbitrarily large, and l arbitrarily close to 1. Hence,

$$\limsup_{\delta \to 0} \frac{1}{h(\delta)} \Pr\left[\mathsf{r}_\delta < h(\delta)\right] \leq \lambda_{\mathsf{w}} + \lambda_{\mathsf{v}}. \tag{B.19}$$

Combining (B.12) with (B.19), the lemma is proved. ■

Proof:(Of Claim 1)

$$\begin{aligned}
&\Pr\left[\mathsf{u} + \epsilon\, f\left(\mathsf{v}/\epsilon,\, \mathsf{w}/\epsilon\right) < g(\epsilon)\right] \\
&= \Pr\left[\mathsf{u} + \mathsf{r}_\epsilon < g(\epsilon)\right] \\
&= \int_0^{g(\epsilon)} \Pr\left[\mathsf{r}_\epsilon < g(\epsilon) - u\right] p_{\mathsf{u}}(u) du \\
&= g(\epsilon) \int_0^{g(\epsilon)} \Pr\left[\mathsf{r}_\epsilon < g(\epsilon)(1-u')\right] \lambda_{\mathsf{u}} e^{-\lambda_{\mathsf{u}} g(\epsilon) u'} du' \\
&= g^2(\epsilon) \int_0^1 (1-u') \left[\frac{\Pr\left[\mathsf{r}_\epsilon < g(\epsilon)(1-u')\right]}{g(\epsilon)(1-u')}\right] \lambda_{\mathsf{u}} e^{-\lambda_{\mathsf{u}} g(\epsilon) u'} du',
\end{aligned} \tag{B.20}$$

where in the second equality we have used the change of variables $u' = u/g(\epsilon)$. But by Lemma 2 with $\delta = \epsilon$ and $h(\delta) = g(\delta)(1-u')$, the quantity in brackets approaches $\lambda_v + \lambda_w$ as $\epsilon \to 0$, so we expect

$$\lim_{\epsilon\to 0} \frac{1}{g^2(\epsilon)} \Pr\left[u + r_\epsilon < g(\epsilon)\right] = \lambda_u(\lambda_v + \lambda_w) \int_0^1 (1-u)du = \frac{\lambda_u(\lambda_v+\lambda_w)}{2} \ . \tag{B.21}$$

To fully verify (B.21), we must utilize the lower and upper bounds developed in Lemma 2.

Using the lower bound (B.12), (B.20) satisfies

$$\begin{aligned}
&\liminf_{\epsilon\to 0} \frac{1}{g^2(\epsilon)} \Pr\left[u + r_\epsilon < g(\epsilon)\right] \\
&\quad \geq \lim_{\epsilon\to 0} \int_0^1 \left(\frac{1-\exp\left[-(\lambda_v+\lambda_w)g(\epsilon)(1-u')\right]}{g(\epsilon)}\right) \lambda_u e^{-\lambda_u g(\epsilon)u'} du' \\
&\quad = \lambda_u(\lambda_v+\lambda_w)\int_0^1 (1-u')du' = \frac{\lambda_u(\lambda_v+\lambda_w)}{2}
\end{aligned} \tag{B.22}$$

where the first equality results from the Dominated Convergence Theorem [3] after noting that the integrand is both bounded by and converges to the function $\lambda_u(\lambda_v+\lambda_w)(1-u')$.

Using the upper bound (B.19), (B.20) satisfies

$$\begin{aligned}
&\limsup_{\epsilon\to 0} \frac{1}{g^2(\epsilon)} \Pr\left[u + r_\epsilon < g(\epsilon)\right] \\
&\quad \leq \limsup_{\epsilon\to 0} (\lambda_v/(1-1/k) + \lambda_w l) \int_0^1 (1-u')\lambda_u e^{-\lambda_u g(\epsilon)u'} du' \\
&\qquad + \limsup_{\epsilon\to 0} \epsilon/g(\epsilon) \int_0^1 \lambda_v\lambda_u e^{-\lambda_u g(\epsilon)u'}/(k-1) du' \\
&\qquad + \limsup_{\epsilon\to 0} g(\epsilon) \underbrace{\int_0^1 (1-u')^2 B(\epsilon, g(\epsilon)(1-u'), k, l)\lambda_u e^{-\lambda_u g(\epsilon)u'} du'}_{D(\epsilon, g(\epsilon), k, l)} \\
&\quad = \frac{\lambda_u\left[\lambda_v\left(\frac{1+c/k}{1-1/k}\right) + \lambda_w l\right]}{2} \ ,
\end{aligned} \tag{B.23}$$

where the last equality results from the fact $\epsilon/g(\epsilon) \to c$ and $D(\epsilon, g(\epsilon), k, l)$ remains finite for all $k > l > 1$ even as $\epsilon \to 0$.

Again, the constants $k > l > 1$ are arbitrary. In particular, k can be chosen arbitrarily large, and l arbitrarily close to 1. Hence,

$$\limsup_{\epsilon\to 0} \frac{1}{g^2(\epsilon)} \Pr\left[u + r_\epsilon < g(\epsilon)\right] \leq \frac{\lambda_u(\lambda_v+\lambda_w)}{2} \tag{B.24}$$

Combining (B.22) and (B.24) completes the proof. ■

Claim 2 *Let u and v be independent exponential random variables with parameters λ_u and λ_v, respectively. Let ϵ be positive and let $g(\epsilon) > 0$ be continuous with $g(\epsilon) \to 0$ as $\epsilon \to 0$. Define*

$$h(\epsilon) \stackrel{\triangle}{=} \epsilon^2 \left[(g(\epsilon)/\epsilon + 1) \ln(g(\epsilon)/\epsilon + 1) - g(\epsilon)/\epsilon \right] \ . \tag{B.25}$$

Then

$$\lim_{\epsilon \to 0} \frac{1}{h(\epsilon)} \Pr\left[u + v + uv/\epsilon < g(\epsilon) \right] = \lambda_u \lambda_v \ . \tag{B.26}$$

Moreover, if $\epsilon(t)$ is continuous about $t = t_0$ with $\epsilon(t) \to 0$ as $t \to t_0$, then

$$\lim_{t \to t_0} \frac{1}{h(\epsilon(t))} \Pr\left[u + v + uv/\epsilon(t) < g(\epsilon(t)) \right] = \lambda_u \lambda_v \ . \tag{B.27}$$

Proof: First, we write CDF in the form:

$$\begin{aligned}
&\Pr\left[u + v + uv/\epsilon < g(\epsilon) \right] \\
&= \int_0^\infty \Pr\left[u + v + uv/\epsilon < g(\epsilon) \middle| \, v = v \right] p_v(v) \, dv \\
&= \int_0^{g(\epsilon)} \Pr\left[u < \frac{g(\epsilon) - v}{1 + v/\epsilon} \middle| \, v = v \right] \lambda_v e^{-\lambda_v v} \, dv \\
&= \int_0^{g(\epsilon)} \left[1 - \exp\left(-\lambda_u \left[\frac{g(\epsilon) - v}{1 + v/\epsilon} \right] \right) \right] \lambda_v e^{-\lambda_v v} \, dv \\
&= g(\epsilon) \int_0^1 \left[1 - \exp\left(-\lambda_u \left[\frac{g(\epsilon)(1 - w)}{1 + g(\epsilon) w/\epsilon} \right] \right) \right] \lambda_v e^{-\lambda_v g(\epsilon) w} \, dw \ ,
\end{aligned} \tag{B.28}$$

where the last equality follows from the change of variables $w = v/g(\epsilon)$.

To upper bound (B.28), we use the identities $1 - e^{-x} \leq x$ for all $x \geq 0$ and $e^{-y} \leq 1$ for all $y \geq 0$, so that (B.28) becomes

$$\begin{aligned}
&\Pr\left[u + v + uv/\epsilon < g(\epsilon) \right] \\
&\leq g^2(\epsilon) \lambda_u \lambda_v \int_0^1 \frac{1 - w}{1 + g(\epsilon) w/\epsilon} \, dw \\
&= \lambda_u \lambda_v \, g^2(\epsilon) \, \frac{(g(\epsilon)/\epsilon + 1) \ln(g(\epsilon)/\epsilon + 1) - g(\epsilon)/\epsilon}{(g(\epsilon)/\epsilon)^2} \\
&= \lambda_u \lambda_v h(\epsilon) \ ,
\end{aligned}$$

whence

$$\limsup_{\epsilon\to 0} \frac{1}{h(\epsilon)} \Pr\left[u + v + uv/\epsilon < g(\epsilon)\right] \le \lambda_u \lambda_v \ . \tag{B.29}$$

To lower bound (B.28), we use the concavity of $1 - e^{-x}$, *i.e.*, for any $t > 0$, $1 - e^{-x} \ge \frac{1-e^{-t}}{t} x$ for all $x \le t$, and the identity $e^{-y} \ge 1 - y$ for all $y \ge 0$, so that (B.28) becomes

$$\begin{aligned}
&\Pr\left[u + v + uv/\epsilon < g(\epsilon)\right] \\
&\quad \ge \quad g(\epsilon) \int_0^1 \left[\left(\frac{1 - e^{-\lambda_u g(\epsilon)}}{\lambda_u g(\epsilon)} \right) \frac{\lambda_u g(\epsilon)(1-w)}{1 + w g(\epsilon)/\epsilon} \right] \lambda_v (1 - \lambda_v g(\epsilon) w)\, dw \\
&\quad = \quad \lambda_u \lambda_v \, g^2(\epsilon) \left(\frac{1 - e^{-\lambda_u g(\epsilon)}}{\lambda_u g(\epsilon)} \right) \int_0^1 \left[\frac{1-w}{1 + w g(\epsilon)/\epsilon} \right] (1 - \lambda_v g(\epsilon) w)\, dw \\
&\quad \ge \quad \lambda_u \lambda_v g^2(\epsilon) \left(\frac{1 - e^{-\lambda_u g(\epsilon)}}{\lambda_u g(\epsilon)} \right) (1 - \lambda_v g(\epsilon)) \int_0^1 \frac{1-w}{1 + w g(\epsilon)/\epsilon}\, dw \\
&\quad = \quad \lambda_u \lambda_v \left(\frac{1 - e^{-\lambda_u g(\epsilon)}}{\lambda_u g(\epsilon)} \right) (1 - \lambda_v g(\epsilon)) \ g^2(\epsilon) \, \frac{(g(\epsilon)/\epsilon + 1)\ln(g(\epsilon)/\epsilon + 1) - g(\epsilon)/\epsilon}{(g(\epsilon)/\epsilon)^2} \\
&\quad = \quad \lambda_u \lambda_v \left(\frac{1 - e^{-\lambda_u g(\epsilon)}}{\lambda_u g(\epsilon)} \right) (1 - \lambda_v g(\epsilon)) \ h(\epsilon) \ .
\end{aligned}$$

Thus,

$$\begin{aligned}
&\liminf_{\epsilon\to 0} \frac{1}{h(\epsilon)} \Pr\left[u + v + uv/\epsilon < g(\epsilon)\right] \\
&\quad \ge \quad \lambda_u \lambda_v \lim_{\epsilon\to 0} \left(\frac{1 - e^{-\lambda_u g(\epsilon)}}{\lambda_u g(\epsilon)} \right) (1 - \lambda_v g(\epsilon)) \\
&\quad = \quad \lambda_u \lambda_v \ .
\end{aligned} \tag{B.30}$$

Since the bounds in (B.29) and (B.30) are equal, the claim is proved. ∎

Claim 3 *Suppose $f_t(s) \to g(s)$ pointwise as $t \to t_0$, and that $f_t(s)$ is monotone increasing in s for each t. Let $h_t(s)$ be such that $h_t(s) \le s$, $h_t(s) \to s$ pointwise as $t \to t_0$, and $h_t(s)/s$ is monotone decreasing in s for each t. Define $\tilde{h}_t^{-1}(r) \stackrel{\triangle}{=} \min h_t^{-1}(r)$. Then*

$$\lim_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) = g(r) \ . \tag{B.31}$$

Proof: Since $h_t(s) \le s$ for all t, we have $r \le \tilde{h}_t^{-1}(r)$, and consequently $f_t(r) \le f_t(\tilde{h}_t^{-1}(r))$

because $f_t(\cdot)$ is monotone increasing. Thus,

$$\liminf_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) \geq g(r) \ . \tag{B.32}$$

The upper bound is a bit more involved. Fix $\delta > 0$. Lemma 3 shows that for each r there exists t^* such that $\tilde{h}_t^{-1}(r) \leq r/(1-\delta)$ for all t such that $|t - t_0| < |t^* - t_0|$. Then we have

$$f_t(\tilde{h}_t^{-1}(r)) \leq f_t(r/(1-\delta)) \ .$$

Thus,

$$\limsup_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) \leq g(r/(1-\delta)) \ ,$$

and since δ can be made arbitrarily small,

$$\limsup_{t\to t_0} f_t(\tilde{h}_t^{-1}(r)) \leq g(r) \ . \tag{B.33}$$

Combining (B.32) with (B.33), we obtain the desired result. ■

The following Lemma is used in the proof of the upper bound of Claim 3.

Lemma 3 *Let $h_t(s)$ be such that $h_t(s) \leq s$, $h_t(s) \to s$ pointwise as $t \to t_0$, and $h_t(s)/s$ is monotone decreasing in s for each t. Define $\tilde{h}_t^{-1}(r) \stackrel{\triangle}{=} \min h_t^{-1}(r)$. For each $r_0 > 0$ and any $\delta > 0$, there exists t^* such that*

$$\tilde{h}_t^{-1}(r_0) \leq r_0/(1-\delta) \ ,$$

for all t such that $|t - t_0| < |t^* - t_0|$.

Proof: Fix $r_0 > 0$ and $\delta > 0$, and select s_0 such that $s_0 > r_0/(1-\delta)$.

Because $h_t(s)/s \to 1$ point-wise as $t \to t_0$, for each $s > 0$ and any $\delta > 0$, there exists a t^* such that

$$h_t(s) \geq s(1-\delta), \quad \text{all } t : |t - t_0| < |t^* - t_0|.$$

Moreover, since $h_t(s)/s$ is monotone decreasing in s, if t^* is sufficient for convergence at s_0, then it is sufficient for convergence at all $s \leq s_0$. Thus, for any $s_0 > 0$ and $\delta > 0$ there

exists a t^* such that

$$h_t(s) \geq s(1-\delta), \quad \text{all } s \leq s_0, t : |t-t_0| < |t^*-t_0| \ .$$

Throughout the rest of the proof, we only consider $s \leq s_0$ and t such that $|t-t_0| < |t^*-t_0|$.

Consider the interval $I = [r_0, r_0/(1-\delta)]$, and note that $s \in I$ implies $s < s_0$. Since $h_t(s) < s$, we have $h_t(r_0) < r_0$. Also, since $h_t(s) > s(1-\delta)$ by the above construction, we have $h_t(r_0/(1-\delta)) > r_0$. By continuity, $h_t(s)$ assumes all intermediate values between $h_t(r_0)$ and $h_t(r_0/(1-\delta))$ on the interval $(r_0, r_0/(1-\delta))$ [64, Theorem 4.23]; in particular, there exists an $s_1 \in (r_0, r_0/(1-\delta))$ such that $h_t(x_1) = r_0$. The result follows from $\tilde{h}_t^{-1}(r_0) \leq x_1 \leq r_0/(1-\delta)$, where the first inequality follows from the definition of $\tilde{h}_t^{-1}(\cdot)$ and the second inequality follows from the fact that $x_1 \in I$. ■

B.2 Results for Chapter 6

We gather in this section the analytical results for Chapter 6, in order to focus the body of the chapter on discussion and interpretation of the results. We begin in Section B.2.1 by developing a general result about asymptotic properties of the CDF of a sum of independent random variables. We then apply this result to obtain large SNR approximations for repetition decode-and-forward cooperative diversity in Section B.2.2 and for space-time coded cooperative diversity in Section B.2.3.

B.2.1 The Basic Result

Both of the arguments later in this appendix rely upon the following result, which is a generalization of Fact 2 to several random variables with fairly general PDFs.

Claim 4 *Let u_k, $k = 1, 2, \ldots, m$, be positive, independent random variables with*

$$\liminf_{\epsilon \to 0} p_{u_k}(\epsilon u) \geq \lambda_k \ , \tag{B.34}$$

and

$$p_{u_k}(\epsilon u) \leq \lambda_k \ . \tag{B.35}$$

Then

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^m} \Pr\left[\sum_{k=1}^{m} u_k < \epsilon\right] = \frac{1}{m!} \prod_{k=1}^{m} \lambda_k \ . \tag{B.36}$$

Before proving Claim 4, we note that the exponential distribution satisfies both requirements (B.34) and (B.35). More generally, however, this result suggests that many of our results hold for a much larger class of PDFs, and, in particular, depend mainly upon properties of the PDFs near the origin. Although we do not provide a proof, we conjecture that random variables resulting from amplify-and-forward transmission, *i.e.*, $u_k = (v_k\, w_k)/(v_k + w_k + 1)$, with v_k and w_k independent exponential random variables with parameters λ_{v_k} and λ_{w_k}, respectively, satisfy (B.34) and (B.35) with $\lambda_k = \lambda_{v_k} + \lambda_{w_k}$.

Proof: Let $s_n \stackrel{\triangle}{=} \sum_{k=1}^{n} u_k$, $n \leq m$. Then

$$\begin{aligned} \Pr\left[\sum_{k=1}^{m} u_k < \epsilon\right] &= \Pr\left[s_m < \epsilon\right] \\ &= \int_0^{\epsilon} p_{s_m}(s) ds \qquad \text{(B.37)} \\ &= \epsilon \int_0^1 p_{s_m}(\epsilon w) dw \ , \qquad \text{(B.38)} \end{aligned}$$

where the last equality results from the change of variables $w = s/\epsilon$. Thus, it is sufficient for us to compute the limit

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} \int_0^1 p_{s_m}(\epsilon w) dw \ . \tag{B.39}$$

To lower bound the lim inf, we exploit Fatou's lemma [3] to obtain

$$\liminf_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} \int_0^1 p_{s_m}(\epsilon w) dw \geq \int_0^1 \left\{ \liminf_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} p_{s_m}(\epsilon w) \right\} dw \ . \tag{B.40}$$

Now, $s_m = s_{m-1} + u_m$, and by independence the PDF of s_m is the convolution of the PDFs of s_{m-1} and u_m. Specifically, since u_m is positive, we have

$$\begin{aligned} p_{s_m}(s) &= \int_0^s p_{s_{m-1}}(s - r) p_{u_m}(r) dr \\ &= s \int_0^1 p_{s_{m-1}}(s(1-y)) p_{u_m}(sy) dy \ , \qquad \text{(B.41)} \end{aligned}$$

where the last equality results from the change of variables $y = r/s$.

Letting

$$A_m(w) \stackrel{\triangle}{=} \liminf_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} p_{\mathsf{s}_m}(\epsilon w) \ , \tag{B.42}$$

substituting into (B.41), and again exploiting Fatou's lemma, we obtain the recursion

$$\begin{aligned} A_m(w) &= \liminf_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} p_{\mathsf{s}_m}(\epsilon w) dw \\ &\geq w \int_0^1 \left\{ \liminf_{\epsilon \to 0} \frac{1}{\epsilon^{(m-2)}} p_{\mathsf{s}_{m-1}}(\epsilon w(1-y)) \right\} \left\{ \liminf_{\epsilon \to 0} p_{\mathsf{u}_m}(\epsilon w y) \right\} dy \\ &\geq \lambda_m w \int_0^1 A_{(m-1)}(w(1-y)) dy \ , \end{aligned} \tag{B.43}$$

where the last inequality follows from (B.34) and substitution of $A_{(m-1)}(w(1-y))$. Beginning with $A_1(w) \geq \lambda_1$ from (B.34), the recursion (B.43) yields

$$A_m(w) \geq \frac{1}{(m-1)!} w^{(m-1)} \prod_{k=1}^{m} \lambda_k \ . \tag{B.44}$$

As a result, (B.40) with (B.42) and (B.44) yields,

$$\liminf_{\epsilon \to 0} \frac{1}{\epsilon^m} \Pr\left[\mathsf{s}_m < \epsilon\right] \geq \frac{1}{m!} \prod_{k=1}^{m} \lambda_k \ . \tag{B.45}$$

To upper bound the lim sup, we obtain a recursive upper bound for the PDF of s_m similar to the lower bound developed above. Specifically, letting

$$B_m(w, \epsilon) \stackrel{\triangle}{=} p_{\mathsf{s}_m}(\epsilon w) \ , \tag{B.46}$$

we have

$$\begin{aligned} &B_m(w, \epsilon) \\ &\quad = \epsilon w \int_0^1 p_{\mathsf{s}_{m-1}}(\epsilon w(1-y)) \, p_{\mathsf{u}_m}(\epsilon w y) \, dy \\ &\quad \leq \epsilon \lambda_m w \int_0^1 B_{m-1}(w(1-y), \epsilon) dy \ , \end{aligned} \tag{B.47}$$

where the equality comes from the convolution (B.41), and the inequality follows from (B.35) and substitution of $B_{m-1}(w(1-y))$. Beginning with $B_1(w, \epsilon) \leq \lambda_1$ from (B.35),

(B.47) yields an upper bound very similar to the lower bound in (B.44), namely,

$$B_m(w,\epsilon) \leq \epsilon^{(m-1)} w^{(m-1)} \frac{1}{(m-1)!} \prod_{k=1}^{m} \lambda_k \ . \tag{B.48}$$

Then

$$\begin{aligned} \limsup_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} \int_0^1 p_{\mathsf{s}_m}(\epsilon w) dw &\leq \limsup_{\epsilon \to 0} \frac{1}{\epsilon^{(m-1)}} \int_0^1 B_m(w,\epsilon) dw \\ &\leq \frac{1}{m!} \prod_{k=1}^{m} \lambda_k \ . \end{aligned} \tag{B.49}$$

Together with the fact that, in general $\liminf \leq \limsup$, (B.45) and (B.49) yield the desired result (B.36). ■

B.2.2 Repetition Decode-and-Forward Cooperative Diversity

In this section, we utilize the result of Claim 4 to obtain a large $\mathtt{SNR}$ approximation for $\Pr\left[I_{\text{rep}} < \mathtt{R} | \mathcal{D}(s)\right]$, the conditional outage probability for repetition decode-and-forward cooperative diversity for source s given a set of decoding relays $\mathcal{D}(s)$. As in (6.7), I_{rep} is of the form

$$I_{\text{rep}} = \frac{1}{m} \log \left(1 + \mathtt{SNR} \sum_{k=1}^{m} u_k \right) \ , \tag{B.50}$$

where u_k are independent exponential random variables with parameters λ_k, $k = 1, 2, \ldots, m$.

After some algebraic manipulations, the outage probability reduces to exactly the same form as in Claim 4,

$$\Pr\left[I_{\text{rep}} < \mathtt{R} | \mathcal{D}(s)\right] = \Pr\left[\sum_{k=1}^{m} u_k < \epsilon \right] \ , \tag{B.51}$$

with $\epsilon = (2^{m\mathtt{R}} - 1)/\mathtt{SNR} \to 0$ as $\mathtt{SNR} \to \infty$. Thus, Claim 4 and continuity yield the approximation

$$\Pr\left[I_{\text{rep}} < \mathtt{R} | \mathcal{D}(s)\right] \sim \left[\frac{2^{m\mathtt{R}} - 1}{\mathtt{SNR}} \right]^m \frac{1}{m!} \prod_{k=1}^{m} \lambda_k \ , \tag{B.52}$$

for large $\mathtt{SNR}$.

B.2.3 Space-Time Coded Cooperative Diversity

In this section, we compute a large $\mathtt{SNR}$ approximation for $\Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right]$, the conditional outage probability for space-time coded cooperative diversity for source s given a set of decoding relays $\mathcal{D}(s)$. As in (6.15), I_{stc} is of the form

$$I_{\text{stc}} = \frac{1}{2}\log\left(1 + \frac{2}{m}\mathtt{SNR}\, u_m\right) + \frac{1}{2}\log\left(1 + \frac{2}{m}\mathtt{SNR}\sum_{k=1}^{m-1} u_k\right) , \tag{B.53}$$

where again u_k are independent exponential random variables with parameters λ_k, $k = 1, 2, \ldots, m$.

Let $s_{m-1} \triangleq \sum_{k=1}^{m-1} u_k$, $t \triangleq (2^{2\mathtt{R}} - 1)$, and $\epsilon \triangleq (2^{2\mathtt{R}} - 1)/(2\mathtt{SNR}/m)$. Then

$$\begin{aligned}
\Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right] &= \Pr\left[u_m + s_{m-1} + \frac{2}{m}\mathtt{SNR} u_m s_{m-1} < \epsilon\right] \\
&= \int_0^{\epsilon} \Pr\left[u_m < \frac{\epsilon - s}{1 + (2\mathtt{SNR}/m)s}\right] p_{s_{m-1}}(s) ds \\
&= \epsilon \int_0^1 \Pr\left[u_m < \frac{\epsilon(1-w)}{1+tw}\right] p_{s_{m-1}}(\epsilon w) dw \\
&= \epsilon \int_0^1 \left[1 - \exp\left(-\lambda_m \frac{\epsilon(1-w)}{1+tw}\right)\right] p_{s_{m-1}}(\epsilon w) dw \ .
\end{aligned} \tag{B.54}$$

Note the penultimate equality in (B.54) follows from the change of variables $w = s/\epsilon$, and the last equality follows from substituting the CDF for u_m.

We now compute the limit

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^m} \Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right] = \frac{1}{(m-2)!} \prod_{k=1}^{m} \lambda_m \int_0^1 \left[\frac{1-w}{1+tw}\right] w^{(m-2)} dw \ , \tag{B.55}$$

that, along with continuity, provides the large $\mathtt{SNR}$ approximation

$$\Pr\left[I_{\text{stc}} < \mathtt{R} | \mathcal{D}(s)\right] \sim \left[\frac{2^{2\mathtt{R}} - 1}{2\mathtt{SNR}/m}\right]^m \frac{1}{(m-2)!} \prod_{k=1}^{m} \lambda_m \int_0^1 \left[\frac{1-w}{1+(2^{2\mathtt{R}}-1)w}\right] w^{(m-2)} \ . \tag{B.56}$$

To lower bound the lim inf, we exploit Fatou's lemma in (B.54) to obtain

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\frac{1}{\epsilon^m}\Pr\left[I_{\text{stc}}<\mathtt{R}|\mathcal{D}(s)\right] \\
&\qquad\geq \int_0^1 \left\{\liminf_{\epsilon\to 0}\frac{1}{\epsilon}\left[1-\exp\left(-\lambda_m\frac{\epsilon(1-w)}{1+tw}\right)\right]\right\}\left\{\liminf_{\epsilon\to 0}\frac{1}{\epsilon^{(m-2)}}p_{\boldsymbol{s}_{m-1}}(\epsilon w)\right\}dw \\
&\qquad= \int_0^1 \lambda_m\left[\frac{1-w}{1+tw}\right]A_{m-1}(w)dw \\
&\qquad\geq \frac{1}{(m-2)!}\prod_{k=1}^{m}\lambda_k\int_0^1\left[\frac{1-w}{1+tw}\right]w^{(m-2)}dw\ ,
\end{aligned} \tag{B.57}$$

where the first equality follows from Fact 1 and substitution of $A_{m-1}(w)$ from (B.42), and the second equality follows from the result (B.44) in the proof of Claim 4.

To upper bound the lim sup, we derive

$$\begin{aligned}
&\limsup_{\epsilon\to 0}\frac{1}{\epsilon^m}\Pr\left[I_{\text{stc}}<\mathtt{R}|\mathcal{D}(s)\right] \\
&\qquad\leq \limsup_{\epsilon\to 0}\int_0^1 \left\{\frac{1}{\epsilon}\left[1-\exp\left(-\lambda_m\frac{\epsilon(1-w)}{1+tw}\right)\right]\right\}\left\{\frac{1}{\epsilon^{(m-2)}}B_{m-1}(w,\epsilon)\right\}dw \\
&\qquad\leq \limsup_{\epsilon\to 0}\int_0^1 \left\{\lambda_m\left[\frac{(1-w)}{1+tw}\right]\right\}\left\{\frac{1}{\epsilon^{(m-2)}}B_{m-1}(w,\epsilon)\right\}dw \\
&\qquad\leq \limsup_{\epsilon\to 0}\int_0^1 \left\{\lambda_m\left[\frac{(1-w)}{1+tw}\right]\right\}\left\{\frac{1}{(m-2)!}w^{(m-2)}\prod_{k=1}^{m-1}\lambda_k\right\}dw \\
&\qquad= \frac{1}{(m-2)!}\prod_{k=1}^{m}\lambda_k\int_0^1\left[\frac{(1-w)}{1+tw}\right]w^{(m-2)}dw\ ,
\end{aligned} \tag{B.58}$$

where the first inequality follows from substitution of $B_{m-1}(w,\epsilon)$ from (B.46), the second inequality follows from the fact that $1-\exp(-x)\leq x$ for all $x\geq 0$, and the third inequality follows from the result (B.48) in Claim 4.

Taken together with the fact that lim inf $\leq$ lim sup, (B.57) and (B.58) yield the desired result (B.55).

Appendix C

Mutual Information Calculations

C.1 Amplify-and-Forward Mutual Information

In this section, we compute the maximum average mutual information for amplify-and-forward transmission. We write the equivalent channel (5.2)–(5.4), with relay processing (5.8), in vector form as

$$\underbrace{\begin{bmatrix} y_d[n] \\ y_d[n+N/4] \end{bmatrix}}_{\mathbf{y}_d[n]} = \underbrace{\begin{bmatrix} a_{s,d(s)} \\ a_{r,d(s)}\beta a_{s,r} \end{bmatrix}}_{A} x_s[n] + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ a_{r,d(s)}\beta & 0 & 1 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} z_r[n] \\ z_d[n] \\ z_r[n+N/4] \end{bmatrix}}_{\mathbf{z}[n]}$$

where the source signal has power constraint $\mathrm{E}\left[x_s\right] \leq P_s$, and relay amplifier has constraint

$$\beta \leq \sqrt{\frac{P_r}{|a_{s,r}|^2 P_s + N_r}} \ , \tag{C.1}$$

and the noise has covariance $\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right] = \mathrm{diag}(N_r, N_d, N_d)$. Note that we determine the mutual information for arbitrary transmit powers and noise levels, even though we utilize the result only for the symmetric case. Since the channel is memoryless, the average mutual information satisfies

$$I_{AF} \leq I(x_s; \mathbf{y}_d) \leq \log\det\left(I + (P_s A A^\dagger)(B\,\mathrm{E}\left[zz^\dagger\right]\ B^\dagger)^{-1}\right) \ ,$$

with equality for x_s zero-mean, circularly symmetric complex Gaussian. Noting that

$$AA^\dagger = \begin{bmatrix} |a_{s,d(s)}|^2 & a_{s,d(s)}(a_{r,d(s)}\beta a_{s,r})^* \\ a^*_{s,d(s)}a_{r,d(s)}\beta a_{s,r} & |a_{r,d(s)}\beta a_{s,r}|^2 \end{bmatrix}$$

$$B\,\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]\,B^\dagger = \begin{bmatrix} N_d & 0 \\ 0 & |a_{r,d(s)}\beta|^2 N_r + N_d \end{bmatrix},$$

we have

$$\det\left(I_2 + (P_s AA^\dagger)(B\,\mathrm{E}\left[zz^\dagger\right]\,B^\dagger)^{-1}\right) = 1 + \frac{P_s|a_{s,d(s)}|^2}{N_d} + \frac{P_s|a_{r,d(s)}\beta a_{s,r}|^2}{\left(|a_{r,d(s)}\beta|^2 N_r + N_d\right)}\,. \qquad \text{(C.2)}$$

It is apparent that (C.2) is increasing in β, so the amplifier power constraint (C.1) should be active, yielding, after substitutions and algebraic manipulations,

$$\begin{aligned} I_{AF} &= \log\left(1 + |a_{s,d(s)}|^2\left(\frac{P_s\sigma^2_{s,d(s)}}{N_d}\right) + \frac{\left[|a_{s,r}|^2\left(\frac{P_s\sigma^2_{s,r}}{N_r}\right)\right]\cdot\left[|a_{r,d(s)}|^2\left(\frac{P_r\sigma^2_{r,d(s)}}{N_d}\right)\right]}{\left[|a_{s,r}|^2\left(\frac{P_s\sigma^2_{s,r}}{N_r}\right)\right] + \left[|a_{r,d(s)}|^2\left(\frac{P_r\sigma^2_{r,d(s)}}{N_d}\right)\right] + 1}\right) \\ &= \log\left(1 + |a_{s,d(s)}|^2\mathtt{SNR}_{s,d(s)} + f\left(|a_{s,r}|^2\mathtt{SNR}_{s,r},\ |a_{r,d(s)}|^2\mathtt{SNR}_{r,d(s)}\right)\right)\,, \end{aligned}$$

with $f(\cdot,\cdot)$ given by (5.13).

C.2 Input Distributions for Transmit Diversity Bound

In this section, we derive the input distributions that minimize outage probability for transmit diversity schemes in the high SNR regime. Our derivation extends that of [78, 79] to deal with asymmetric fading variances.

An equivalent channel model for the two-antenna case can be summarized as

$$y[n] = \underbrace{\begin{bmatrix} a_1 & a_2 \end{bmatrix}}_{\mathbf{a}} \underbrace{\begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix}}_{\mathbf{x}[n]} + z[n]\,, \qquad \text{(C.3)}$$

where $\mathbf{a}$ represents the fading coefficients and $\mathbf{x}[n]$ the transmit signals from the two transmit antennas, and $z[n]$ is a zero-mean, white complex Gaussian process with variance N_0 that captures the effects of noise and interference. Let $Q = \mathrm{E}\left[\mathbf{x}\mathbf{x}^\dagger\right]$ be the covariance matrix for

the transmit signals. Then the power constraint on the inputs may be written in the form $\mathrm{tr}(Q) \leq P$.

We are interested in determining a distribution on the input vector $\mathbf{x}$, subject to the power constraint, that minimizes outage probability, *i.e.*,

$$\max_{p_{\mathbf{x}}:\mathrm{tr}(Q)\leq\mathcal{P}} \Pr\left[I(\mathbf{x};y|\mathbf{a}=\mathbf{a}) \leq \mathtt{R}\right] \ . \tag{C.4}$$

As [78, 79] develops, the optimization (C.4) can be restricted to optimization over zero-mean, circularly symmetric complex Gaussian inputs, because Gaussian codebooks maximize the mutual information for each value of the fading coefficients $\mathbf{a}$. Thus, (C.4) is equivalent to maximizing over the covariance matrix of the complex Gaussian inputs subject to the power constraint, *i.e.*,

$$\max_{Q:\mathrm{tr}(Q)\leq P} \Pr\left[\log\left(1+\frac{\mathbf{a}Q\mathbf{a}^\dagger}{N_0}\right) \leq \mathtt{R}\right] \ . \tag{C.5}$$

We now argue that Q diagonal is sufficient, even if the components of $\mathbf{a}$ are independent but not identically distributed. We note that this argument is a slight extension of [78, 79], in which i.i.d. fading coefficients are treated. Although we treat the case of two transmit antennas, the argument should extend naturally to more than two antennas.

We write $\mathbf{a} = \tilde{\mathbf{a}}\Sigma$, where $\tilde{\mathbf{a}}$ is a zero mean, i.i.d. complex Gaussian vector with unit variances and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2)$. Thus, the outage probability in (C.5) may be written as

$$\Pr\left[\log\left(1+\frac{\tilde{\mathbf{a}}\Sigma Q\Sigma^\dagger\tilde{\mathbf{a}}^\dagger}{N_0}\right) \leq \mathtt{R}\right] \ .$$

Now consider an eigen-decomposition of the matrix $\Sigma Q\Sigma^\dagger = UDU^\dagger$, where U is unitary and D is diagonal. Using the fact that the distribution of $\tilde{\mathbf{a}}$ is rotationally invariant, *i.e.*, $\tilde{\mathbf{a}}U$ has the same distribution as $\tilde{\mathbf{a}}$ for any unitary U [78, 79], we observe that the outage probability for covariance matrix $\Sigma Q\Sigma^\dagger$ is the same as the outage probability for the diagonal matrix D.

For $D = \mathrm{diag}(d_1, d_2)$, the outage probability can be written in the form

$$\Pr\left[d_1|a_1|^2 + d_2|a_2|^2 \leq \frac{2^{\mathtt{R}}-1}{\mathtt{SNR}}\right] \ ,$$

which, using Fact 2, decays proportional to $1/(\mathtt{SNR}^2 \det D)$ for large $\mathtt{SNR}$ if $d_1, d_2 \neq 0$. Thus,

minimizing the outage probability for large SNR is equivalent to maximizing

$$\det D = \det \Sigma Q \Sigma^{\dagger} = \sigma_1^2 \sigma_2^2 (Q_{1,1} Q_{2,2} - |Q_{1,2}|^2) \tag{C.6}$$

such that $Q_{1,1} + Q_{2,2} \leq P$. Clearly, (C.6) is maximized for $Q_{1,1} = Q_{2,2} = P/2$ and $Q_{1,2} = Q_{2,1} = 0$. Thus, zero-mean, i.i.d.complex Gaussian inputs minimize the outage probability in the high SNR regime.

Bibliography

[1] Ibrahim Abou-Faycal, Mitchell D. Trott, and Shlomo Shamai (Shitz). The capacity of discrete-time Rayleigh fading channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 473, 29 June – 4 July 1997.

[2] Ibrahim C. Abou-Faycal, Mitchell D. Trott, and Shlomo Shamai (Shitz). The capacity of discrete-time memoryless rayleigh-fading channels. *IEEE Trans. Inform. Theory*, 47(4):1290–1301, May 2001.

[3] Malcolm Adams and Victor Guillemin. *Measure Theory and Probability*. Birkhäuser, Boston, 1996.

[4] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[5] Siavash Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Select. Areas of Commun.*, 16(8):1451–1458, October 1998.

[6] John G. Apostolopoulos. Reliable video compression over lossy packet networks using multiple state encoding and path diversity. In *Proc. SPIE Visual Communications and Image Processing (VCIP)*, volume 4310, San Jose, CA, January 2001.

[7] Hari Balakrishnan, Venkata N. Padmanabhan, Srinivasan Seshan, and Randy H. Katz. A comparison of mechanisms for improving TCP performance over wireless links. *IEEE/ACM Trans. Networking*, 5(6):756–769, December 1997.

[8] Toby Berger, Zhen Zhang, and Harish Viswanathan. The CEO problem: Multiterminal source coding. *IEEE Trans. Inform. Theory*, 42(3):887–902, May 1996.

[9] Randall A. Berry. *Power and Delay Trade-Offs in Fading Channels*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2000.

[10] Dimitri P. Bertsekas and Robert G. Gallager. *Data Networks*. Prentice Hall, second edition, 1991.

[11] Ezio Biglieri, Giuseppe Caire, and Giorgio Taricco. Limiting performance of block-fading channels with multiple antennas. *IEEE Trans. Inform. Theory*, 47(4):1273–1289, May 2001.

[12] Ezio Biglieri, John Proakis, and Shlomo Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2692, October 1998.

[13] Giuseppe Caire, Giorgio Taricco, and Ezio Biglieri. Optimum power control over fading channels. *IEEE Trans. Inform. Theory*, 45(5):1468–1489, July 1999.

[14] Aydano B. Carleial. Multiple-access channels with different generalized feedback signals. *IEEE Trans. Inform. Theory*, 28(6):841, November 1982.

[15] Benjie Chen, Kyle Jamieson, Hari Balakrishnan, and Robert Morris. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. *ACM Wireless Networks Journal*, 8(5), September 2002.

[16] Thomas M. Cover. Broadcast channels. *IEEE Trans. Inform. Theory*, 18(1):2–14, January 1972.

[17] Thomas M. Cover and Abbas A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory*, 25(5):572–584, September 1979.

[18] Thomas M. Cover and Cyril S.K. Leung. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Inform. Theory*, 27(3):292–298, May 1981.

[19] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[20] Bevan Das and Vaduver Bharghavan. Routing in ad-hoc networks using minimum connected dominating sets. In *Proc. IEEE Int. Conf. Communications (ICC)*, volume 1, pages 376–380, Montreal, Canada, June 1997.

[21] Stark Christiaan Draper. *Successive Structuring of Source Coding Algorithms for Data Fusion, Buffering, and Distribution in Networks*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2002.

[22] Thomas H. E. Ericson. A Gaussian channel with slow fading. *IEEE Trans. Inform. Theory*, 16(3):353–355, May 1970.

[23] G. David Forney, Jr. Introduction to digital communications. M.I.T. Course 6.450 Lecture Notes, September 2000.

[24] Gerard J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Syst. Tech. J.*, pages 41–59, Autumn 1996.

[25] Gerard J. Foschini and Michael J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3):311–335, March 1998.

[26] Abbas A. El Gamal and Thomas M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inform. Theory*, 28(6):851–857, November 1982.

[27] Michael Gastpar, Gerhard Kramer, and Piyush Gupta. The multiple-relay channel: Coding and antenna-clustering capacity. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 136, Lausanne, Switzerland, July 1–5 2002.

[28] Michael Gastpar and Martin Vetterli. On the asymptotic capacity of Gaussian relay networks. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 195, Lausanne, Switzerland, July 1–5 2002.

[29] Michael Gastpar and Martin Vetterli. On the capacity of wireless networks: The relay case. In *Proc. IEEE INFOCOM*, New York, NY, June 2002.

[30] Andrea J. Goldsmith and Pravin P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory*, 43(6):1986–1992, November 1995.

[31] Matthias Grossglauser and David N.C. Tse. Mobility increases the capacity of ad-hoc wireless networks. Submitted to *IEEE/ACM Trans. Networking*, March 2001.

[32] Piyush Gupta and P.R. Kumar. The capacity of wireless networks. *IEEE Trans. Inform. Theory*, 46(2):388–404, March 2000.

[33] Piyush Gupta and P.R. Kumar. Towards and information theory of large networks: An achievable rate region. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 150, Washington DC, June 2001.

[34] Stephen V. Hanly and David N.C. Tse. Multiaccess fading channels — Part II: Delay-limited capacities. *IEEE Trans. Inform. Theory*, 44(7):2816–2831, November 1998.

[35] V. Hayes. IEEE standard for wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 1997.

[36] Wendi Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. Energy-efficient communication protocols for wireless microsensor networks. In *Proc. of the Hawaii Int. Conf. on System Sciences*, pages 3005–3014, Maui, HW, January 2000.

[37] Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. To appear in *IEEE Trans. on Wireless Comm.*, 2002.

[38] Bertand M. Hochwald and Wim Sweldens. Differential unitary space-time modulation. *IEEE Trans. Commun.*, 48(12):2041–2052, December 2000.

[39] Bertrand M. Hochwald and Thomas L. Marzetta. Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading. *IEEE Trans. Inform. Theory*, 46(2):543–564, March 2000.

[40] Bertrand M. Hochwald, Thomas L. Marzetta, Thomas J. Richardson, Wim Sweldens, and Rüdiger Urbanke. System design of unitary space-time constellations. *IEEE Trans. Inform. Theory*, 46(6):1962–1973, September 2000.

[41] Todd E. Hunter and Aria Nosratinia. Cooperation diversity through coding. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 220, Lausanne, Switzerland, July 2002.

[42] William C. Jakes and Donald C. Cox, editors. *Microwave Mobile Communications*. IEEE Press, New York, 1994.

[43] R. E. Kahn. The organization of computer resources into a packet radio network. *IEEE Trans. Commun.*, 25(1):169–178, January 1977.

[44] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman. Advances in packet radio technology. *Proc. IEEE*, 66(11):1468–1496, November 1978.

[45] Hisham Ibrahim Kassab. *Low-Energy Mobile Packet Radio Networks: Routing, Scheduling, and Architecture*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, November 1999.

[46] Raymond Knopp and Pierre A. Humblet. Information capacity and power control for single-cell multiuser communications. In *Proc. IEEE Int. Conf. Communications (ICC)*, volume 1, pages 331–335, Seattle, WA, 1995.

[47] Ralf Koetter and Muriel Medard. An algebraic approach to network coding. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 104, Washington, DC, June 2001.

[48] Gerhard Kramer and Adriaan J. van Wijngaarden. On the white Gaussian multiple-access relay channel. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 40, Sorrento, Italy, June 2000.

[49] J. Nicholas Laneman and Gregory W. Wornell. Energy-efficient antenna sharing and relaying for wireless networks. In *Proc. IEEE Wireless Comm. and Networking Conf. (WCNC)*, Chicago, IL, September 2000.

[50] Amos Lapidoth and Prakash Narayan. Reliable communications under channel uncertainty. *IEEE Trans. Inform. Theory*, 44(6):2148–2177, October 1998.

[51] Barry M. Leiner, Donald L. Nielson, and Foud A. Tobagi. Issues in packet radio network design. *Proc. IEEE*, 75(1):6–20, January 1987.

[52] Lifang Li and Andrea Goldsmith. Outage capacities and optimal power allocation for fading multiple-access channels. Submitted to *IEEE Trans. Inform. Theory*, 2001.

[53] Lifang Li and Andrea J. Goldsmith. Capacity and optimal resource allocation for fading broadcast channels – Part I: Ergodic capacity. *IEEE Trans. Inform. Theory*, 47(3):1083–1102, March 2001.

[54] Lifang Li and Andrea J. Goldsmith. Capacity and optimal resource allocation for fading broadcast channels – Part II: Outage capacity. *IEEE Trans. Inform. Theory*, 47(3):1103–1127, March 2001.

[55] Thomas L. Marzetta and Bertrand M. Hochwald. Capacity of a mobile multiple-antenna communication link in rayleigh flat fading. *IEEE Trans. Inform. Theory*, 45(1):139–157, January 1999.

[56] Patrick Maurer and Vahid Tarokh. Transmit diversity when the receiver does not know the number of transmit antennas. In *Proc. International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Aalborg, Denmark, September 2001.

[57] Aradhana Narula, Mitchell D. Trott, and Gregory W. Wornell. Performance limits of coded diversity methods for transmitter antenna arrays. *IEEE Trans. Inform. Theory*, 45(7):2418–2433, November 1999.

[58] Lawrence H. Ozarow. The capacity of the white Gaussian multiple access channel with feedback. *IEEE Trans. Inform. Theory*, 30(4):623–629, July 1984.

[59] Lawrence H. Ozarow, Shlomo Shamai (Shitz), and Aaron D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Veh. Technol.*, 43(5):359–378, May 1994.

[60] John G. Proakis. *Digital Communications*. McGraw-Hill, Inc., New York, Third edition, 1995.

[61] Theodore S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1996.

[62] Bixio Rimoldi. Successive refinement of information: Characterization of the achievable rates. *IEEE Trans. Inform. Theory*, 40(1):253–259, January 1994.

[63] Kenneth H. Rosen, editor. *Handbook of Discrete and Combinatorial Mathematics*. CRC Press, Boca Raton, FL, 2000.

[64] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Book Company, New York, second edition, 1964.

[65] Brett Schein. *Distributed Coordination in Network Information Theory*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, August 2001.

[66] Brett Schein and Robert G. Gallager. The Gaussian parallel relay network. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 22, Sorrento, Italy, June 2000.

[67] Andrew Sendonaris. *Advanced Techniques for Next-Generation Wireless Systems*. PhD thesis, Rice University, Houston, TX, August 1999.

[68] Andrew Sendonaris, Elza Erkip, and Behnaam Aazhang. Increasing uplink capacity via user cooperation diversity. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, MA, August 1998.

[69] Andrew Sendonaris, Elza Erkip, and Behnaam Aazhang. User cooperation diversity, Part I: System description. Submitted to *IEEE Trans. Commun.*, 1999.

[70] Andrew Sendonaris, Elza Erkip, and Behnaam Aazhang. User cooperation diversity, Part II: Implementation aspects and performance analysis. Submitted to *IEEE Trans. Commun.*, 1999.

[71] Shlomo Shamai (Shitz). A broadcast strategy for the Gaussian slowly fading channel. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 150, Ulm, Germany, June 29 – July 4 1997.

[72] Shlomo Shamai (Shitz) and Aaron D. Wyner. Information theoretic considerations for symmetric, cellular, multiple-access fading channels — Part I. *IEEE Trans. Inform. Theory*, 43(6):1877–1894, November 1997.

[73] Shlomo Shamai (Shitz) and Aaron D. Wyner. Information theoretic considerations for symmetric, cellular, multiple-access fading channels — Part II. *IEEE Trans. Inform. Theory*, 43(6):1895–1911, November 1997.

[74] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, IL, 1949.

[75] Timothy J. Shepard. *Decentralized Channel Management in Scalable Multihop Spread-Spectrum Packet Radio Networks*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, July 1995.

[76] Vahid Tarokh, H. Jafarkhani, and A. Robert Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. Inform. Theory*, 45(5):1456–1467, July 1999.

[77] Vahid Tarokh, Nambi Seshadri, and A. Robert Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Trans. Inform. Theory*, 44(2):744–765, March 1998.

[78] I. Emre Telatar. Capacity of multi-antenna Gaussian channels. Technical report, Bell Labs, Lucent Technologies, 1995.

[79] I. Emre Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. on Telecomm.*, 10(6):585–596, November-December 1999.

[80] Fouad A. Tobagi. Modeling and performance analysis of multihop packet radio networks. *Proc. IEEE*, 75(1):135–155, January 1987.

[81] S. Toumpis and Andrea J. Goldsmith. Ad-hoc network capacity. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2000.

[82] S. Toumpis and Andrea J. Goldsmith. Some capacity results for ad-hoc networks. In *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, October 2000.

[83] 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, November 1998.

[84] Edward C. van der Meulen. *Transmission of Information in a T-Terminal Discrete Memoryless Channel*. Department of Statistics, University of California, Berkeley, CA, 1968.

[85] Edward C. van der Meulen. Three-terminal communication channels. *Adv. Appl. Prob.*, 3:120–154, 1971.

[86] Sergio Verdú and T. S. Han. A general formula for channel capacity. *IEEE Trans. Inform. Theory*, 40(5):1147–1157, July 1994.

[87] Udo Wachsmann, Robert F.H. Fischer, and Johannes B. Huber. Multilevel codes: Theoretical concepts and practical design rules. *IEEE Trans. Inform. Theory*, 45(5):1361–1391, July 1999.

[88] Frans M.J. Willems. *Informationtheoretical Results for the Discrete Memoryless Multiple Access Channel*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, October 1982.

[89] Frans M.J. Willems. The discrete memoryless multiple access channel with partially cooperating encoders. *IEEE Trans. Inform. Theory*, 29(3):441–445, May 1983.

[90] Frans M.J. Willems and Edward C. van der Meulen. The discrete memoryless multiple-access channel with cribbing encoders. *IEEE Trans. Inform. Theory*, 31(3):313–327, May 1985.

[91] Frans M.J. Willems, Edward C. van der Meulen, and J. Pieter M. Schalkwijk. An achievable rate region for the multiple access channel with generalized feedback. In *Proc. Allerton Conf. Communications, Control, and Computing*, pages 284–292, Monticello, IL, October 1983.

[92] Aaron D. Wyner. Shannon-theoretic approach to a Gaussian cellular multiple access channel. *IEEE Trans. Inform. Theory*, 40(6):1713–1727, November 1994.

[93] Lizhong Zheng and David N.C. Tse. Optimal diversity-multiplexing tradeoff in multiple antenna channels. In *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, October 2001.

[94] Lizhong Zheng and David N.C. Tse. Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inform. Theory*, 48(2):359–383, February 2002.