OUTAGE LIMITED COOPERATIVE CHANNELS: PROTOCOLS AND ANALYSIS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Kambiz Azarian, B.Sc., M.Sc.

* * * * *

The Ohio State University

2006

Dissertation Committee:

Hesham El Gamal, Adviser

Philip Schniter

Randolph L. Moses

Andrea Serrani

Approved by

Adviser

Graduate Program in
Electrical Engineering

ABSTRACT

We propose novel cooperative protocols for various coherent flat-fading channels composed of half-duplex nodes. We consider relay, cooperative broadcast (CB), multiple-access relay (MAR) and cooperative multiple-access (CMA) channels and devise efficient protocols for them. We also present automatic repeat request (ARQ) variants of these protocols. We evaluate the proposed protocols using the diversity-multiplexing tradeoff (DMT).

For the relay channel, we investigate two classes of cooperation protocols, i.e., amplify and forward (AF) and decode and forward (DF). For the first class, we establish an upper bound on the achievable DMT with a single relay. We then propose a AF protocol that achieves this upper bound. The proposed algorithm is then extended to the case of arbitrary number of relays. For the class of DF protocols, we propose a dynamic decode and forward (DDF) protocol that achieves the optimal DMT for a range of multiplexing gains. We further show that, with a single relay, the DDF protocol dominates the class of AF protocols for all multiplexing gains.

The superiority of the DDF protocol is shown to be more significant in the CB and MAR channels. The situation is reversed in the CMA scenario, where we propose a novel AF protocol that achieves the optimal tradeoff for all multiplexing gains. This result highlights the fundamental difference between the relay and CMA channels.

We also consider ARQ cooperative channels where users are provided with ACK/ NACK signals indicating success or failure of destination in decoding their messages. We show that utilization of ARQ techniques not only improves the tradeoff achieved by non-ARQ protocols such as DDF relay and MAR, but also provides novel opportunities for cooperation that are otherwise unavailable. This is, for example, the case with the cooperative vector multiple-access (CVMA) channel where the destination is equipped with multiple receiving-antennas. As we will see, achieving the full-rate and full-diversity in this channel is only possible through ARQ techniques.

A distinguishing feature of the protocols proposed in this dissertation is that they do not rely on orthogonal subspaces, allowing for a more efficient use of resources. In fact, based on our results one can argue that the sub-optimality of previously proposed protocols stems from their use of orthogonal subspaces rather than the half-duplex constraint.

We also provide a better understanding of the asymptotic relationship between the probability of error, transmission rate, and signal-to-noise ratio, as compared to what DMT offers. In particular, we identify the limitation imposed by the multiplexing gain notion and provide a new formulation for the throughput-reliability tradeoff (TRT) that avoids this limitation. The new characterization is then used to elucidate the asymptotic trends exhibited by the outage probability curves of MIMO channels.

To my mother, Mrs. Fereshteh Mahabadi and my late father, Dr. Khodarahm Azarian Yazdi, for their love and support.

ACKNOWLEDGMENTS

First and for most, I thank my parents, sister and brother for their love and support. I also thank them for the motivation and inspiration that they gave me to do well in my studies.

I am very grateful to my advisor, Prof. Hesham El Gamal, for his enthusiasm, guidance, confidence in my abilities, and placing a priority on my research in the midst of a busy schedule. I also would like to thank my co-advisor, Prof. Philip Schniter, for the innumerable things that I have learned from him. This thesis would not have taken shape and I would not have completed the Ph.D program without their constant motivation, guidance and support. In addition I would like to thank Prof. Randolph L. Moses and Prof. Andrea Serrani, for agreeing to be in my thesis committee, and for providing me with valuable comments and feedback, all through the program.

I would like to thank all the people I have interacted with at The Ohio State University, specifically everyone associated with the IPS laboratory. They created a wonderful environment for conducting research. I also would like to thank my Iranian friends at Ohio State. Time spent together with them has always been so much fun. I also thank Ms. Jeri McMichael, IPS administrative assistant, for being so kind and helpful to me.

VITA

August, 1973	Born - Tehran, Iran
1996	B.Sc. Electrical Engineering, Shahid Beheshti University, Tehran, Iran
1999	M.Sc. Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
2002-2006	Graduate Research Associate, The Ohio State University.

PUBLICATIONS

Research Publications

1. K. Azarian, H. El Gamal and P. Schniter, "On the Achievable Diversity-Multiplexing Tradeoff in Half-Duplex Cooperative Channels," *IEEE Trans. Info. Theory*, vol. 51, no. 12, Dec. 2005, pp. 4152-4172.

2. K. Azarian and H. El Gamal, "The Throughput-Reliability Tradeoff in MIMO Channels," *IEEE Trans. Info. Theory*, accepted for publication subject to revisions, Aug. 2005.

3. K. Azarian, H. El Gamal and P. Schniter, "On the Optimality of ARQ-DDF Protocols," *IEEE Trans. Info. Theory*, submitted, Jan. 2006.

4. A. Murugan, K. Azarian and H. El Gamal, "Cooperative Lattice Coding and Decoding," *IEEE JSAC Special Issue on Cooperative Communications and Networking*, submitted, Feb. 2006.

5. K. Azarian and H. El Gamal, "On the Utility of a 3dB SNR Gain in MIMO Channels," *2006 IEEE International Symposium on Information Theory*, 9-14 Jul., Seattle, WA.

6. K. Azarian and H. El Gamal, "What Does a 3dB Buy in MIMO Channels?" **Invited Paper**, *2006 UCSD Workshop on Information Theory and Its Applications, 6-10 Feb 2006*, La Jolla, CA.

7. K. Azarian and H. El Gamal, "Cooperation in Outage-limited Channels," **Invited Paper**, *2006 IEEE International Zurich Seminar on Communications.*

8. K. Azarian, H. El Gamal, "Beyond the Multiplexing-Gain: The Throughput-Reliability Tradeoff," *Allerton Conf. on Communication, Control, and Computing*, 2005, Monticello, IL.

9. K. Azarian, Y. Nam and H. El Gamal, "Multi-User Diversity without Transmitter CSI," *2005 IEEE International Symposium on Information Theory*, 4-9 Sept. 2005, Adelaide, Australia, pp. 2055-2059.

10. K. Azarian, H. El Gamal , "From Diversity-Multiplexing to Diversity-Rate Tradeoff," **Invited Talk**, *2005 IEEE Communication Theory Workshop (CTW05)*, June 12-15 2005, Park City, Utah.

11. Y. Nam, K. Azarian and H. El Gamal , "Cooperation Through ARQ," **Invited Paper**, *2005 IEEE 6th Workshop on Signal Processing Advances in Wireless Communications (SPAWC05)*, June 5-8 2005, New York City, New York, pp. 1023-1027.

12. K. Azarian, H. El Gamal and P. Schniter, "Achievable Diversity-vs-Multiplexing Tradeoffs in Half-Duplex Cooperative Channels," *Proc. 2004 IEEE Information Theory Workshop*, Oct. 24-29 2004, San Antonio, TX, pp. 292-297.

13. K. Azarian, H. El Gamal and P. Schniter, "On the Achievable Diversity-Multiplexing Tradeoff in Half Duplex Cooperative Channels," **Invited Paper**, *Proc. Allerton Conf. on Communication, Control, and Computing*, Oct. 2004, Monticello, IL.

14. K. Azarian, H. El Gamal and P. Schniter, "On the Achievable Diversity-vs-Multiplexing Tradeoff in Cooperative Channels" *Proc. Conference on Information Sciences and Systems*, Mar. 2004, Princeton, NJ.

15. K. Azarian, H. El Gamal, and P. Schniter, "On the Design of Cooperative Transmission Schemes," *Proc. Allerton Conf. on Communication, Control, and Computing*, Oct. 2003, Monticello, IL.

FIELDS OF STUDY

Major Field: Electrical and Computer Engineering

Studies in:

Comm. and Signal Proc.	Prof. Hesham El Gamal
Comm. and Signal Proc.	Prof. Philip Schniter
Comm. and Signal Proc.	Prof. Randolph L. Moses
Control Theory	Prof. Andrea Serrani

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation

Recently, there has been a growing interest in the design and analysis of cooperative transmission protocols [1]-[16]. This is partly due to the ever-increasing popularity of potential applications such as cellular communications, wireless LANs and ad-hoc networks, and partly due to the attainable performance gains in terms of improved throughput, reliability and energy-efficiency. While user cooperation finds applications in a rather wide range of scenarios (e.g., fading vs. AWGN, ergodic vs. quasi-static and full-duplex vs. half-duplex), the basic idea is invariably that users, through pooling together their transmission resources such as power and antennas, can achieve a better performance compared to the case when they operate individually.

The term *cooperative diversity* appeared first in the work by Sendonaris, Erkip and Aazhang [1, 2]. In this work, a multiple-access (MA) scenario was considered where each user, through employing spreading codes, had been devoted a number of *orthogonal* sub-channels. As a result of orthogonality, a user's signal could be received at the end of the sub-channel, without interference from other users. Each

user then devoted one of its sub-channels for communication to the destination, and the rest for communication with the other users. This way, users could transmit their own messages and at the same time help others by relaying (an estimation of) their messages. The performance of this protocol was analyzed in terms of achievable rate-region and outage probability. Although this work takes the credit for introducing the concept of cooperative diversity, the proposed protocol suffered from unnecessary complications induced by the spreading codes, which in turn made the analysis not very straightforward. Also, some of the assumptions made, such as availability of channel state information (CSI) at the users, or, users being able to simultaneously transmit and receive (i.e. full-duplex communication), were not very practical.

The next and maybe the most influential work in the context of cooperative communications was the one by Laneman, Tse and Wornell [3]. One of the great contributions of this work was to provide a neat system model, which, together with the intuitive protocols proposed, inspired many people for conducting follow-on research (e.g., [5, 6]). Laneman *et al.*, too, considered a (two-user) MA scenario, however, they restricted the users to be half-duplex (i.e., no simultaneous transmission and reception was allowed). Furthermore, they adopted a coherent communication model, meaning that CSI was not available to the users. In this work, too, the channel was divided into (two) orthogonal sub-channels, however instead of using spreading codes, a Time Division Multiple Access (TDMA) framework was used. This made both understanding and analyzing of protocols much easier. The basic idea behind the protocols proposed in this work was to leverage the antenna available at the other user as a source of *virtual* spatial diversity. The proposed protocols were broadly classified as either amplify and forward (LTW-AF), where the helping node retransmitted a scaled version of its

soft observation, or decode and forward (LTW-DF), where the helping node decoded the information stream before repeating it. It was shown in [3] that the LTW-AF protocol achieved full-diversity gain (i.e., two levels of diversity for two users). However, the LTW-DF protocol failed in achieving a diversity gain greater than one. This, as was shown later, was a consequence of *not* encoding the information stream, rather than the protocol itself; In [4], Laneman and Wornell generalized the LTW-DF protocol into arbitrary number of users (LW-STC). Furthermore, they used a space-time code, originally designed for a multiple-input multiple-output (MIMO) channel, to encode the information stream in a distributed fashion. They then showed that the LW-STC protocol (and therefore the *coded* version of LTW-DF protocol) achieved full cooperative diversity, in the number of cooperating users.

Other follow-on works mainly focused on incorporation of various types of channel codes (i.e., outer codes such as convolutional and turbo codes), into the LTW-DF protocol in an attempt to exploit the promised information-theoretic gains (e.g., [5, 6]). It is important to realize that since these protocols use LTW-DF protocol's underlying signalling (i.e., inner code), from an information-theoretic point of view, they are *not* new protocols.

It was observed by Laneman *et. al.*, that the protocols proposed in [3, 4] (and subsequently the follow-on works) suffered from a significant loss of performance in the high spectral efficiency scenario. In fact, the authors of [3] posed the following open problem: *"a key area of further research is exploring cooperative diversity protocols in the high spectral efficiency regime."* This remark motivates our work here, where we present efficient (and in many cases optimal) AF and DF protocols for the relay,

cooperative broadcast (CB), multiple-access relay (MAR) and cooperative multiple-access (CMA) channels, along with their automatic repeat request (ARQ) variants.

Another difficulty in the area of (cooperation) protocol design is that, except for the simplest protocols, traditional performance evaluation techniques such as bit-error-rate and outage probability analysis, become mathematically intractable. To get around this problem, we adopt the diversity-multiplexing tradeoff (DMT) as our analysis tool. This powerful tool was introduced by Zheng and Tse for point-to-point MIMO channels [17] and later used by Tse, Viswanath, and Zheng to study the (non-cooperative) MA channel [18]. It should be noted, however, that there is no established relation, in the literature, between the DMT of a protocol and its *fixed-rate* outage probability. This may cause difficulties in comparing two protocols' performances based on their tradeoff curves. As an example, consider the relay protocol proposed by Nabar, Bolcskei and Kneubuhler (NBK-DF) [7]. In this protocol, the source splits its message into two parts. It then transmits the first part to the destination and the relay during the first phase. During the second phase, the source transmits the second part while the relay retransmits the first part. It is immediate to realize that since the second part of the message is not being relayed, this protocol does not achieve a diversity gain greater than one. Yet, the *upper-bound* on the DMT achieved by this protocol in the vicinity of $r = 1$[1], as characterized by [8], is maximal among all known protocols (refer to Fig. 1.1). Now, the natural question to ask is whether there exist any circumstances, under which this maximality (for multiplexing gains close to one) translates into superiority (in terms of outage probability) of this protocol over protocols that achieve full-diversity. With today's understanding

[1]The multiplexing gain r will be defined rigorously in Chapter 2

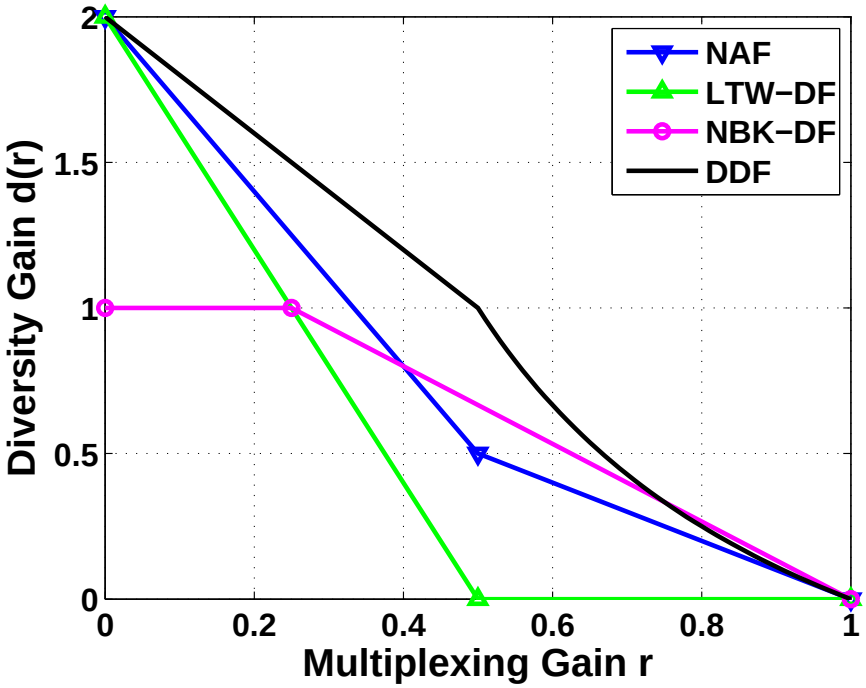

Figure 1.1: Comparison between the DMT's achieved by the LTW-DF, NAF, DDF and NBK protocols for the relay channel.

of DMT, giving an answer to this question is very difficult, if not impossible. As we will see in Chapter 3, this same problem shows up when we try to choose the segment lengths in a variant of the dynamic decode and forward (DDF) protocol such that the outage probability is minimized (refer to Fig. 1.2). These observations motivate us to examine the relation between outage probability and DMT of a protocol more closely. This investigation results in the proposal of the throughput-reliability trade-off (TRT), which fully characterizes the aforementioned relationship in MIMO and a few other channels (i.e., V-BLAST, orthogonal constellations and ARQ MIMO).

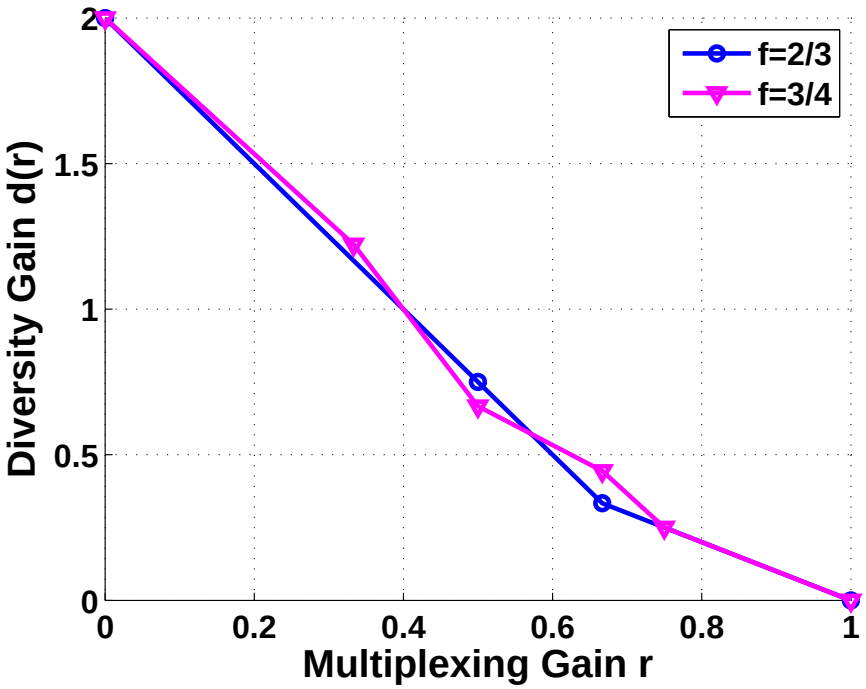

Figure 1.2: Comparison between DMT's achieved by the DDF relay protocol with three segments (fractions are $f_1 = \frac{1}{2}$ and $f_2 = f$).

1.2 Contributions and Outline

In the sequel, we give the dissertation outline and its main contributions.

In Chapter 2, we detail our modeling assumptions and review some of the results that are extensively used in the rest of the dissertation.

In Chapter 3, we consider the AF and DF relay channels and propose protocols for each one. The main contributions in this chapter can be summarized as follows.

- For the AF relay channel (single relay), we first establish an upper bound on the achievable DMT. We then propose a nonorthogonal amplify and forward (NAF) protocol that achieves this upper bound. Finally, we generalize the NAF protocol to the case of arbitrary number of relays and characterize its DMT. Notably, we show that the NAF protocol outperforms the LW-STC protocol without requiring decoding/encoding at the relays.

- For the DF relay channel (single relay), we propose a dynamic decode and forward (DDF) protocol and prove its *optimality* (with respect to DMT), over the range of multiplexing gains $1/2 \geq r \geq 0$. Furthermore, we show that the DDF protocol outperforms all AF protocols at any multiplexing gain. Finally, we extend the DDF protocol to the case of arbitrary number of relays and characterize its tradeoff curve.

- We present a novel variant of the (single relay) DDF protocol, which is particularly suitable for implementation. In this variant, the channel seen by the destination is reduced to a single-input single-output (SISO) time-selective channel. Furthermore, to reduce the complexity of the protocol, we confine the relay to start transmission only at the beginning of a *finite* number of segments. We

give the rule for determining these segments such that the variant is Pareto[2] optimal with respect to DMT.

In Chapter 4, we propose protocols for the CB, MAR and CMA channels. The main contributions in this chapter are as follows.

- For the CB channel, we present a variant of the DDF protocol that allows for efficient transmission of common information. We then establish the superiority of this protocol, over AF protocols, by characterizing its tradeoff curve. In fact, it turns out that the gain offered by the DDF protocol is more significant in this scenario, when compared to the relay channel.

- For the symmetric MAR channel (two users), we first derive an upper bound on the achievable DMT. We then modify the DDF protocol to match it to the MAR channel and derive its tradeoff curve. This characterization shows that the DDF protocol is DMT optimal over the range of multiplexing gains $3/4 \geq r \geq 0$. It also reveals that in the MAR channel, a *single* relay can be utilized by *several* users to simultaneously improve the diversity gain achieved by *all* of them.

- For the symmetric CMA channel, we propose a novel AF protocol where an *artificial* inter-symbol interference (ISI) channel is created. We prove the optimality (in terms of DMT) of this protocol by showing that, for all multiplexing gains (i.e., $1 \geq r \geq 0$), it achieves the DMT of the corresponding $N \times 1$ MIMO channel. We then use this result to argue that the sub-optimality of the protocols proposed in [3] is a consequence of using orthogonal sub-channels, rather

[2]The notion of Pareto optimality will be rigorously defined in Chapter 2.

than the half-duplex constraint. We also utilize this result to elucidate the fundamental difference between the relay and CMA channels.

In Chapter 5, we consider the ARQ relay, MAR and cooperative vector multiple-access (CVMA) channels and quantify the significant performance gains, attained through providing the users with a few feedback bits. The main contributions in this chapter are as follows.

- For the ARQ relay and MAR (two users) channels, we first modify the corresponding DDF protocols (proposed in Chapter 3 and Chapter 4, respectively) to incorporate ACK/NACK feedback signals. We then characterize their achieved tradeoff curves, which prove their optimality.

- For the ARQ CVMA channel (two users and two receiving antennas), we develop a new variant of the DDF protocol where the users are *purposefully* instructed not to cooperate in the first round of transmission. Lower and upper bounds on the achievable DMT are then derived. These bounds are shown to converge to the optimal tradeoff as the number of transmission rounds increases.

In Chapter 6, we consider an outage limited MIMO channel and build on Zheng and Tse's elegant formulation of DMT to develop a better understanding of the asymptotic relationship between the probability of error, transmission rate, and signal to noise ratio (SNR). The main contributions in this chapter are summarized in the following.

- We identify the limitation imposed by the notion of multiplexing gain and develop a new formulation for the throughput-reliability tradeoff that avoids this

limitation. In this formulation, the multiplexing gain notion is replaced by the more general concept of *operating* regions.

- We use the proposed TRT formulation to elucidate the asymptotic trends exhibited by MIMO channels. In particular, we devise a piecewise linear approximation to the outage probability of MIMO channels (at *fixed* rates), which becomes progressively more accurate as rate and SNR grow.

- We characterize the TRT, along with the corresponding piecewise linear outage approximation, for the V-BLAST, orthogonal constellations and ARQ MIMO channels.

Finally in Chapter 7, we offer some concluding remarks. To enhance the flow of the dissertation, we collect all the proofs in the Appendix.

CHAPTER 2

BACKGROUND

Throughout the dissertation, we use $(x)^+$ to mean $\max\{x, 0\}$, $(x)^-$ to mean $\min\{x, 0\}$, $\lceil x \rceil$ to mean nearest integer to x towards plus infinity and $\lfloor x \rfloor$ to mean nearest integer to x towards minus infinity. $\mathbb{R}^N$ and $\mathbb{C}^N$ denote the set of real and complex N-tuples, respectively, while $\mathbb{R}^{N+}$ denotes the set of non-negative N-tuples. We denote the complement of set $O \subseteq \mathbb{R}^N$, in $\mathbb{R}^N$, by O^c, while O^+ means $O \cap \mathbb{R}^{N+}$. I_N denotes the $N \times N$ identity matrix, $\Sigma_{\mathbf{x}}$ denotes the autocovariance matrix of vector $\mathbf{x}$, and $\log(.)$ denotes the base-2 logarithm.

Next, we state the general assumptions that apply to all of the *cooperative* channels considered in this dissertation (A separate system model will be given for the outage-limited MIMO channels considered in Chapter 6). Assumptions pertaining to a specific scenario will be given in the related Chapter.

- All channels are assumed to be flat Rayleigh-fading and quasi-static, i.e., the channel gains remain constant during a coherence-interval and change independently from one coherence-interval to another. Furthermore, the channel gains are mutually independent with unit variance. The additive noises at different nodes are zero-mean, mutually-independent, circularly-symmetric and

white complex-Gaussian. Furthermore, the variances of these noises are proportional to one another such that there are always *fixed* offsets between the different channels' SNRs.

- All nodes have the same power constraint, operate synchronously and unless otherwise stated, have a single antenna (In Section 5.2 we consider the CVMA scenario where the destination is equipped with two receiving antennas). Throughout the dissertation, we consider coherent communications meaning that only the receiving node of any link knows the channel gain. Also, except for Chapter 5 (where we study ARQ channels), no feedback to the transmitting node is permitted. Following in the footsteps of [3], all cooperating partners operate in the half-duplex mode, i.e., at any point in time, a node can either transmit or receive, but not both. This constraint is motivated by, e.g., the typically large difference between the incoming and outgoing signal power levels.

- In this work, we exclusively use random Gaussian code-books where a codeword spans the entire coherence-interval of the channel. Furthermore, we assume asymptotically large code-lengths, implying that the established DMTs only serve as upper-bounds on the performance of the protocols that use finite code-lengths. Results related to the design of practical coding/decoding schemes that approach the fundamental limits established here are reported in [31].

In the sequel, we summarize several important definitions and results that will be used throughout the dissertation.

- The SNR of a link, ρ, is defined as

$$\rho \triangleq \frac{E}{\sigma^2}, \tag{2.1}$$

where E denotes the average energy available for transmission of a symbol across the link and σ^2 denotes the variance of the noise observed at the receiving end of the link. We say that $f(\rho)$ is *exponentially equal to* ρ^b, denoted by $f(\rho)\dot{=}\rho^b$, when

$$\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log \rho} = b. \tag{2.2}$$

In (2.2), b is called the *exponential order* of $f(\rho)$. $\dot{\leq}$ and $\dot{\geq}$ are defined similarly.

- Assume that g is a Gaussian random variable with zero mean and unit variance. If v denotes the exponential order of $1/|g|^2$, i.e.,

$$v = -\lim_{\rho\to\infty} \frac{\log |g|^2}{\log \rho}, \tag{2.3}$$

then the probability density function (PDF) of v can be shown to be:

$$p_v = \lim_{\rho\to\infty} \ln(\rho)\rho^{-v} \exp(-\rho^{-v}).$$

Careful examination of the previous expression reveals that

$$p_v \dot{=} \begin{cases} \rho^{-\infty} = 0, & \text{for } v < 0, \\ \rho^{-v}, & \text{for } v \geq 0 \end{cases}. \tag{2.4}$$

Thus, for independent random variables $\{v_j\}_{j=1}^N$ distributed identically to v, the probability P_O that $(v_1, \ldots, v_N)$ belongs to set O can be characterized by

$$P_O \dot{=} \rho^{-d_o} \quad \text{where} \quad d_o \triangleq \inf_{(v_1,\ldots,v_N)\in O^+} \sum_{j=1}^{N} v_j, \tag{2.5}$$

provided that O^+ is not empty. In other words, the exponential order of P_O only depends on O^+. This is due to the fact that the probability of any set, consisting of N-tuples $(v_1, \ldots, v_N)$ with at least one negative element, decreases exponentially with SNR and therefore can be neglected compared to P_{O^+} which decreases polynomially with SNR.

- Consider a family of codes $\{C_\rho\}$ indexed by operating SNR ρ, such that the code C_ρ has a rate of $R(\rho)$ bits per channel use (BPCU) and a maximum likelihood (ML) error probability $P_E(\rho)$. For this family, the *multiplexing gain* r and the *diversity gain* d are defined as

$$r \triangleq \lim_{\rho\to\infty} \frac{R(\rho)}{\log \rho}, \qquad\qquad d \triangleq -\lim_{\rho\to\infty} \frac{\log P_E(\rho)}{\log \rho}. \tag{2.6}$$

- The problem of characterizing the optimal tradeoff between the reliability and throughput of a point-to-point communication system over a coherent quasi-static flat Rayleigh-fading channel was posed and solved by Zheng and Tse in [17]. For a MIMO communication system with M transmit and N receive antennas, they showed that, for any $\min\{M, N\} \geq r \geq 0$, the optimal diversity gain $d^*(r)$ is given by the piecewise linear function joining the (r, d) pairs $(k, (M-k)(N-k))$ for $k = 0, ..., \min\{M, N\}$, provided that the code-length l satisfies $l \geq M + N - 1$.

- We say that protocol A is *uniformly optimal*, if for every protocol B, $d_A(r) \geq d_B(r), \forall r$.

- We say that protocol A is *Pareto optimal*, if there is no protocol B dominating protocol A in the Pareto sense. Protocol B is said to dominate protocol A in the Pareto sense if there is some r_0 for which $d_B(r_0) > d_A(r_0)$, but no r such that $d_B(r) < d_A(r)$.

- Consider a coherent linear Gaussian channel, i.e.

$$\mathbf{y} = \mathbf{s} + \mathbf{n},$$

where, $\mathbf{s} \in \mathbb{C}^N$ and $\mathbf{n} \in \mathbb{C}^N$ denote the signal and noise components of the observed vector, respectively. For this channel, the pairwise error probability (PEP) of the ML decoder, denoted by P_{PE}, averaged over the ensemble of random Gaussian codes, is upper bounded by

$$P_{PE} \leq \det(I_N + \frac{1}{2}\Sigma_{\mathbf{s}}\Sigma_{\mathbf{n}}^{-1})^{-1}. \tag{2.7}$$

- The following lemma will be used in characterizing the DMT of DDF protocols.

Lemma 1 *Consider a coherent linear Gaussian channel of data rate R and codeword length l. The error probability of the ML decoder utilizing a fraction of the code-word such that the mutual information between the received and transmitted signals exceeds lR, averaged over the ensemble of random Gaussian codes, can be made arbitrarily small provided that the codeword length l is sufficiently large.*

Proof: Please refer to Appendix A.1.

CHAPTER 3

RELAY CHANNEL

In this chapter, we consider the relay channel, where $N-1$ relays help a single source to better transmit its message to the destination. As the vague descriptions *help* and *better transmit* suggest, the general relay problem is rather broad and only certain sub-problems have been studied (for example see [24]). In this chapter, we focus on two important classes of relay protocols. The first is the class of amplify and forward (AF) protocols, where a relaying node can only process the observed signal *linearly* before re-transmitting it. The second is the class of decode and forward (DF) protocols, where the relays are allowed to decode and re-encode the message using (a possibly different) code-book. Here we emphasize that, a priori, it is not clear which class (i.e., AF or DF) offers a better performance (e.g., [3]).

3.1 Amplify and Forward Protocols

We first consider the single relay scenario (i.e., $N = 2$). For this scenario, we derive the optimal DMT and identify a specific protocol within this class, i.e., the NAF protocol that achieves this optimal tradeoff. We then extend the NAF protocol to the general case with an arbitrary number of relays.

Under the half-duplex constraint, it is easy to see that any single-relay AF protocol can be mathematically described by some choice of the matrices A_1, A_2, and B in the following model

$$\mathbf{y} = \begin{bmatrix} g_1 A_1 & 0 \\ g_2 h B A_1 & g_1 A_2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ g_2 B \end{bmatrix} \mathbf{w} + \mathbf{v}. \tag{3.1}$$

In (3.1), $\mathbf{y} \in \mathbb{C}^l$ represents the vector of observations at the destination, $\mathbf{x} \in \mathbb{C}^l$ the vector of source symbols, $\mathbf{w} \in \mathbb{C}^{l'}$ the vector of noise samples (of variance σ_w^2) observed by the relay, and $\mathbf{v} \in \mathbb{C}^l$ the vector of noise samples (of variance σ_v^2) observed by the destination. The variables h, g_1 and g_2 denote the source-relay channel gain, source-destination channel gain, and relay-destination channel gain, respectively. $A_1 \in \mathbb{C}^{l' \times l'}$ and $A_2 \in \mathbb{C}^{(l-l') \times (l-l')}$ are diagonal matrices. In this protocol, the source can potentially transmit a new symbol in every symbol-interval of the codeword, while the relay listens during the first l' symbols and then, for the remaining $l - l'$ symbols, transmits linear combinations of the l' noisy observations using the coefficients in $B \in \mathbb{C}^{(l-l') \times l'}$. In fact, by letting $l' = l/2$, $A_1 = I_{l'}$, $A_2 = 0$ and $B = bI_{l'}$ (with $b \leq \sqrt{E/(|h|^2 E + \sigma_w^2)}$ denoting the relay repetition gain), we obtain the LTW-AF protocol [3]. Finally, we note that when the source symbols are independent, the average energy constraint translates to

$$|h|^2 E \sum_{i=1}^{l'} |b_{ji}|^2 |a_i|^2 + \sigma_w^2 \sum_{i=1}^{l'} |b_{ji}|^2 \leq E, \quad j = 1, \ldots, l - l', \tag{3.2}$$

where $B = [b_{ji}]$ and $A_1 = \mathrm{diag}(a_1, \cdots, a_{l'})$.

Theorem 2 *The optimal diversity gain for the relay channel with a single AF relay is upper-bounded by*

$$d^*(r) \leq (1 - r) + (1 - 2r)^+. \tag{3.3}$$

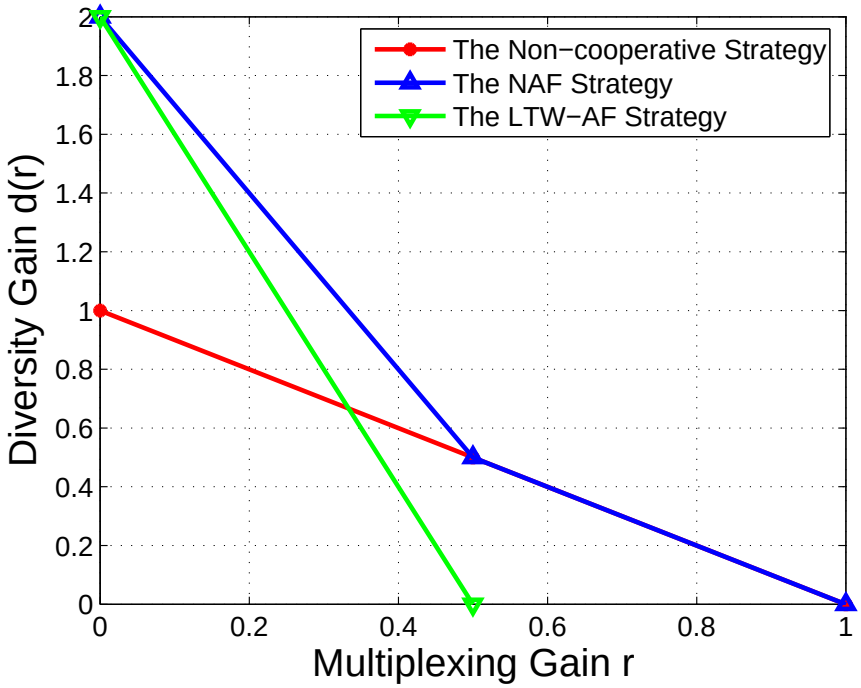

Figure 3.1: The optimal DMT for a single-relay AF protocol.

Proof: Please refer to Appendix A.2.

The upper-bound on $d^*(r)$, as given by (3.3), is shown in Fig. 3.1. Having Theorem 2 at hand, it now suffices to identify an AF protocol that achieves this upper-bound in order to establish its optimality. Towards this end, we observe that, in the proof of Theorem 2, the only requirements on B such that the protocol described by (3.1) could *potentially* achieve the optimal DMT are that B is square (of dimension $l/2 \times l/2$) and full-rank. Furthermore, B should not violate the relay average energy constraint as given by (3.2). Thus, the simple choices

$$A_1 = I_{l/2} \qquad A_2 = I_{l/2} \qquad B = bI_{l/2} \;\text{ for }\; b \leq \sqrt{\frac{E}{|h|^2 E + \sigma_w^2}} \tag{3.4}$$

inspire the NAF protocol. In particular, the source transmits on every symbol-interval in a cooperation frame, where a cooperation frame is defined as two consecutive symbol-intervals. The relay, on the other hand, transmits only once per cooperation frame; it simply repeats the (noisy) signal it observed during the previous symbol-interval. It is important to realize that this design is dictated by the half-duplex constraint, which implies that the relay can repeat at most once per cooperation frame. We denote the repetition gain by b and, for frame k, we denote the information symbols by $\{x_{j,k}\}_{j=1}^{2}$. The signals received by the destination during frame k are thus:

$$\begin{aligned} y_{1,k} &= g_1 x_{1,k} + v_{1,k} \\ y_{2,k} &= g_1 x_{2,k} + g_2 b(h x_{1,k} + w_{1,k}) + v_{2,k} \end{aligned}$$

where the repetition gain b satisfies (3.4). Note that, in order to decode the message, the destination needs to know the relay repetition gain b, the source-relay channel gain h, the source-destination channel gain g_1, and the relay-destination channel gain

g_2. Now, we are ready to establish the optimality of the NAF protocol with respect to DMT.

Theorem 3 *The NAF protocol achieves the optimal DMT of the AF single-relay scenario, which is*

$$d^*(r) = (1 - r) + (1 - 2r)^+. \tag{3.5}$$

Proof: Please refer to Appendix A.3.

Three remarks are now in order:

1. As shown in Fig. 3.1, the NAF protocol enjoys uniform dominance over the direct transmission scheme (i.e., no cooperation) and LTW-AF protocol. This dominance can be attributed to relaxing the orthogonality constraint whereby one can reap two distinct benefits: rate enhancement via continuous transmission and diversity enhancement via cooperation. It is interesting to note that this dominance is achieved while only half of the symbols are repeated by the relay.

2. From Fig. 3.1, one can see that for multiplexing gains greater that 0.5, the diversity gain achieved by the NAF relay protocol is identical to that of the non-cooperative protocol. This is due to the fact that the *AF cooperative* link provided by the relay can not support multiplexing gains greater than 0.5–a consequence of the half-duplex constraint. Hence, for multiplexing gains larger than 0.5, there is only one link from the source to the destination, and thus, the tradeoff curve is identical to that of a point-to-point system with one transmit and one receive antenna. Later, we will show that the proposed DDF strategy avoids this drawback.

3. As shown in the proof of Theorem 3, the achievability of the optimal tradeoff is not very sensitive to the choice of the repetition gain "b" (i.e., for a wide range of choices, the NAF protocol achieves the optimal tradeoff). In practice, one should optimize the repetition gain, experimentally if needed, to minimize the outage probability at the target rate and SNR.

The NAF protocol can be extended to the case of arbitrary number of relays (i.e., $N \geq 2$) as follows. First, we define a super-frame as a concatenation of $N-1$ consecutive cooperation frames. Within each super-frame, the relays take turns repeating the signals they previously observed as they did in the case of a single relay (refer to Fig. 3.2). Thus, the destination's received signals during a super-frame will be

$$\begin{aligned}
y_{1,1} &= g_1 x_{1,1} + v_{1,1} \\
y_{2,1} &= g_1 x_{2,1} + g_2 b_2 (h_2 x_{1,1} + w_{1,1}) + v_{2,1} \\
y_{1,2} &= g_1 x_{1,2} + v_{1,2} \\
y_{2,2} &= g_1 x_{2,2} + g_3 b_3 (h_3 x_{1,2} + w_{1,2}) + v_{2,2} \\
&\vdots \\
y_{1,N-1} &= g_1 x_{1,N-1} + v_{1,N-1} \\
y_{2,N-1} &= g_1 x_{2,N-1} + g_N b_N (h_N x_{1,N-1} + w_{1,N-1}) + v_{2,N-1},
\end{aligned}$$

where the source-relay channel gain, relay-destination channel gain, relay repetition gain, and relay-observed noise for relay $i \in \{1, ..., N-1\}$ are denoted by h_{i+1}, g_{i+1}, b_{i+1} and $w_{1,i}$, respectively. As before, g_1 represents the source-destination channel gain. The quantities $y_{j,k}$, $v_{j,k}$, and $x_{j,k}$ represent the received signal, noise sample, and source symbol, respectively, during the j^{th} symbol-interval of the k^{th} cooperation frame. Note that there is nothing to be gained by having more than one relay

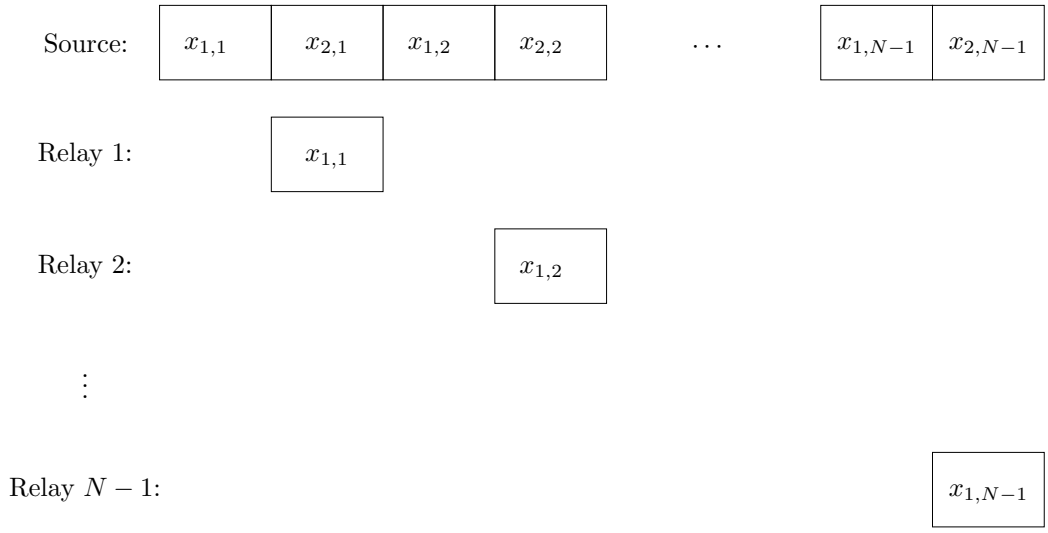

Figure 3.2: The super-frame in the NAF protocol with $N-1$ relays.

transmitting the same symbol simultaneously. Also, similar to the single-relay NAF scenario, the destination needs to know all relay repetition gains $\{b_i\}_{i=2}^{N}$ as well as all channel gains $\{g_i\}_{i=1}^{N}$ and $\{h_i\}_{i=2}^{N}$. The following theorem characterizes the DMT achieved by this protocol.

Theorem 4 *The DMT achieved by the NAF protocol with $N-1$ relays is characterized by*

$$d(r) = (1-r) + (N-1)(1-2r)^{+}.$$

Proof: The proof is virtually identical to that of Theorem 3, and hence, is omitted for brevity.

It is interesting to note that the generalized NAF protocol uniformly dominates the LW-STC [4]. This can be attributed to the fact that in the generalized NAF protocol, in contrast to the LW-STC protocol, the source transmits over the whole duration of

the codeword. The generalized NAF protocol offers the additional advantage of low complexity since it does not require decoding/encoding at the relays.

3.2 Decode and Forward Protocols

In this class of protocols, we allow for the possibility of decoding/encoding at the different relays. In [3], Laneman-Tse-Wornell presented a particular variant of DF protocols (LTW-DF) where the source transmits in the first half of the codeword. Based on its received signal in this interval, the relay attempts to decode the message. It then re-encodes and transmits the encoded stream in the second half of the codeword. In [4], Laneman and Wornell derived the DMT achieved by this scheme (i.e., $d(r) = 2(1-2r)$), which is depicted in Fig. 3.3. Here, we propose a dynamic decode and forward (DDF) protocol and characterize its tradeoff curve. This characterization reveals the uniform dominance of this protocol over all known *full-diversity* (i.e., $d(0) = 2$) protocols proposed for the half-duplex single-relay channel and furthermore establishes its optimality, over a certain range of multiplexing gains (i.e., $1/2 \geq r \geq 0$). We first describe and analyze the protocol for the case of a single relay. Generalization to $N-1$ relays will then follow.

We assume that a codeword consists of l consecutive symbol-intervals, during which all the channel gains remain unchanged. In the DDF protocol, the source transmits data at a rate of R BPCU during every symbol-interval in the codeword. The relay, on the other hand, listens to the source until the mutual information between its received signal and source signal exceeds lR. It then decodes and re-encodes the message using an *independent* Gaussian code-book and transmits it during the rest of the codeword. The dynamic nature of the protocol is manifested in the fact

that we allow the relay to listen for a time duration that depends on the instantaneous channel realization to maximize the probability of successful decoding. We denote the signals transmitted by the source and relay as $\{x_k\}_{k=1}^{l}$ and $\{\tilde{x}_k\}_{k=l'+1}^{l}$, respectively, where l' is the number of symbol-intervals the relay waits before starting transmission. Using this notation, the received signals (at the destination) can be written as:

$$y_k = \begin{cases} g_1 x_k + v_k & \text{for} \quad l' \geq k \geq 1 \\ g_1 x_k + g_2 \tilde{x}_k + v_k & \text{for} \quad l \geq k > l' \end{cases}.$$

From the protocol description, it is clear that the number of symbols where the relay listens should be chosen as:

$$l' = \min \left\{ l, \left\lceil \frac{lR}{\log\left(1 + |h|^2 c\rho\right)} \right\rceil \right\}, \tag{3.6}$$

where h is the source-relay channel gain, and $c = \sigma_v^2/\sigma_w^2$. One can now see the dependence of this choice of l' on the instantaneous channel realization and that this choice, together with the asymptotically large l, guarantees that when $l' < l$, the relay average probability of error with a Gaussian code ensemble is arbitrarily small (recall Lemma 1). Clearly, when $l' = l$ the relay does not contribute to the transmission of the message, and hence, incorrect decoding at the relay in this case does not affect performance. Here, we observe that, in contrast to the NAF protocol, the destination does not need to know the source-relay channel gain. It does, however, need to know the relay waiting time l', along with the source-destination and relay-destination channel gains. The following theorem describes the DMT achievable with this cooperation protocol.

Theorem 5 *The DMT achieved by the single relay DDF protocol is given by*

$$d(r) = \begin{cases} 2(1-r) & \text{if } \frac{1}{2} \geq r \geq 0 \\ (1-r)/r & \text{if } 1 \geq r \geq \frac{1}{2} \end{cases} . \tag{3.7}$$

Proof: Please refer to Appendix A.4.

The DMT described by (3.7) is shown in Fig. 3.3. It is now clear that the DDF protocol is optimal for $0.5 \geq r \geq 0$ since it achieves the genie aided diversity (where the relay is assumed to know the information message *a-priori*). For $r > 0.5$, the DDF protocol suffers from a loss, compared to the genie aided strategy, since, on the average, the relay will only be able to help during a small fraction of the codeword. It is easy to see that, the performance for this range of multiplexing gains can not be improved through employing a mixed AF and DF strategy. In fact, the DDF strategy dominates all such strategies[3]. It remains to be seen whether there exists a strategy that closes the gap to the genie aided strategy when $r > 0.5$ or not. Note also that the gain offered by the DDF protocol, compared to AF protocols, can be attributed to the ability of this strategy to transmit independent Gaussian symbols after successful decoding. In AF strategies, on the other hand, the relay is limited to repeating the noisy Gaussian symbols it receives from the source. Fig. 3.3 also compares the DDF protocol with the NBK-DF protocol proposed in [7]. In this comparison, we utilize the upper-bound derived by Prasad and Varanasi on the DMT of the NBK-DF, which was reported in [8]. One can see from Fig. 3.3 that the NBK-DF protocol does not achieve any diversity gain greater than one. This can be attributed to the fact that in this protocol, the message is split-up into two parts, out of which, only one is retransmitted by the relay. Fig. 3.3 also shows that for multiplexing gains close to one, the NBK-DF upper-bound outperforms the DDF protocol. Therefore, in this

[3]The proof for this is rather straightforward, and hence, is omitted here for brevity.

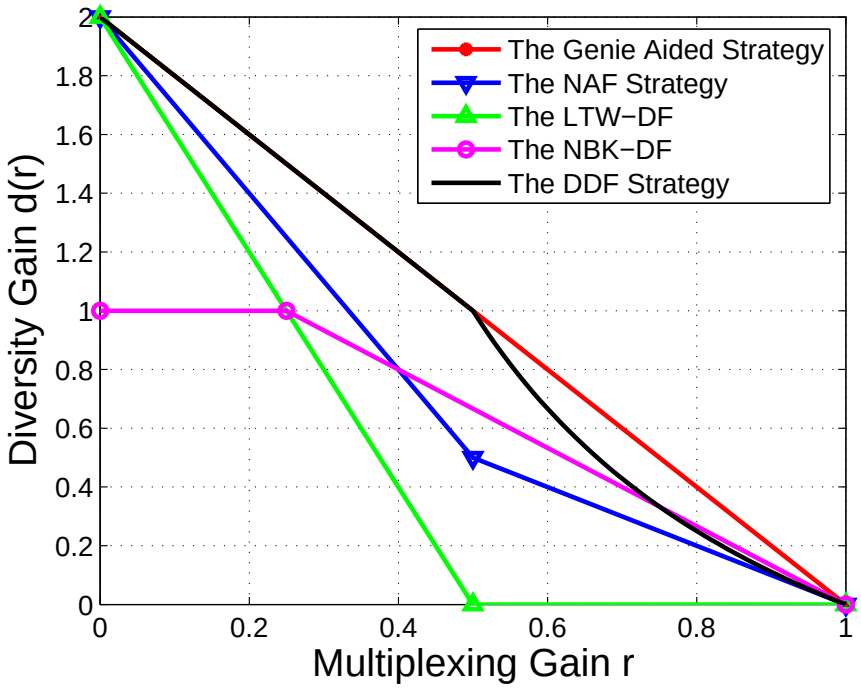

Figure 3.3: DMT for the DDF protocol with one relay.

range, the comparison between the two protocols depends on the tightness of the NBK-DF upper-bound which was not discussed in [8].

Next, we describe the generalization of the DDF protocol to the case of multiple relays. In this case, the source and relays cooperate in nearly the same manner as in the single relay case. Specifically, the source transmits during the whole codeword while each relay listens until the mutual information between its received signal and the signals transmitted by the source and other relays exceeds lR. It is assumed that every relay knows the code-books used by the source and other relays. Once a relay decodes the message, it uses an independent code-book to re-encode the message,

which it then transmits for the rest of the codeword. Note that, since the source-relay channel gains may differ, the relays may require different waiting times for decoding. This complicates the protocol, since a given relay's ability to decode the message requires precise knowledge of the times at which every other relay begins its transmission. To address this problem, the codeword is divided into a number of segments, and relays are allowed to start transmission only at the beginning of a segment. In between the segments, every relay is allowed to broadcast a (well protected) beacon, informing all other relays whether or not it will start transmission. Judicious choice of the segment length, relative to the codeword length, results in only a small loss compared to the genie-aided case, whereby all relays know all decoding times *a-priori*. Here, we assume that the number of segments is sufficiently large and the length of the beacon signals is much smaller than the segment length. Therefore, in characterizing the DMT achieved by this protocol, we ignore the losses associated with the beacons and the quantization of the starting times for the different relays.

Theorem 6 *The DMT achieved by the DDF protocol with $N-1$ relays is characterized by:*

$$d(r) = \begin{cases} N(1-r), & \frac{1}{N} \geq r \geq 0, \\ 1 + \frac{(N-1)(1-2r)}{1-r}, & \frac{1}{2} \geq r \geq \frac{1}{N}, \\ \frac{1-r}{r}, & 1 \geq r \geq \frac{1}{2}. \end{cases} \tag{3.8}$$

Proof: Please refer to Appendix A.5.

The DMT (3.8) is shown in Fig. 3.4 and Fig. 3.5 for different values of N. While the loss of the DDF protocol compared to the genie-aided protocol increases with N, it is not clear at the moment if this loss is due to the half-duplex constraint or due to the sub-optimality of the DDF strategy.

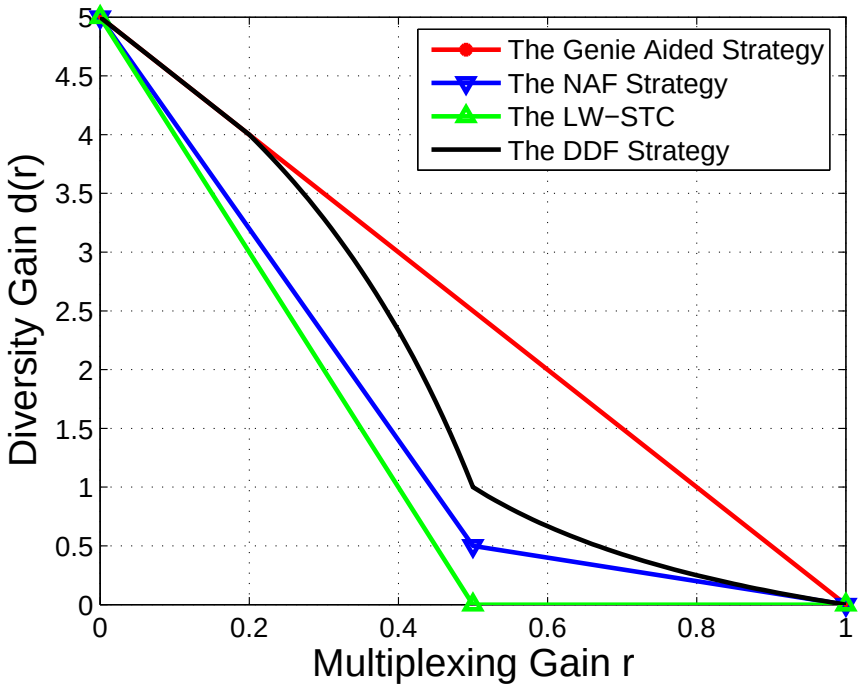

Figure 3.4: DMTs achieved by the NAF, DDF, LW-STC, and genie aided protocols with 4 relays.

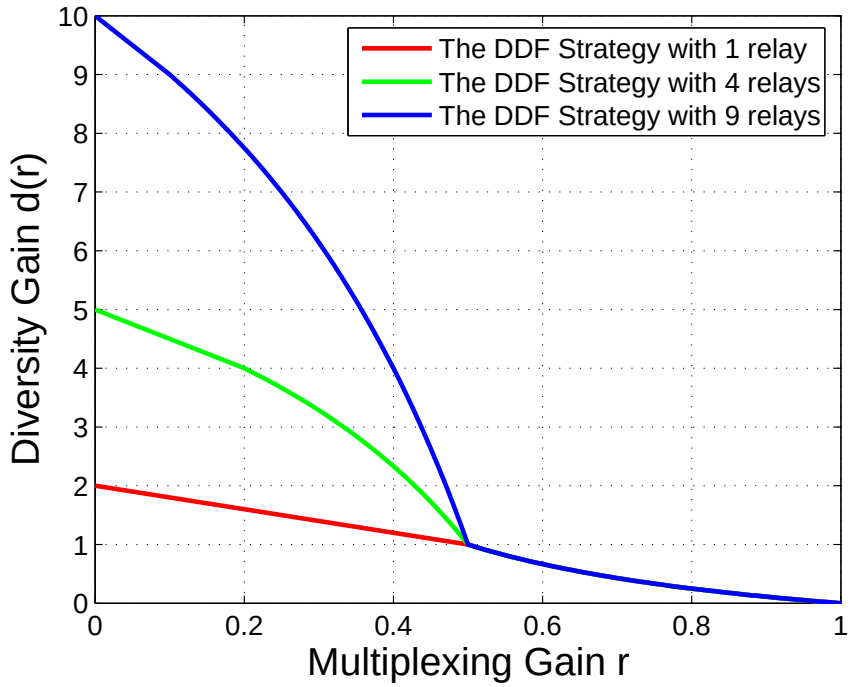

Figure 3.5: DMTs achieved by the DDF protocol with different number of relays.

The excellent performance of the DDF protocol (in terms of DMT), motivates us to develop low-complexity variants of this protocol that are particularly suited for implementation. In doing so, we take a step by step approach where we provide a number of lemmas that characterize the modifications needed for reducing the complexity of the protocol, while maintaining a good performance.

As the first step, we notice that the achievability result in Theorem 5 relies on using independent Gaussian code-books at the source and relay nodes. This approach potentially requires a computationally intensive algorithm at the destination to *jointly* decode the source and relay signals. Allowing the relay node to start transmission at any symbol-interval in the codeword (as characterized by (3.6)), is another potential source of complexity. In practice, this requires the source to use a very high-dimensional constellation (with a very low-rate code) such that the information stream can be uniquely decoded using only a few symbols (provided that the source-relay channel is sufficiently good). This same aspect also impacts the amount of overhead in the relay-destination packet, since the destination needs to be informed of the starting time of the relay. Next, we present two modifications to the DDF protocol that reduce the complexities associated with these two aspects.

1. Since the relay knows the source code-book, it can correctly anticipate the future transmissions from source (i.e., x_k for $l \geq k > l'$), once it successfully decodes the message. Based on this knowledge, the relay implements the following scheme, i.e.

$$\tilde{x}_k = \begin{cases} x_{k+1}^* & \text{for} \quad k = l'+1, l'+3, \cdots \\ -x_{k-1}^* & \text{for} \quad k = l'+2, l'+4, \cdots \end{cases}. \tag{3.9}$$

Note that (3.9) reduces the signal seen by the destination, for $l \geq k > l'$, to an Alamouti constellation.

2. We restrict the relay to start transmission only after the codeword is halfway through, i.e., we replace (3.6) with

$$l' = \min\left\{l, \max\left\{\frac{l}{2}, \left\lceil \frac{lR}{\log\left(1+|h|^2 c\rho\right)} \right\rceil\right\}\right\}. \tag{3.10}$$

As the following lemma shows, these modifications do not entail any loss (at least from the DMT perspective) in the performance of the variant.

Lemma 7 *Modifying the DDF protocol according to* (3.9) *and* (3.10) *does not effect the DMT achieved by the protocol (refer to Theorem 5).*

Proof: Please refer to Appendix A.6.

As shown in the proof of Lemma 7 (refer to (A.59)), the channel seen by the destination in the modified DDF protocol is a time-selective SISO. This allows for employing standard SISO decoding architectures, such as belief propagation and Fano decoding, at the destination. In addition, restricting the relay to start transmission according to (3.10) relaxes the constraint on the constellation-size, since the information stream now needs to be uniquely decodable only after the codeword is halfway through.

The next lemma investigates the effect of further restricting the relay to start transmission only at a *finite* number of symbol-intervals $\{l_j\}_{j=1}^{N}$. These symbol-intervals partition the codeword into $N+1$ segments, which are not necessarily equal in length. We further define the corresponding set of waiting fractions, $\{f_j\}_{j=0}^{N+1}$,

according to

$$f_j \triangleq \begin{cases} 0 & \text{for} \quad j = 0 \\ \frac{l_j}{l} & \text{for} \quad N \geq j \geq 1 \\ 1 & \text{for} \quad j = N+1 \end{cases}.$$

Thus

$$f_0 \triangleq 0 < f_1 < \cdots < f_N < f_{N+1} \triangleq 1.$$

The question now is, for a finite N, how to choose $\{f_j\}_{j=1}^N$ such that the protocol achieves the *optimal* DMT. The following lemma shows that this problem does not have a *uniformly* optimal solution and instead characterizes a *Pareto* optimal set of waiting fractions.

Lemma 8 *For the DDF protocol with $N+1$ segments,*

1. *there exists no uniformly optimal set of waiting fractions $\{f_j^u\}_{j=1}^N$.*

2. *let $f_1^p = \frac{1}{2}$ and*

$$f_j^p = \frac{1 - f_{j-1}^p}{2 - (1 + \frac{1}{f_N^p}) f_{j-1}^p}, \quad \textit{for} \quad N \geq j > 1 \tag{3.11}$$

then the set of waiting fractions $\{f_j^p\}_{j=1}^N$ is Pareto optimal, with

$$d^p(r) = 1 - r + (1 - \frac{r}{f_N^p})^+. \tag{3.12}$$

Proof: Please refer to Appendix A.7.

Fig. 3.6 shows the DMTs achieved by the Pareto optimal DDF protocols with $N = 1(\{\frac{1}{2}\})$, $N = 2(\{\frac{1}{2}, \frac{2}{3}\})$ and $N = \infty$.

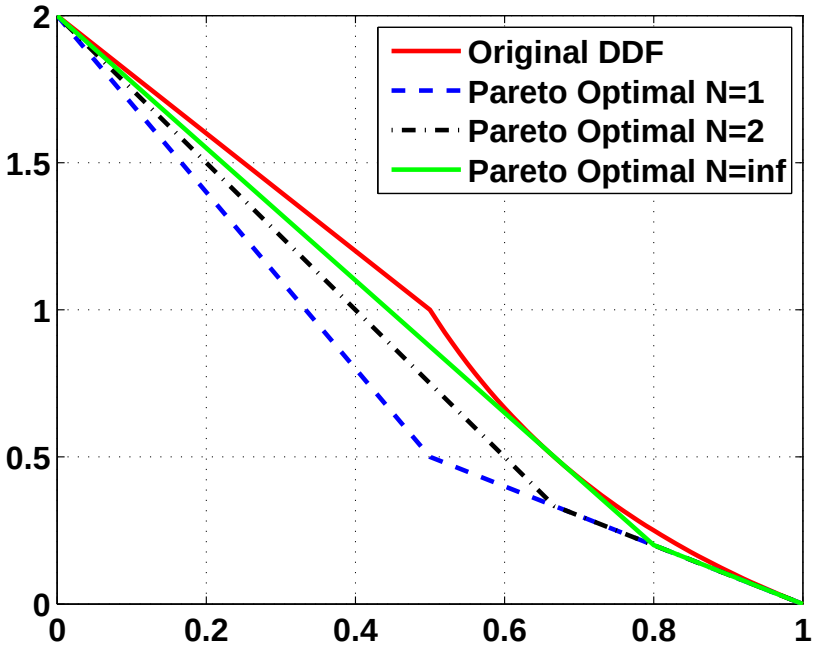

Figure 3.6: DMTs achieved by the Pareto optimal DDF protocols with $N = 1, 2$ and ∞.

3.3 Numerical Results

In this section, we report numerical results that quantify the performance gains offered by the NAF and DDF protocols. These numerical results correspond to outage probabilities and are meant to show that the superiority of the proposed protocols in terms of DMT translates into significant SNR gains (please refer to [30] and [31] for exemplar implementations of these protocols and performance measures in terms of bit and frame error rates). In Fig. 3.7 and Fig. 3.8, we compare the proposed protocols with the non-cooperative (direct transmission) and the LTW-AF protocols. To ensure fairness, we have imposed strict power constraints on the NAF and the DDF protocols; specifically, we lowered the average transmission energy of the source and the relay from E to $E/2$ during the interval when both are transmitting. This way, the total average energy per symbol-interval, spent by any of the protocols considered here is E. While one may find other energy allocation strategies that offer performance improvement (in terms of the outage probability), any such optimization will not affect the achievable DMT, and hence, will not be pursued here. To obtain a lower bound on the gain offered by the DDF protocol, we assume a noiseless source-relay channel for the LTW-AF and NAF protocols. For the DDF protocol, though, the SNR of the link between the two cooperating partners was assumed to be only 3 dB better than that of the relay-destination or source-destination channels. In all the considered cases, the outage probabilities are computed through Monte-Carlo simulations.

Fig. 3.7 shows the performance gain offered by the NAF protocol over both the non-cooperative protocol and the LTW-AF protocol at high SNRs and two different data rates. The same comparison is repeated in Fig. 3.8 with the DDF protocol where,

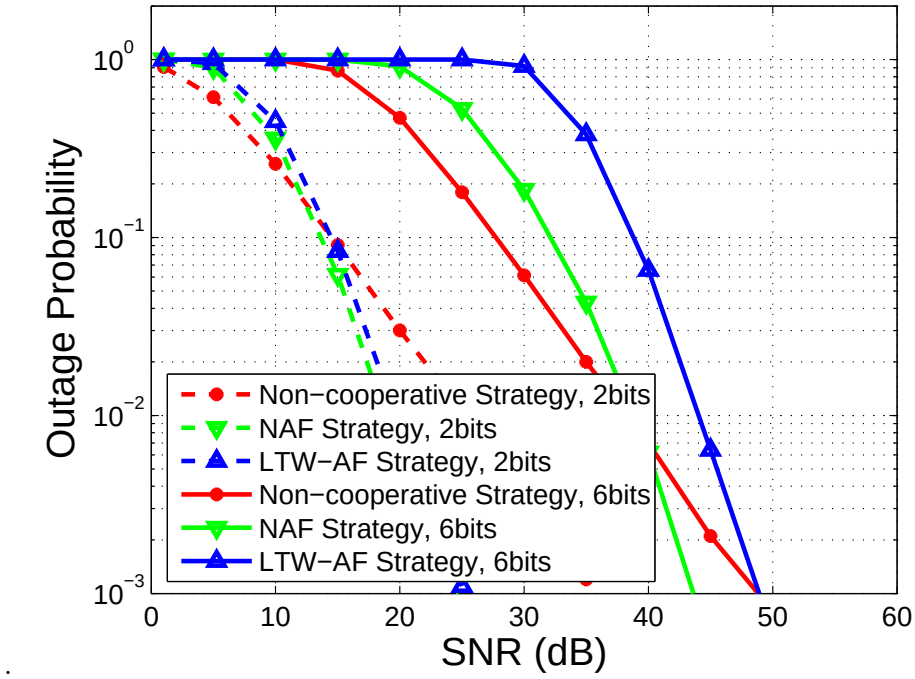

Figure 3.7: Comparison of the outage probability for the NAF, LTW-AF, and non-cooperative 1×1 protocols.

as expected, the gains are shown to be larger. We observe that the gain offered by the DDF protocol, compared with the LTW-AF protocol, increases with the data rate. This is a direct consequence of the higher multiplexing gains achievable with the DDF protocol. Overall, these results re-emphasize the fact that the full-diversity criterion (i.e., $d(0) = 2$) alone, is a rather weak design tool.

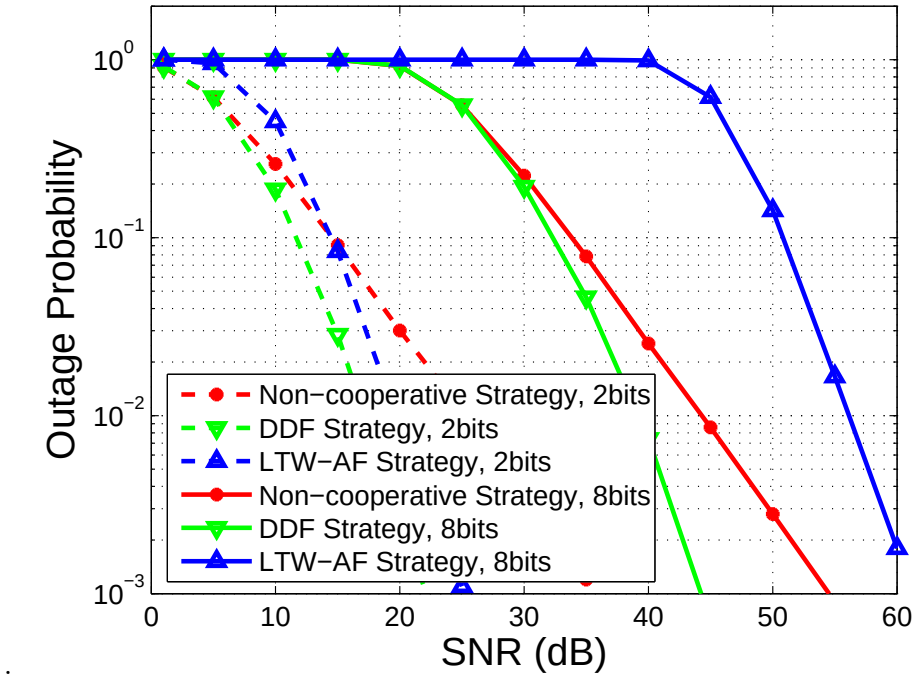

Figure 3.8: Comparison of the outage probability for the DDF, LTW-AF and non-cooperative 1×1 protocols.

CHAPTER 4

MULTIUSER COOPERATIVE CHANNELS

In this chapter we consider various multiuser cooperative channels (i.e., CB, MAR and CMA), and devise appropriate protocols for them.

4.1 Cooperative Broadcast Channel

In the CB scenario, a single source broadcasts to N destinations. This setup is different from a (non-cooperative) broadcast channel in that in the former scenario, the destinations are allowed to cooperate through helping one another in receiving their messages. We assume that the message intended for destination $j \in \{1, \cdots, N\}$ consists of two parts. A common part of rate $R_c = r_c \log \rho$ BPCU, which is intended for all of the destinations and an individual part of rate $R_j = r_j \log \rho$ BPCU, which is specific to the j-th destination. The total rate is then $R = R_c + \sum_{j=1}^{N} R_j$ and the multiplexing gain tuple is given by $\mathbf{r} = (r_c, r_1, ..., r_N)$. We define the overall diversity gain d based on the performance of the worst receiver, i.e.

$$d = \min_{N \geq j \geq 1} \{d_j\},$$

where we require all the receivers to decode the common information[4]. Now, as a first step, one can see that if $r_c = 0$, i.e., if there is no common message, then the

[4]Clearly this definition does not allow for different Quality of Service (QoS) constraints.

techniques developed for the relay channel can be *exported* to this setting through a proportional time sharing strategy. With this assumption, all properties of the NAF and DDF protocols, established in Chapter 3, carry over to this scenario. The problem becomes slightly more challenging when $r_c > 0$. In fact, it is easy to see that, for a fixed total rate R, the highest probability of error corresponds to the case where all destinations are required to decode all the messages. This translates to the following condition (that applies to any cooperation scheme)

$$d(r_c, r_1, r_2, ..., r_N) \geq d(r_c + r_1 + ... + r_N, 0, 0, .., 0).$$

So, we focus the following discussion on this worst case scenario, i.e.,

$$\mathbf{r} = (r_c, 0, 0, ..., 0), \quad \text{for} \quad 1 \geq r_c \geq 0.$$

As the first observation, we notice that in this scenario, the only AF strategy that achieves full-rate (i.e., $d(r) > 0, \forall r < 1$), is the non-cooperative protocol. Any other AF strategy will require some of the nodes to re-transmit, and therefore not to *listen* during parts of the codeword[5], which prevents the protocol from achieving full-rate. Fortunately, this problem can be avoided using a modified version of the DDF protocol. The reason is that, in the DDF protocol, a node will start helping only after it has successfully decoded the message. Therefore, cooperation does not come at the price of reduced rate. The modified DDF protocol, which will be called the CB-DDF protocol, is very similar to the DDF protocol for the multiple-relay scenario. The only modification needed, is that now every destination can act as a relay for the other destinations, based on its instantaneous channel gain. Specifically, the source transmits during the whole codeword while each destination listens until

[5]This follows from the half-duplex constraint.

the mutual information between its received signal and the signals transmitted by the source and other destinations exceeds lR. Once a destination decodes the message, it uses an independent code-book to re-encode the message, which it then transmits for the rest of the codeword. Similar to the relay channel, it is assumed that every destination knows the code-books used by the source and other destinations. Also, the protocol must include a mechanism that keeps every destination informed of the re-transmission starting times of all the other destinations. Again, in deriving the following result, we ignore the associated cost of this mechanism, relying on the asymptotic assumptions.

Theorem 9 *The DMT achieved by the CB-DDF protocol with N destinations is given by*

$$d(r_c) = \begin{cases} N(1-r_c), & \frac{1}{N} \geq r_c \geq 0, \\ 1 + \frac{(N-1)(1-2r_c)}{1-r_c}, & \frac{1}{2} \geq r_c \geq \frac{1}{N}, \\ \frac{1-r_c}{r_c}, & 1 \geq r_c \geq \frac{1}{2}. \end{cases} \tag{4.1}$$

Proof: Please refer to Appendix A.8.

It is interesting to note that this is exactly the same tradeoff obtained in the relay channel (compare to (3.8)). This implies that requiring all nodes to decode the message does not entail a price in terms of the achievable tradeoff.

4.2 Multiple-Access Relay Channel

In the (two-user) MAR channel, a relay is assigned to help the two users transmit their messages to a common destination. In this scenario, the users are not allowed to cooperate with one another (e.g, due to practical limitations). In our DDF protocol for this channel, the two users transmit their individual messages during every symbol interval in the codeword, while the relay listens to the users until it collects

sufficient energy to decode *both* of them error-free. After decoding, the relay uses an *independent* code book to encode the two messages *jointly*. The encoded symbols are then transmitted for the rest of the codeword.

We characterize the DMT achieved by the DDF protocol, along with an upper bound on the achievable DMT in the MAR channel, in Theorem 10. However, before proceeding further, we need to define the multiplexing gain r and diversity gain d. Toward this end, we notice that using (2.6), the pair (r_j, d_j) can be defined for each of the two users. However, since we consider the *symmetric* MAR scenario, the two multiplexing gains r_1 and r_2 are equal, i.e. $r_1 = r_2$. We define the overall multiplexing r as $r \triangleq r_1 + r_2$. That is, the multiplexing gain at which the destination receives information. We further notice that as a result of symmetry, $d_1 = d_2$. Therefore, we define the overall diversity gain d, as $d \triangleq d_1 = d_2$. We are now ready to state our main result in this section

Theorem 10 *The optimal diversity gain for the symmetric two-user MAR channel is upper bounded by*

$$d_{MAR}(r) \leq \begin{cases} 2-r & \text{if } \frac{1}{2} \geq r \geq 0 \\ 3(1-r) & \text{if } 1 \geq r \geq \frac{1}{2} \end{cases} . \tag{4.2}$$

Furthermore, the DMT achieved by the DDF protocol is lower bounded by

$$d_{DDF\text{-}MAR}(r) \geq \begin{cases} 2-r & \text{if } \frac{1}{2} \geq r \geq 0 \\ 3(1-r) & \text{if } \frac{2}{3} \geq r \geq \frac{1}{2} \\ 2\frac{1-r}{r} & \text{if } 1 \geq r \geq \frac{2}{3} \end{cases} . \tag{4.3}$$

Proof: Please refer to Appendix A.9.

Fig. 4.1 compares the upper and lower bounds in Theorem 10 where the optimality of the DDF protocol for $2/3 \geq r \geq 0$ is evident. As a final remark, We note that the results and ideas in Theorem 10 extend to the N-user MAR channel. However,

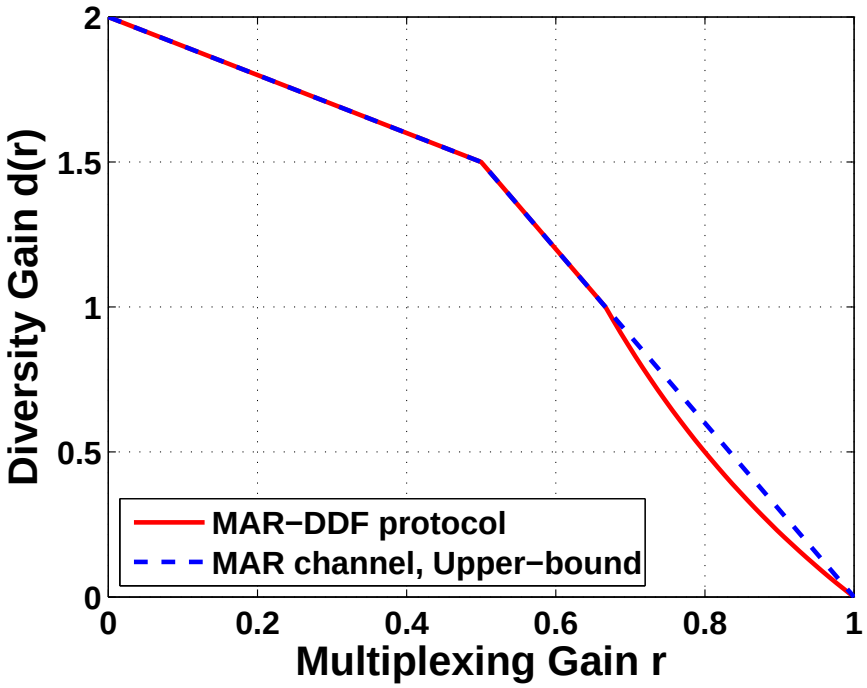

Figure 4.1: The DMT achieved by the DDF protocol in the MAR channel, along with an upper-bound on the achievable DMT .

the mathematics becomes tedious and does not provide further insights. Overall, the main conclusion in this section is that a *single* relay can be efficiently shared by *several* multiple-access users such that it enhances the diversity gain achieved by *all* of them.

4.3 Cooperative Multiple-Access Channel

In this section, we consider the CMA scenario, where N sources transmit their independent messages to a common destination. This scenario is a generalization of the multiple-access channel where the sources are allowed to cooperate. As was the case in the MAR scenario, we assume symmetry so that all sources transmit information at the same rate. The basic idea of the proposed protocol, which we refer to as the CMA-NAF protocol, is to create an artificial ISI channel. Towards this end, each of the N sources transmits once per cooperation frame, where a cooperation frame is defined as N consecutive symbol-intervals (refer to part (a) of Fig. 4.2). Each source is assigned unique *transmission* and *reception* symbol-intervals within the cooperation frame. During its transmission symbol-interval, a source transmits a linear combination of its own symbol and the signal it observed during its most recent reception symbol-interval. In other words, every source, in addition to sending its own symbol, *helps* another source by repeating the (noisy) signal it last received from it. Without loss of generality, we set the j^{th} source transmission symbol-interval equal to j.

We now provide an illustrative example for the $N = 3$ case. Here we assume that sources 1, 2, and 3 help sources 3, 1, and 2, respectively. For the j^{th} source and the k^{th} cooperation frame, $t_{j,k}$ denotes the transmission, $r_{j,k}$ the (assigned) reception, and $x_{j,k}$

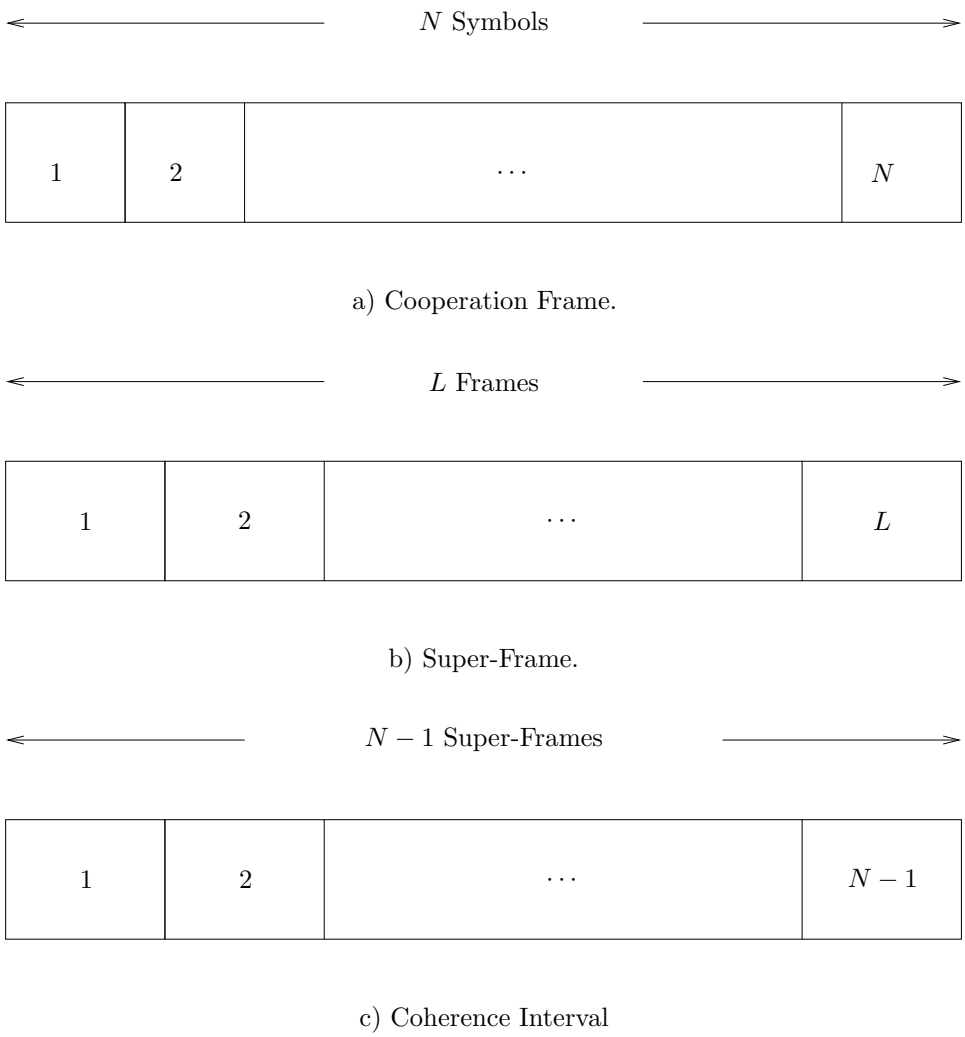

Figure 4.2: The cooperation frame, super-frame and coherence-interval in the CMA-NAF protocol with N sources.

the originating symbol. Using a_j and b_j to denote the broadcast and repetition gains of the j^{th} source, respectively, the signals transmitted during the first two cooperation frames would be (in chronological order)

$$t_{1,1} = a_1 x_{1,1}$$

$$t_{2,1} = a_2 x_{2,1} + b_2 r_{2,1}$$

$$t_{3,1} = a_3 x_{3,1} + b_3 r_{3,1}$$

$$t_{1,2} = a_1 x_{1,2} + b_1 r_{1,1}$$

$$t_{2,2} = a_2 x_{2,2} + b_2 r_{2,2}$$

$$t_{3,2} = a_3 x_{3,2} + b_3 r_{3,2}.$$

Using h_{ji} to denote the i^{th}-source-to-j^{th}-source channel gain, and $w_{j,k}$ to denote the noise observed by the j^{th} source during its k^{th}-frame reception symbol-interval, the assigned receptions become

$$r_{2,1} = h_{21} t_{1,1} + w_{2,1}$$

$$r_{3,1} = h_{32} t_{2,1} + w_{3,1}$$

$$r_{1,1} = h_{13} t_{3,1} + w_{1,1}$$

$$r_{2,2} = h_{21} t_{1,2} + w_{2,2}$$

$$r_{3,2} = h_{32} t_{2,2} + w_{3,2}.$$

Using g_j to denote the j^{th}-source-to-destination channel gain, and $v_{j,k}$ to denote the noise observed by the destination during the j^{th} symbol-interval of the k^{th} frame, the signals observed at the destination would be

$$y_{j,k} = g_j t_{j,k} + v_{j,k}.$$

The source-observed noises $\{w_{j,k}\}$ have variance σ_w^2 for all j, k, and the destination-observed noises $\{v_{j,k}\}$ have variance σ_v^2 for all j, k. Note that, as mandated by our half-duplex constraint, no source transmits and receives simultaneously. The broadcast and repetition gains $\{a_j, b_j\}$ should be chosen to satisfy the average power constraint

$$E\{|t_{j,k}|^2\} \leq E. \tag{4.4}$$

Let us now define L consecutive cooperation frames as a super-frame (refer to part b of Fig. 4.2). We will assume that helper assignments are fixed within a super-frame but are scheduled to change across super-frames. We impose the following requirements on helper scheduling.

1. In each super-frame, every source is helped by a different source.

2. Across super-frames, every source is helped equally by every other source.

Among the many scheduling rules that satisfy these requirements, we choose the following circular rule. In super-frame i, sources with indices $(1, \ldots, N)$ are assigned helpers with indices given by the j^{th} right circular shift of $(1, \ldots, N)$, where $j = \langle i-1 \rangle_{N-1} + 1$. For example, when $N = 4$, the helper configurations are given by the following table. Since this scheduling algorithm generates $N-1$ distinct helper configurations, the length of the super-frames, L, is chosen such that a coherence-interval consists of $N-1$ consecutive super-frames (refer to part c of Fig. 4.2). To achieve maximal diversity for a given multiplexing gain, it is required that all codewords span the entire coherence-interval. For this reason, we choose codes of length l given by

$$l = (N-1)L. \tag{4.5}$$

Super-frame index	Helper assigned to 1	2	3	4
1	4	1	2	3
2	3	4	1	2
3	2	3	4	1
4	4	1	2	3

Table 4.1: An exemplar scheduling rule for CMA-NAF protocol.

Similar to the broadcast channel, defining the multiplexing gain r and diversity gain d for the cooperative multiple-access channel requires some care. Note that, using (2.6), the pair (r_j, d_j) can be defined for communication between the j^{th} source and the destination. However, since we assumed a symmetric CMA setup, all multiplexing gains are equal, i.e., $r = r_j$ for all j. Furthermore, since CMA-NAF mandates that only one source transmits in any symbol-interval, the destination's multiplexing gain is also equal to r. That is, the destination receives information at rate R given by

$$R = r \log(\rho). \tag{4.6}$$

We define the overall diversity gain d based on the worst case probability of error for the N information streams, i.e.,

$$d = \min_{1 \leq j \leq N} \{d_j\}.$$

With these definitions, Theorem 11 establishes the optimality of the CMA-NAF in the symmetric scenario with N sources.

Theorem 11 *The CMA-NAF protocol achieves the optimal (genie-aided) DMT for the symmetric scenario with N sources, given by*

$$d^*(r) = N(1 - r). \tag{4.7}$$

Proof: Please refer to Appendix A.10.

Theorem 11 not only establishes the optimality of the CMA-NAF protocol, but also shows that the half-duplex constraint does not entail any cost (in terms of DMT) in the symmetric CMA channel. One can now attribute the sub-optimality of the CMA schemes reported in [3, 4] to the use of orthogonal subspaces. It is also interesting to observe that one can achieve the optimal tradeoff in the symmetric CMA channel with a simple AF strategy. In fact, by comparing Theorems 2 and Theorem 11, one can see the fundamental difference between the half duplex CMA and relay channels.

Next we report numerical results (in terms of outage probability[6]) that quantify the performance gains offered by our CMA-NAF protocol. In particular, Fig. 4.3 shows that the optimality of the CMA-NAF protocol, as compared to NAF-AF, translates into significant SNR gains. As was the case with the DDF protocol, this gain increases with the data rate, which is again a direct consequence of the higher multiplexing gains achievable through the CMA-NAF protocol. Note also that our comparison with the LTW-AF protocol here is fair, since in the CMA-NAF protocol, the constant average energy per symbol-interval requirement is automatically implied. In fact, to obtain a lower bound on the gain offered by the CMA-NAF protocol, we assumed a *noiseless* source-relay channel for the LTW-AF protocol, while for the CMA-NAF protocol, we assumed the SNR of the inter-user link to be only 3 dB better than those of the user-destination channels. We optimized the broadcast and repetition gains for the CMA-NAF protocol experimentally. As was the case

[6]Please refer to [31] for an exemplar implementation of the CMA-NAF protocol and the associated bit and frame error rates.

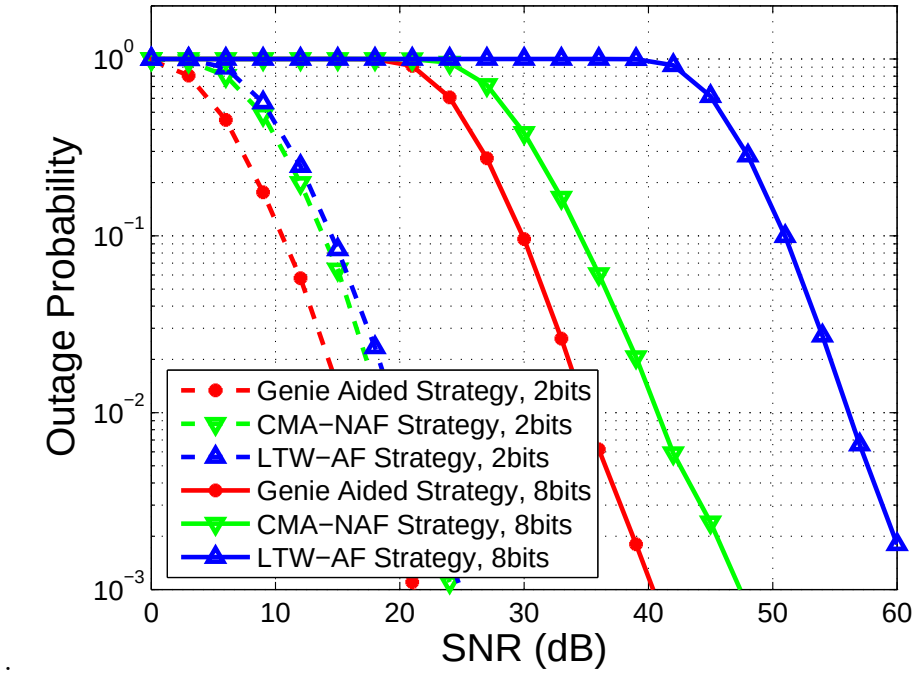

Figure 4.3: Comparison of the outage probability for the CMA-NAF, LTW-AF and genie-aided 2×1 protocols ($N = 2$).

with relay protocols, the outage probabilities were computed through Monte-Carlo simulations.

Also depicted in Fig. 4.3 are the outage probability curves for the genie-aided strategy (i.e., 2×1 MIMO). It is interesting to notice that the gap between the outage curves of the CMA-NAF and genie-aided strategies is less than 3 dB when the date rate is equal to 2 BPCU.

CHAPTER 5

ARQ COOPERATIVE CHANNELS

In this chapter we consider ARQ cooperative channels where users, after each transmission round, are provided with ACK/NACK signals indicating success or failure of destination in decoding their messages. As we will show, utilization of ARQ techniques not only improves the DMT achieved by some of the proposed protocols (e.g., DDF relay and MAR), but also provides opportunities for cooperation that are otherwise unavailable (e.g., DDF CVMA protocol). We start, though, with a few definitions and assumptions that are used throughout the chapter.

- We denote the maximum allowable number of ARQ rounds by L, where $L = 1$ corresponds to the non-ARQ scenario. A transmission round consists of l consecutive symbol-intervals. In order to highlight the benefits of ARQ, as opposed to temporal interleaving, we adopt the long-term static channel model of [29] where all the ARQ rounds corresponding to a certain message take place over the same coherence-interval.

- We denote the first-round rate of transmission (at destination) by R_1 and the average throughput by η. These two quantities are related through [29]

$$\eta = \frac{R_1}{1 + \sum_{\ell=1}^{L-1} p(\ell)}, \tag{5.1}$$

where $p(\ell)$ is the probability that the destination requests for the $(\ell+1)$th round of transmission. We define the first-round multiplexing gain, r_1, and effective multiplexing gain, r_e, as

$$r_1 \triangleq \lim_{\rho\to\infty} \frac{R_1}{\log \rho}, \qquad\qquad r_e \triangleq \lim_{\rho\to\infty} \frac{\eta}{\log \rho}. \tag{5.2}$$

From (5.1), we note that if $\{p(\ell)\}_{\ell=1}^{L-1}$ decay polynomially with ρ, then $\eta \dot{=} R_1$ and thus $r_e = r_1$.

- As was the case in non-ARQ scenarios, we adopt the coherent transmission paradigm where only the receiving node of any link knows the channel gain; Except for the ACK/NACK feedback bits, no other CSI is available to the transmitting nodes.

- In characterizing DMT for ARQ cooperative protocols, we replace r with the *effective* multiplexing gain r_e to capture the variable-rate nature of ARQ schemes [29].

5.1 ARQ Multiple-Access Relay Channel

In this section, we adopt the DDF protocol to the ARQ MAR channel and prove its optimality. The proposed ARQ-DDF protocol is identical to the one devised for the MAR channel (refer to Section.4.2), except that users only move on to sending their next message if either all of them have received ACK signals, or that the maximum allowable number of ARQ rounds, L, has already been requested. To prove the optimality of this protocol, we first prove the optimality of the ARQ-DDF protocol in the relay channel (i.e., an ARQ MAR channel with a single user). The proof for this result provides the machinery necessary to handle the ARQ mechanism. Then,

we combine the ideas in Lemma 12 and Theorem 10 to prove the optimality of the ARQ-DDF protocol in the (two-user) ARQ MAR channel.

Lemma 12 *The optimal DMT for the relay channel with $L \geq 2$ ARQ rounds is given by*

$$d_R(r_e, L) = 2(1 - \frac{r_e}{L}) \quad for \quad 1 > r_e \geq 0. \tag{5.3}$$

Furthermore, this optimal tradeoff is achieved by the proposed ARQ-DDF protocol.

Proof: Please refer to Appendix A.11.

Theorem 13 *The optimal DMT for the symmetric two-user MAR channel with $L \geq 2$ ARQ rounds is given by*

$$d_{MAR}(r_e, L) = 2 - \frac{r_e}{L} \quad for \quad 1 > r_e \geq 0. \tag{5.4}$$

Furthermore, this optimal tradeoff is achieved by the proposed ARQ-DDF protocol.

Proof: Please refer to Appendix A.12.

We note that the optimality in Theorem 13 extends to the N-user MAR channel. As was the case with the (non-ARQ) MAR scenario, the main conclusion here is that a *single* relay can be efficiently shared by *several* multiple-access users such that it enhances the diversity gain achieved by *all* of them.

Before concluding this section we notice that the DMTs achieved by the ARQ-DDF protocol for relay and MAR channels are stretched versions (with a factor of L) of those for (non-ARQ) DDF relay and MAR protocols (compare (5.3) to (3.7) and (5.4) to(4.3)). This is because of the following two characteristics of ARQ protocols. 1) The average throughput, as a function of SNR, converges asymptotically to the

throughput corresponding to one round of transmission, and 2) the error event is dominated by those for which L rounds of transmission has been requested. As a result, the ARQ protocols enjoy the diversity gain that the non-ARQ versions achieve at a multiplexing gain L times smaller.

5.2 ARQ Cooperative Vector Multiple-Access Channel

In the cooperative vector multiple access (CVMA) channel, the two single-antenna users are allowed to assist each other, as long as they do not violate the half duplex constraint. The challenge in this scenario stems from the availability of two receiving antennas at the destination, which increases the channel's degrees of freedom to two. Loosely speaking, in order to exploit these two degrees of freedom, the two users need to transmit *new* independent symbols continuously, which prevents them from cooperation under the half duplex constraint (non-ARQ case). More rigorously, it is straightforward to see that with $L = 1$, any half-duplex cooperation protocol that achieves full diversity (i.e., $d(0) = 3$), falls short of achieving full rate (i.e., $d(r) > 0$, for all $r < 2$). To get around this problem, in the case of $L \geq 2$, we purposefully instruct the users *not* to cooperate in the first round of transmission. In fact, a user continues transmitting its message while it receives NACK signals. Only when a user receives an ACK signal, it starts listening to the other user (assuming the other user has not been successfully decoded yet). Once the cooperating user decodes the message of its partner, it starts helping (i.e., the typical DDF methodology). The following result establishes lower and upper bounds on the DMT achieved by this protocol.

Theorem 14 *The optimal diversity gain for the symmetric two-user CVMA channel with L ARQ rounds is upper bounded by*

$$d_{CVMA}(r_e, L) \leq \min\left\{3(1-\frac{r_e}{2L}), 4-\frac{3r_e}{L}\right\}, \quad \textit{for} \;\; 2 > r_e \geq 0. \tag{5.5}$$

For $L = 2$, the diversity gain achieved by the proposed ARQ-DDF protocol satisfies

$$d_{DDF-CVMA}(r_e, 2) \geq \begin{cases} 3-r_e, & 1 > r_e \geq 0 \\ 4-2r_e, & \frac{4}{3} > r_e \geq 1 \\ 2-\frac{r_e}{2}, & 2 > r_e \geq \frac{4}{3} \end{cases}. \tag{5.6}$$

Furthermore, as L increases, the ARQ-DDF diversity gain converges to the optimal value, i.e.,

$$\lim_{L\to\infty} d_{DDF-CVMA}(r_e, L) = 3, \quad \textit{for} \;\; 2 > r_e \geq 0. \tag{5.7}$$

Please refer to Appendix A.13.

Fig. 5.1 compares the upper and lower bounds in (5.5) and (5.6) for $L = 2$. Clearly, this figure shows the full diversity and full rate properties of the proposed ARQ-DDF protocol. Finally, we observe that the analysis of the ARQ-DDF protocol in Theorem 14 can be repeated for $L > 2$. We have, however, chosen to omit this analysis since it is rather tedious and not necessarily inspiring.

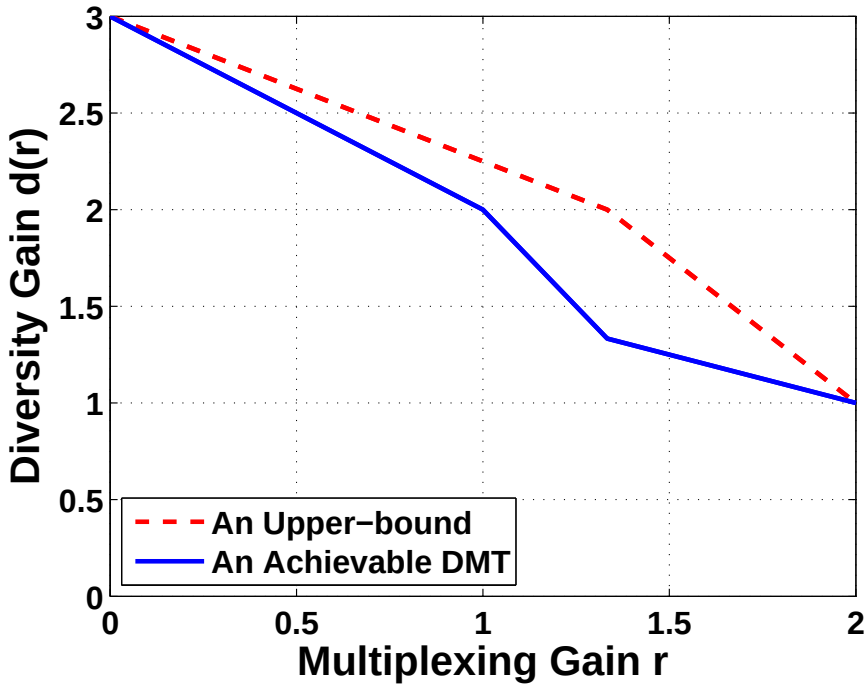

Figure 5.1: The DMTachieved by the ARQ-DDF protocol in the CVMA channel, along with an upper-bound on the achievable DMT ($L = 2$).

CHAPTER 6

THE THROUGHPUT-RELIABILITY TRADEOFF

6.1 Problem Formulation

This chapter revolves around the following question: What does a 3 dB increase in SNR buy in an outage limited MIMO channel? In an AWGN setting, it is well known that a 3 dB increase in SNR translates into an extra bit in channel capacity in the high SNR regime. The scenario considered in this chapter, however, is more involved. We study an outage limited channel where the randomness of the instantaneous mutual information results in a non-zero lower bound on the probability of error, for non-zero constant transmission rates. Hence, a fundamental tradeoff between the throughput, as quantified by the transmission rate, and reliability, as quantified by the so-called outage probability, arises. Our work explores this tradeoff in the high SNR regime.

To be more specific, we consider a MIMO wireless communication system with m transmit and n receive antennas. We adopt a quasi-static flat-fading setup where the path gains remain constant over l consecutive symbol-intervals (*i.e.,* a block), but change independently from one block to another. We further assume a coherent communication model implying the availability of channel state information (CSI) only at the receiver. Under these assumptions, the channel input-output relation is

given by:

$$\mathbf{y} = \sqrt{\frac{\rho}{m}}\mathbf{H}\mathbf{x} + \mathbf{w}. \tag{6.1}$$

In (6.1), $\mathbf{y} \in \mathbb{C}^n$ has entries y_i representing the signal received at antenna $i \in \{1, \cdots, n\}$, $\mathbf{x} \in \mathbb{C}^m$ has entries x_j denoting the signal transmitted by antenna $j \in \{1, \cdots, m\}$, $\mathbf{H} \in \mathbb{C}^{n \times m}$ has entries h_{ij} which represents the path gain between receive antenna $i \in \{1, \cdots, n\}$ and transmit antenna $j \in \{1, \cdots, m\}$, and $\mathbf{w} \in \mathbb{C}^n$ represents the unit-variance additive white Gaussian noise. We model $\{h_{ij}\}$ as i.i.d zero-mean and unit-variance complex Gaussian random variables. Finally, ρ corresponds to the SNR at each receive antenna.

Our work builds on Zheng and Tse's formulation of the DMT [17]. This formulation assumes a family of space-time codes $\{\mathcal{C}(\rho)\}$ indexed by their operating SNR ρ, such that the code $\mathcal{C}(\rho)$ has rate $R(\rho)$ BPCU, and error probability $P_e(\rho)$. For this family, the multiplexing gain r and the diversity gain d are defined by

$$r \triangleq \lim_{\rho \to \infty} \frac{R(\rho)}{\log \rho} \quad \text{and} \quad d \triangleq -\lim_{\rho \to \infty} \frac{\log P_e(\rho)}{\log \rho}. \tag{6.2}$$

The optimal DMT yields the maximum possible diversity gain for every value of r. The main result of [17] is summarized in the following theorem:

Theorem 15 *The optimal diversity gain for the coherent quasi-static MIMO channel with m transmit and n receive antennas, at multiplexing gain r, is given by $d(r) = f(r)$, where $f(\cdot)$ is the piecewise linear function joining the points $(k, (m-k)(n-k))$ for $k = 0, \ldots, \min\{m, n\}$. Moreover, there exists a code that achieves $d(r)$ for all block lengths $l \geq m + n - 1$.*

In the sequel, we will use the notation $d_{max} = mn$ and $r_{max} = \min\{m, n\}$. To motivate our work, we use the DMT to make a first attempt towards answering our central

question on the utility of a 3 dB SNR gain in quasi-static MIMO channels. Using the extreme points of the tradeoff curve, i.e., $(0, d_{max})$ and $(r_{max}, 0)$, a *reasonable* conjecture is

1. At high enough SNRs, one can fix the transmission rate and obtain a decay, by a factor of $10^{-d_{max}}$, in the outage probability, for every 10 dB gain in SNR.

2. At high enough SNRs, one can fix the outage probability and obtain a rate increase of r_{max} BPCU for every 3 dB gain in SNR.

Fig. 6.1 and Fig. 6.2 examine the validity of this conjecture in a 2×2 MIMO channel. In these figures, the transmission rates and SNR ranges are carefully chosen to illustrate the following points.

1. The slope of the outage probability curves in Fig. 6.1 is shown to approach the asymptotic value of $d_{max} = 4$, on the log-log scale, as predicted by the first part of our conjecture. The *surprising* observation, however, is that for a constant outage probability a 4.5 dB gain in SNR is needed to obtain an $r_{max} = 2$ BPCU increase in the throughput (To avoid fractions of a dB, the figure shows a 9 dB spacing for a 4 BPCU throughput increase). This contrasts the second part of our conjecture which predicts the need for only 3 dB for every 2 BPCU. More interestingly, this 4.5 dB *horizontal spacing* seems to persist as the SNR increases.

2. Fig. 6.2, on the other hand, comes in close agreement with the second part of our conjecture. Here, the horizontal spacing, for a 2 BPCU increase in throughput, is seen to be 3 dB. The disagreement in this case, however, is exhibited in the

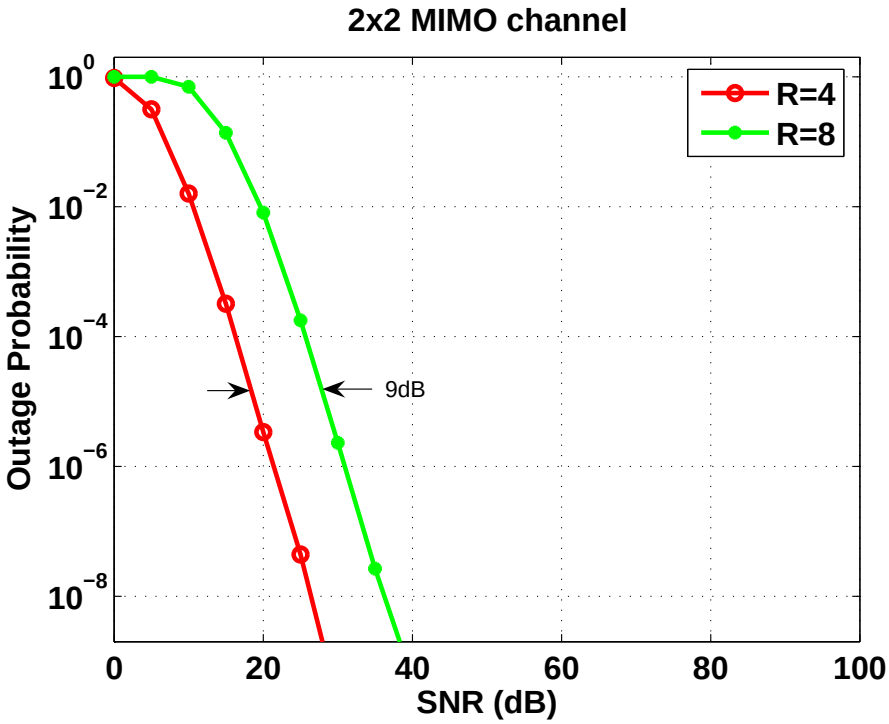

Figure 6.1: Outage curves corresponding to $R = 4, 8$ BPCU, for a 2×2 MIMO channel.

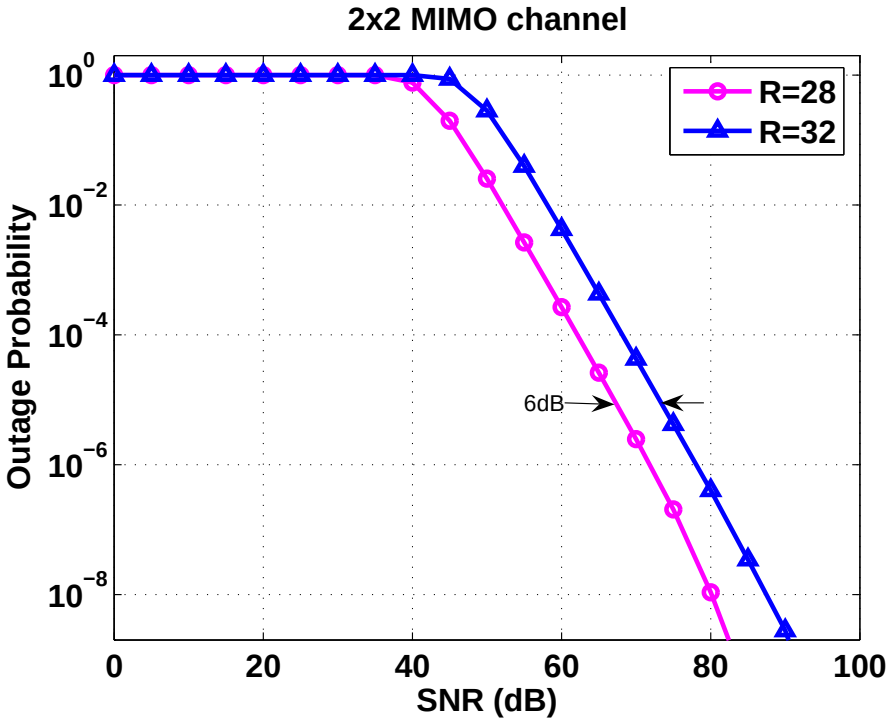

Figure 6.2: Outage curves corresponding to $R = 28, 32$ BPCU, for a 2×2 MIMO channel.

fact that the slope of the outage probability curves, corresponding to fixed rates, seems to *stabilize* for a wide range of SNRs at a value of 2 (instead of 4).

3. Repeating the experiment for different values of m and n reveals the same trends, i.e., 1) Our conjecture seems to offer *partially* accurate predictions in certain *operating regions*[7] and 2) Except for the 1×1 channel, the predictions for the outage probability rate of decay and horizontal spacing are *never simultaneously* accurate.

4. Overall, these disagreements clearly disprove our *naive* conjecture. However, it seems that the conjecture is *not completely* false as it offers some accurate predictions, at least in certain operating regions.

Inspired by these observations, this chapter aims at developing a better understanding of the fundamental throughput-reliability tradeoff in outage limited MIMO channels. It turns out that such an understanding requires a more general formulation which is not limited by the multiplexing gain notion as defined in (6.2). In particular, the multiplexing gain notion limits the scenarios of interest to *asymptotic lines* on the $R - \log \rho$ plane as shown in Fig. 6.3. Our formulation, on the other hand, allows for investigating more general scenarios by relaxing this constraint. Specifically, we shed more light on the relationship between the three quantities $(R, \log \rho, P_e(R, \rho))$, in the asymptotic limit of large ρ, when

$$\limsup_{\rho \to \infty} \frac{R}{\log \rho} \neq \liminf_{\rho \to \infty} \frac{R}{\log \rho}. \tag{6.3}$$

It is clear that (6.3) allows for investigating scenarios defined by arbitrary asymptotic trajectories in the $R - \log \rho$ plane where the multiplexing gain is not defined (Fig. 6.4

[7] A more formal definition of an operating region is presented in the sequel.

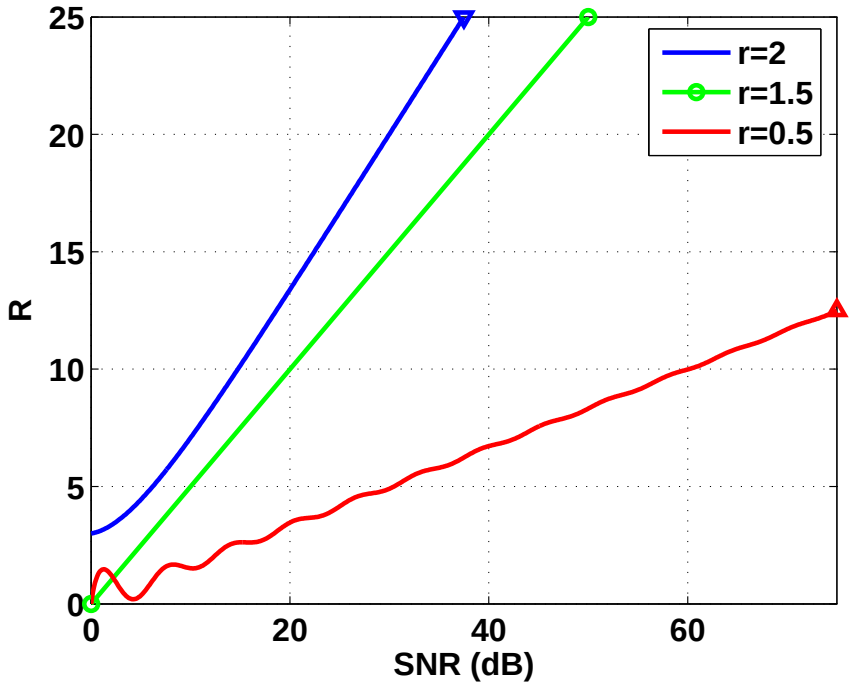

Figure 6.3: The notion of multiplexing gain restricts the scenarios of interest to those in which R *asymptotically* scales linearly with $\log \rho$, i.e. $R \sim r \log \rho$.

depicts such a trajectory). As argued in the sequel, this *freedom* of walking along arbitrary trajectories, on the $R - \log \rho$ plane, is the **key** to obtaining accurate predictions for the outage probability slopes and horizontal spacings in different operating regions. While our characterization is *rigorous* only in the asymptotic scenario where SNR grows to infinity (i.e., $\rho \to \infty$), we will demonstrate, via numerical results, that it yields very accurate predictions for practically relevant values of SNR.

The rest of the chapter is organized as follows. In Section 6.2, we state our main result formulating the throughput-reliability tradeoff (TRT) for the point-to-point coherent MIMO channel, along with a sketch of the main ideas in the proof. In this

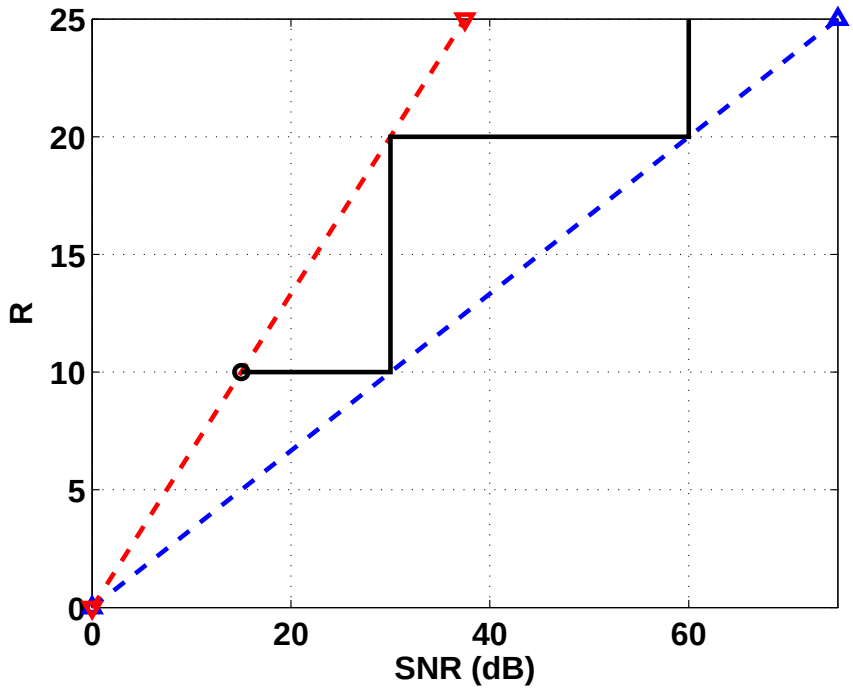

Figure 6.4: Relaxing the constraint imposed by the multiplexing gain notion; A multiplexing gain cannot be defined for the depicted trajectory, however, since it remains well within an operating region (i.e. $\mathcal{R}(1)$) TRT analysis can be applied.

section, we also present numerical results and intuitive arguments that demonstrate the utility of our results in predicting the *behavior* of outage probability curves in the high SNR regime. Section 6.3 utilizes the TRT to shed more light on the performance of various space-time architectures and further extends our investigation to Automatic Repeat reQuest (ARQ) channels. In order to enhance the flow of the chapter, the proofs are collected in Appendix A.

6.2 The Throughput-Reliability Tradeoff (TRT)

An outage is defined as the event that the instantaneous mutual information does not support the intended rate, i.e.,

$$O_{p(\mathbf{x})} \triangleq \{H \in \mathbb{C}^{n\times m} | I(\mathbf{x};\mathbf{y}|\mathbf{H}=H) < R\}.$$

Notice that the mutual information depends on both the channel realization H and the input distribution $p(\mathbf{x})$. The outage probability $P_o(R,\rho)$ is then defined as

$$P_o(R,\rho) = \inf_{p(\mathbf{x})} \Pr\{O_{p(\mathbf{x})}\}.$$

The following theorem characterizes the relationship between R, ρ, and $P_o(R,\rho)$.

Theorem 16 *For the $m \times n$ MIMO channel described by* (6.1)*,*

$$\lim_{\substack{\rho\to\infty \\ R\in\mathcal{R}(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log\rho} = -g(k), \tag{6.4}$$

where $P_o(R,\rho)$ denotes the outage probability at rate R and SNR ρ. $\mathcal{R}(k)$ is defined by

$$\mathcal{R}(k) \triangleq \{R | k+1 > \frac{R}{\log\rho} > k\} \quad for \quad k \in \mathbb{Z}, \min\{m,n\} > k \geq 0. \tag{6.5}$$

In (6.4)*,* $c(k)$ *and* $g(k)$ *are given by*

$$c(k) \triangleq m + n - (2k + 1), \tag{6.6}$$

and

$$g(k) \triangleq mn - k(k + 1). \tag{6.7}$$

Moreover, in the degenerate case $R > \min\{m, n\} \log \rho$*,* $\lim_{\rho \to \infty} \log P_o(R, \rho) / \log \rho = 0$*.*

We refer to $g(k)$ as the *reliability gain coefficient* and $t(k) \triangleq g(k)/c(k)$ as the *throughput gain coefficient*.

Proof: (Sketch) Our proof follows the same lines as the proof of Theorem 15 in [17] except for the fundamental challenge that the multiplexing gain is not defined here. To handle this challenge, we judicially choose the region $\mathcal{R}(k)$ and find a lower bound on

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_o(R, \rho) - c(k)R}{\log \rho},$$

and an upper bound on

$$\limsup_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_o(R, \rho) - c(k)R}{\log \rho}$$

and show that the two bounds coincide for the choice of $c(k)$ and $g(k)$ given by (6.6) and (6.7), respectively. The detailed proof is reported in the Appendix A.14.

It is immediate to check that if the multiplexing gain is well defined, i.e., if

$$r = \lim_{\rho \to \infty} \frac{R(\rho)}{\log \rho}$$

exists, then Theorem 16 reduces to Theorem 15. In the more general case, however, Theorem 16 replaces the restrictive multiplexing gain notion with the new concept of operating regions $\mathcal{R}(k)$. It is worth noting that every operating region corresponds to a line segment in the DMT. In fact, this correspondence inspires the following observation

$$g(k) = d(k) - kd'(k^+), \tag{6.8}$$

$$c(k) = -d'(k^+), \tag{6.9}$$

where $d'(k^+)$ is the slope of the line segment connecting $d(k)$ and $d(k+1)$.

We are now ready to investigate the asymptotic trends of the throughput-reliability tradeoff using Theorem 16. The following discussion hinges on the intuitive interpretation of equation (6.4).

$$\log P_o(R, \rho) \approx c(k)R - g(k)\log\rho \quad \text{for} \quad R \in \mathcal{R}(k), \tag{6.10}$$

where the approximation in (6.10) becomes progressively more accurate as ρ increases. Equation (6.10) implies that the slope of the outage curve, for a constant rate, is given by $g(k)$, while the horizontal spacing in dB between two outage curves with a ΔR rate difference is given by $3\Delta R/t(k)$. The key observation here is that in order to fix the transmission rate (or outage probability), while staying in the same operating region, one **must** deviate from the linear trajectory imposed by the multiplexing gain notion. The following *heuristic* derivation of (6.8) and (6.9) further illustrates this point. To derive (6.8), we start from two approximate relationships obtained from the DMT

$$\log P_o\left(R, 2^{\log\rho}\right) \approx -d\left(\frac{R}{\log\rho}\right)\log\rho, \tag{6.11}$$

$$\log P_o\left(R, 2^{\log\rho+\Delta\log\rho}\right) \approx -d\left(\frac{R}{\log\rho+\Delta\log\rho}\right)(\log\rho+\Delta\log\rho). \tag{6.12}$$

We further approximate $d\left(\frac{R}{\log\rho+\Delta\log\rho}\right)$ with the first two terms of its Taylor series expansion, i.e.,

$$d\left(\frac{R}{\log\rho+\Delta\log\rho}\right) \approx d\left(\frac{R}{\log\rho}\right) - \frac{R\times\Delta\log\rho}{\log\rho(\log\rho+\Delta\log\rho)}d'\left(\frac{R}{\log\rho}\right), \tag{6.13}$$

Now (6.13), together with (6.11) and (6.12), gives

$$\frac{\log P_o(R, 2^{\log\rho}) - \log P_o(R, 2^{\log\rho+\Delta\log\rho})}{\Delta\log\rho} \approx d\left(\frac{R}{\log\rho}\right) - \frac{R}{\log\rho}d'\left(\frac{R}{\log\rho}\right). \tag{6.14}$$

Realizing that the left-hand side of (6.14) gives the slope of $P_o(R,\rho)$ with respect to ρ, i.e. $g(k)$, and that $d(r) - rd'(r)$ remains constant over the line segments of $d(r)$, we get (6.8). Deriving (6.9) follows the same lines. In particular, we first compute the horizontal spacing, $\Delta\log\rho$, between the outage curves corresponding to rates R and $R+\Delta R$. For this purpose, we use (6.11) to write

$$\log P_o\left(R+\Delta R, 2^{\log\rho+\Delta\log\rho}\right) \approx -d\left(\frac{R+\Delta R}{\log\rho+\Delta\log\rho}\right)(\log\rho+\Delta\log\rho) \tag{6.15}$$

Then we expand $d(\frac{R+\Delta R}{\log\rho+\Delta\log\rho})$ in a way similar to (6.13) and equate (6.11) with (6.15) to get

$$\begin{aligned}\Delta\log\rho &\approx \frac{-d'(\frac{R}{\log\rho})}{d(\frac{R}{\log\rho}) - \frac{R}{\log\rho}d'(\frac{R}{\log\rho})}\Delta R,\\ \Delta\log\rho &\approx \frac{-d'(\frac{R}{\log\rho})}{g(k)}\Delta R.\end{aligned} \tag{6.16}$$

Realizing that $\Delta\log\rho = \frac{c(k)}{g(k)}\Delta R$, and that $d'(r)$ remains constant over the line segments of $d(r)$, we get (6.9).

Revisiting our naive conjecture, we can now see that $g(0) = mn = d_{max}$ which agrees with the first part, while $t(\min\{n,m\}-1) = \min\{m,n\} = r_{max}$ agrees with the

second part. This explains the *partial* correctness of the conjecture and the fact that, except for the 1×1 MIMO channel, the two parts are **never** simultaneously accurate, since they correspond to different operating regions. We also observe that both the reliability and throughput gain coefficients exhibit a staircase behavior. Moreover, it is easy to see that $g(k)$ is a decreasing function of k, while $t(k)$ is an increasing function of k. This implies that, at a fixed rate R and for sufficiently large SNRs, as $R/\log\rho$ increases (i.e., ρ decreases), the decay rate of $P_o(R,\rho)$ decreases and the horizontal spacing between the outage curves corresponding to a fixed rate difference shrinks. In the following, we present numerical results that validate this observation.

Before proceeding to the numerical results, we need the following rule of thumb for determining the operating regions for large but finite values of ρ and R such that the approximation in (6.10) is accurate. The operating point (R,ρ) is in operating region $\mathcal{R}(k)$ if and only if

$$\frac{\rho^k}{2^R} \leq \delta, \quad \text{and} \quad \frac{2^R}{\rho^{k+1}} \leq \delta,$$

where δ is a **small** value which determines the accuracy of the approximation. It is now straightforward to see that the high SNR segment of Fig. 6.1 falls in the region $\mathcal{R}(0)$ and, indeed, the 4 levels of diversity and 4.5 dB spacing (for every 2 BPCU throughput increase) in this figure correspond precisely to $g(0) = 4$ and $t(0) = 4/3$. Similarly, the high SNR segment of Fig. 6.2 falls in $\mathcal{R}(1)$ and, again, the 2 levels of diversity and 3 dB spacing, for every 2 BPCU throughput increase, agree with $g(1) = 2$ and $t(1) = 2$. Fig. 6.6 and Fig. 6.7 are different from the previous two figures in that the high SNR segments of the outage curves fall within **both** of the two regions. As a result, the slope of the curves and the spacing between them change as the operating point leaves one operating region and enters another. Again,

the values of the slopes, spacings, and operating points at which the change occurs (which can be read from Fig. 6.5) match nicely with the predictions of the TRT as formulated by Theorem 16. Fig. 6.8 through Fig. 6.11 correspond to a 3×3 MIMO system. As can be seen from Fig. 6.8, the solid segment of the curve corresponding to $R = 10$ BPCU, falls in $\mathcal{R}(1)$ and, as predicted, we observe a slope of $g(1) = 7$. It should be noted, however, that the tail of the curve corresponding to $R = 4$ BPCU is leaving $\mathcal{R}(1)$ and entering $\mathcal{R}(0)$ and thus the slope of this curve is larger than 7 (about 7.7). For the same reason, the horizontal spacing between the two curves (almost 10 dB) is larger than the value predicted for $k = 1$ (i.e., 7.7 dB). The solid segments of the outage curves corresponding to $R = 58$ and 64 BPCU in Fig. 6.9 fall in $\mathcal{R}(2)$, and therefore, we observe 3 levels of diversity and 3 dB of spacing, for every 3 BPCU throughput increase, which correspond precisely to $g(2) = 3$ and $t(2) = 3$. Fig. 6.10 and Fig. 6.11 depict the case where the high SNR segments fall within two different regions ($k = 2$ and $k = 1$). Again the slopes and spacings are in agreement with the predictions of the TRT.

6.3 Applications

The following result establishes the operational significance of Theorem 16 by showing that the optimal space-time code probability of error exhibits the same asymptotic behavior as the outage probability.

Theorem 17 *The probability of error for the* **optimal** *coding/decoding scheme used in conjunction with channel (6.1) satisfies*

$$\lim_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_e(R, \rho) - c(k)R}{\log \rho} = -g(k), \tag{6.17}$$

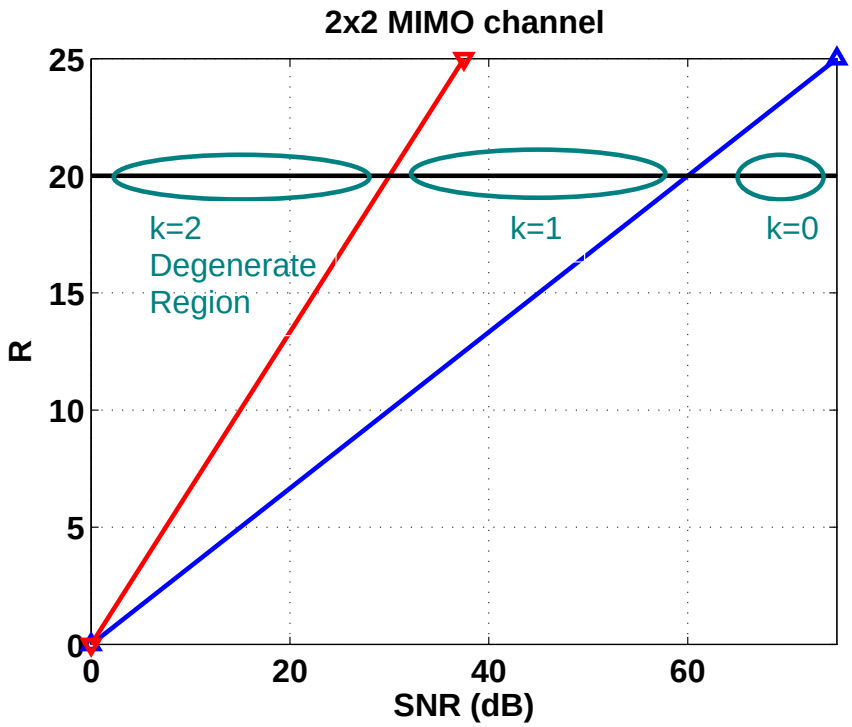

Figure 6.5: The constant rate trajectory with $R = 20$ BPCU passes through different operating regions in a 2×2 MIMO system

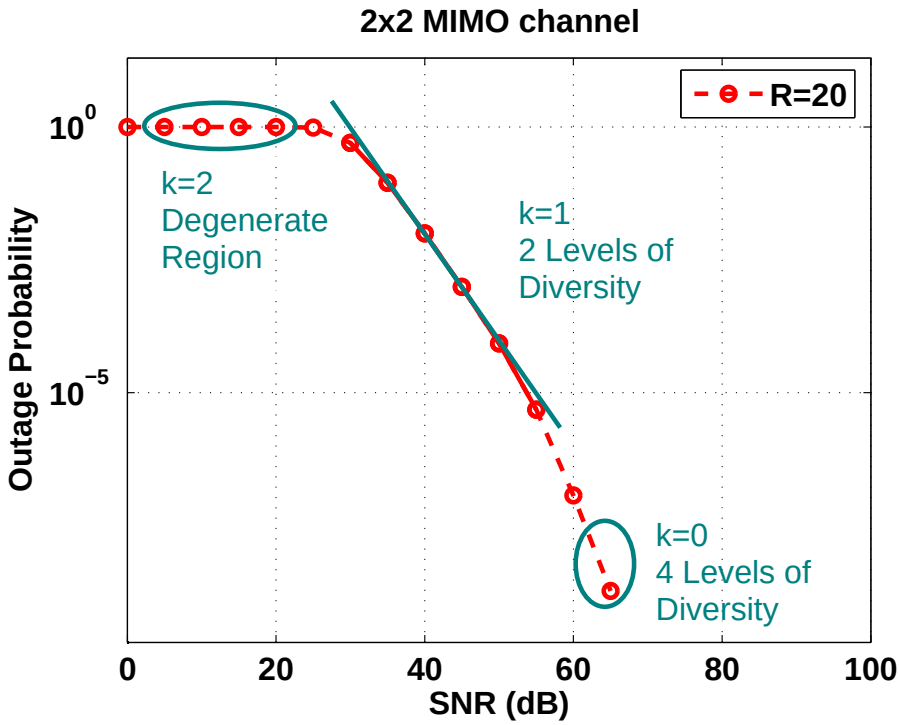

Figure 6.6: Outage curves corresponding to $R = 20$ BPCU for a 2×2 MIMO channel. The solid segment corresponds to the $\mathcal{R}(1)$ operating region.

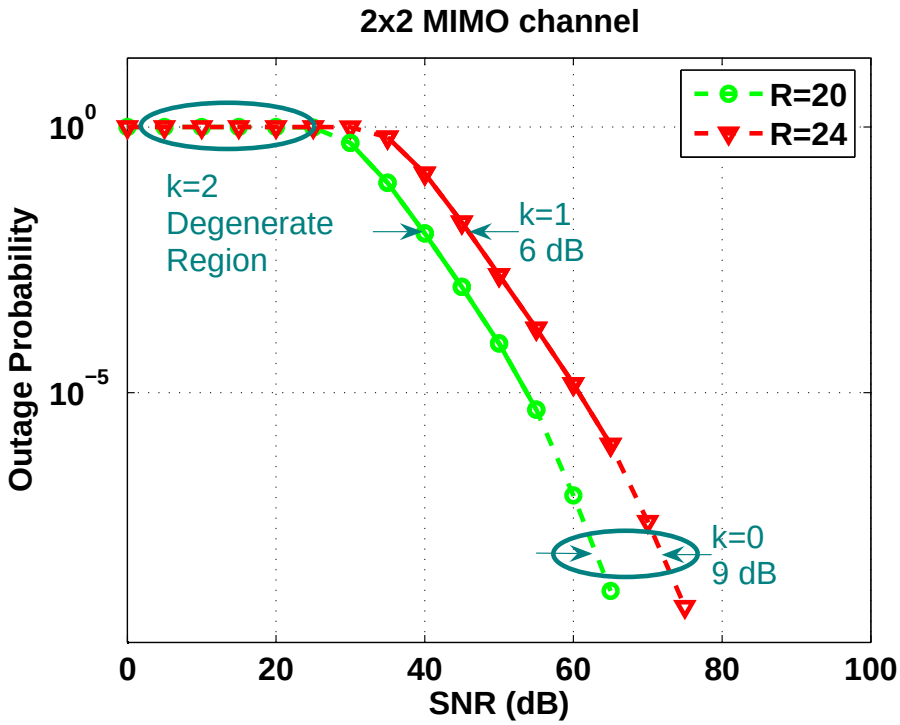

Figure 6.7: Outage curves corresponding to $R = 20, 24$ BPCU for a 2×2 MIMO channel. The solid segments correspond to the $\mathcal{R}(1)$ operating region.

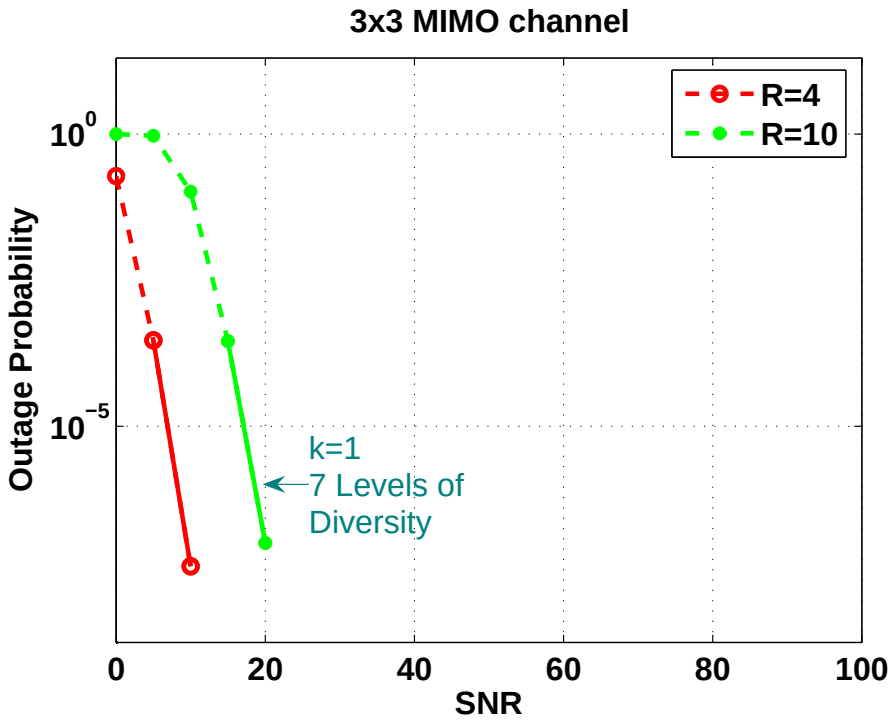

Figure 6.8: Outage curves corresponding to $R = 4, 10$ BPCU for a 3×3 MIMO channel. The solid segments correspond to the $\mathcal{R}(1)$ operating region.

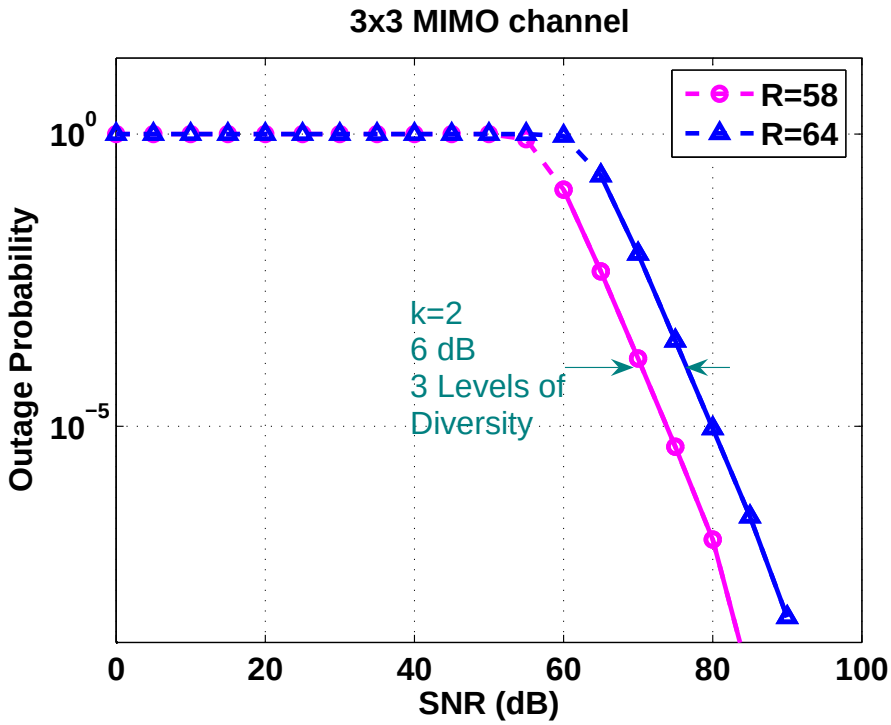

Figure 6.9: Outage curves corresponding to $R = 58, 64$ BPCU for a 3×3 MIMO channel. The solid segments correspond to the $\mathcal{R}(2)$ operating region.

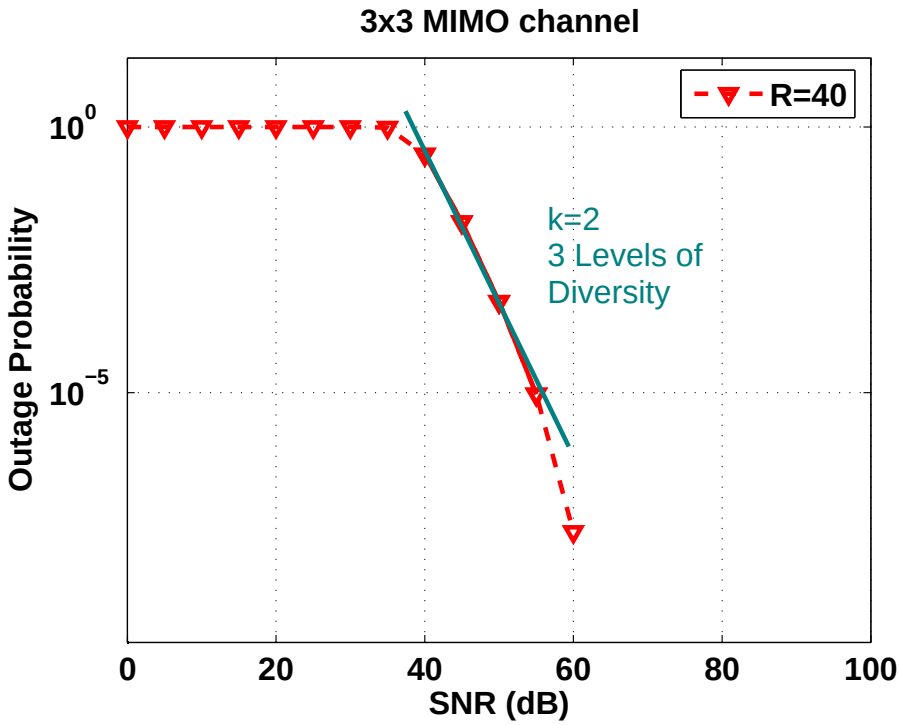

Figure 6.10: Outage curves corresponding to $R = 40$ BPCU for a 3×3 MIMO channel. The solid segment corresponds to the $\mathcal{R}(2)$ operating region.

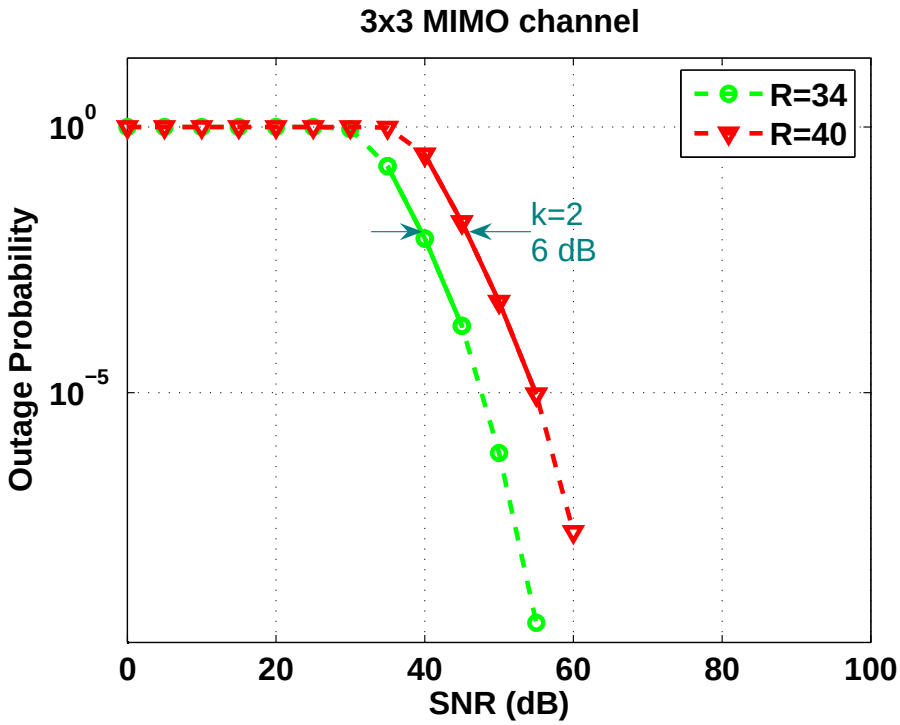

Figure 6.11: Outage curves corresponding to $R = 34, 40$ BPCU for a 3×3 MIMO channel. The solid segments correspond to the $\mathcal{R}(2)$ operating region.

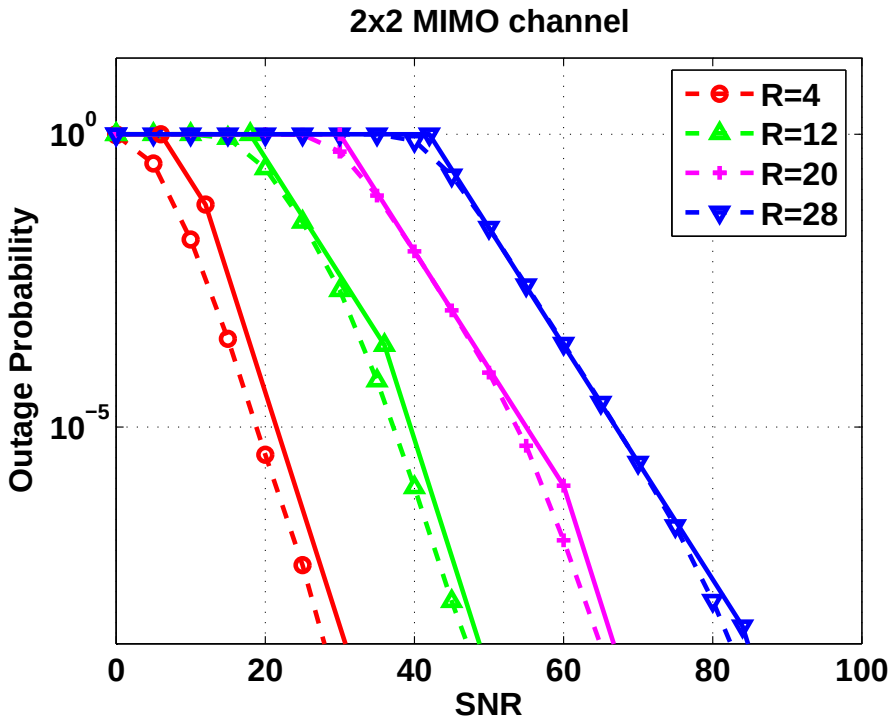

Figure 6.12: Outage probability curves for the 2×2 MIMO channel (dashed), along with the piecewise-linear approximation (solid) suggested by TRT.

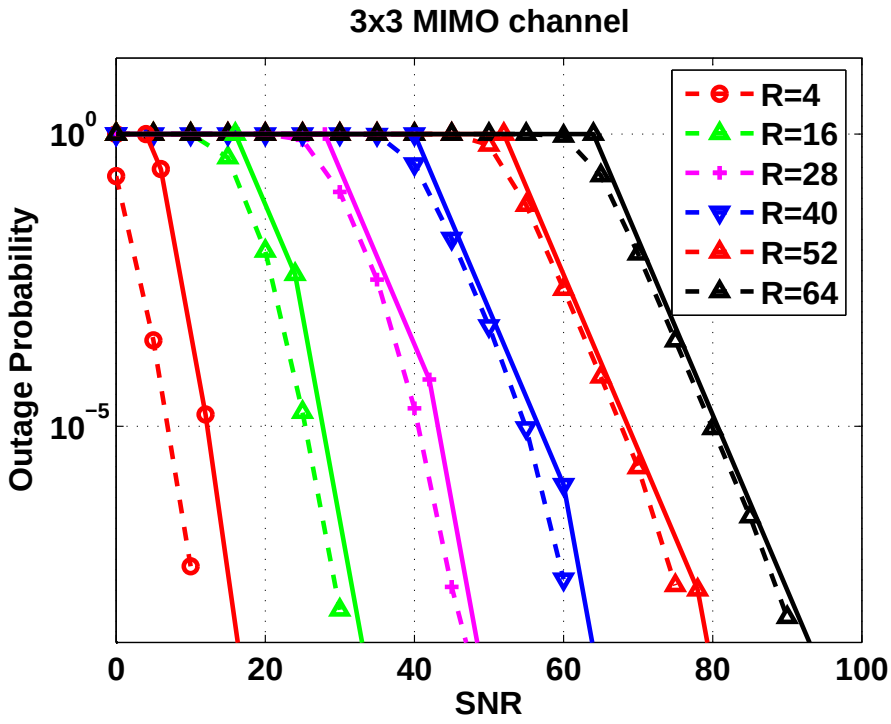

Figure 6.13: Outage probability curves for the 3×3 MIMO channel (dashed), along with the piecewise-linear approximation (solid) suggested by TRT.

where $\mathcal{R}(k)$, $c(k)$, and $g(k)$ are given by (6.5), (6.6), and (6.7), respectively. Moreover, there exists a coding scheme that achieves (6.17) for $l \geq m + n - 1$.

Proof: (Sketch) The proof follows the same lines as [17]. In particular, the converse is obtained via a careful use of Fano's inequality. The achievability is established using an ensemble of Gaussian codebooks along with the appropriate use of the union bound. Please refer to Appendix A.15 for a detailed proof.

One can also derive the TRT achievable by certain suboptimal space-time architectures. In this chapter, we restrict our study to square V-BLAST protocols and orthogonal space-time constellations. In the V-BLAST architecture, the input stream is split into m sub-streams. These sub-streams are then encoded independently and transmitted over the m transmit antennas [26]. The following theorem characterizes the throughput-reliability tradeoff for this protocol when a maximum likelihood decoder is employed.

Theorem 18 *The ML error probability for a V-BLAST communication system with m transmit and m receive antennas satisfies*[8]

$$\lim_{\substack{\rho \to \infty \\ R \in \mathcal{R}_{vb}}} \frac{\log P_e(R, \rho) - R}{\log \rho} = -m, \tag{6.18}$$

where $\mathcal{R}_{vb}$ is given by

$$\mathcal{R}_{vb} \triangleq \{R | m > \frac{R}{\log \rho} > 0\}.$$

Moreover, there exists a coding scheme that achieves (6.18) *for $l \geq 2m - 1$.*

Proof: Please refer to the Appendix A.16.

[8]The subscript "vb" stands for V-BLAST.

Fig. 6.14 depicts the outage curves corresponding to $R = 8$ and 12 BPCU for a 2×2 V-BLAST scheme with ML decoding. As can be seen from this figure, the high SNR segments of the outage curves achieve 2 levels of diversity with a horizontal spacing of 6 dB. These values agree with $g_{vb}(0) = 2$ and $t_{vb}(0) = 2$. Fig. 6.15, on the other hand, compares the outage behavior of 2×2 MIMO and ML V-BLAST schemes for $R = 4, 16$ and 32 BPCU. As can be seen from this figure, the outage curves for the two schemes coincide while $R/\log\rho > 1$. In particular, the curves corresponding to $R = 32$ BPCU are almost identical. However, for $R/\log\rho < 1$, the sub-optimality of the V-BLAST becomes evident. In fact, the curve corresponding to the 2×2 MIMO with $R = 4$ BPCU approaches 4 levels of diversity very rapidly, while the curve corresponding to the V-BLAST only attains 2 levels.

Similarly, orthogonal space-time constellations allow for a simple TRT characterization. An orthogonal constellation of size m, length l, and rate k/l (in symbols per channel use (spcu)) is a space-time code $\mathbf{X} \in \mathbb{C}^{m \times l}$ such that

$$\mathbf{X}\mathbf{X}^H = \left(\sum_{i=1}^{k} |x_i|^2 \right) \times \mathbf{I}_m, \tag{6.19}$$

where $\{x_i\}_{i=1}^k$ denote the symbols to be sent, $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{X}^H$ denotes the hermitian of matrix $\mathbf{X}$ [27]. As an example, consider the orthogonal constellation with $m = 2$, $l = 2$, and rate one, which is known as the Alamouti code [28]. In this case

$$\mathbf{X} = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix},$$

where x^* denotes the complex conjugate of x. Notice that

$$\mathbf{X}\mathbf{X}^H = (|x_1|^2 + |x_2|^2) \times \mathbf{I}_2,$$

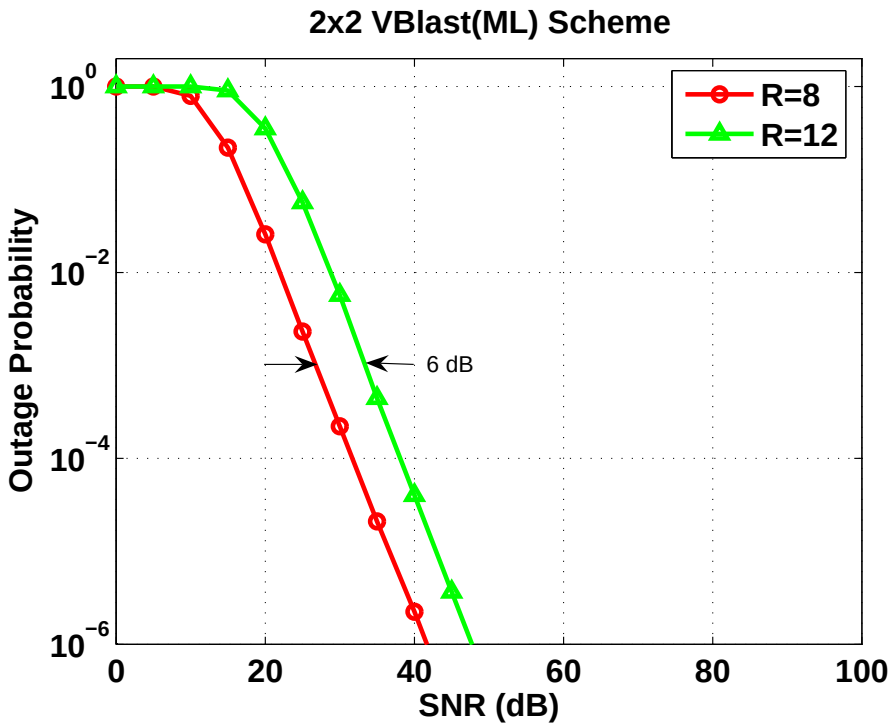

Figure 6.14: Outage curves corresponding to $R = 8, 12$ BPCU for a 2×2 V-BLAST scheme.

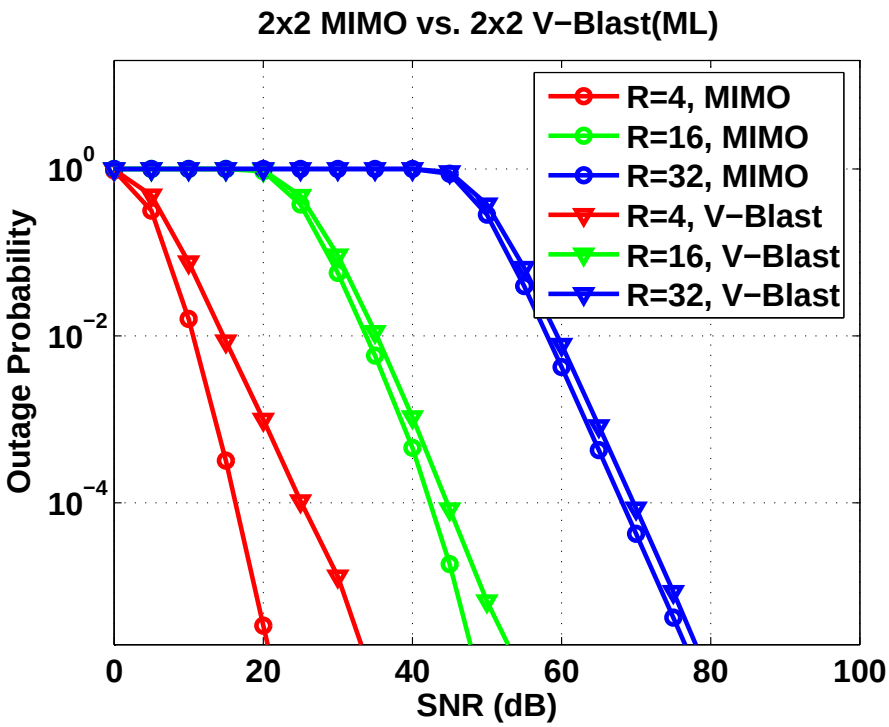

Figure 6.15: Comparison of outage curves corresponding to $R = 4, 16, 32$ BPCU for the 2×2 MIMO channel and the V-BLAST scheme.

as required by (6.19). The received signal matrix $\mathbf{Y} \in \mathbb{C}^{n \times l}$ at the destination can be written as

$$\mathbf{Y} = \sqrt{\frac{\rho}{m}}\mathbf{H}\mathbf{X} + \mathbf{W}. \tag{6.20}$$

The ML receiver performs linear processing on $\mathbf{Y}$ to yield the following equivalent parallel channel model

$$\tilde{y}_i = \sqrt{\frac{\rho}{m}\|\mathbf{H}\|^2}x_i + \tilde{w}_i, \quad \text{for} \quad i = 1, \cdots, k. \tag{6.21}$$

In (6.21), $\|\mathbf{H}\|^2$ denotes the Frobenius norm of $\mathbf{H}$ (i.e., $\|\mathbf{H}\|^2 \triangleq \sum |h_{ij}|^2$, where $\{h_{ij}\}$ are the entries of $\mathbf{H}$) and $\{\tilde{w}_i\}_{i=1}^k$ are i.i.d complex Gaussian random variables of zero mean and unit variance. The following theorem gives the throughput-reliability tradeoff for orthogonal constellations.

Theorem 19 *The optimal throughput-reliability tradeoff for an orthogonal constellation of size m, length l, rate k/l spcu (R BPCU) and n receive antennas satisfies*[9]

$$\lim_{\substack{\rho \to \infty \\ R \in \mathcal{R}_{oc}}} \frac{\log P_e(R, \rho) - \frac{l}{k}mnR}{\log \rho} = -mn, \tag{6.22}$$

where $\mathcal{R}_{oc}$ is given by

$$\mathcal{R}_{oc} \triangleq \{R | \frac{k}{l} > \frac{R}{\log \rho} > 0\},$$

Moreover, there exists an outer coding scheme (one that maps the information bits into symbols $\{x_i\}_{i=1}^k$) that achieves (6.22) *for $k \geq mn$.*

Proof: The proof follows immediately from Theorem 17 and the fact that the orthogonal constellation of interest effectively converts the underlying $m \times n$ MIMO

[9]The subscript "oc" stands for an orthogonal constellation.

channel of rate R (as given by (6.20)) into a $1 \times mn$ channel of rate $\frac{l}{k}R$ (as given by (6.21)).

Fig. 6.16 depicts the outage curves corresponding to $R = 4$ and 8 BPCU for a 2×2 Alamouti scheme. As can be seen from this figure, the high SNR segments of the outage curves achieve 4 levels of diversity with a horizontal spacing of 12 dB. These values agree with $g_{oc}(0) = 4$ and $t_{oc}(0) = 1$. Fig. 6.17, on the other hand, compares the outage behavior of the 2×2 MIMO channel and the Alamouti constellation for $R = 4, 16$ and 32 BPCU. As can be seen from this figure, while the outage curves for the two schemes coincide for small rates, the sub-optimality of the Alamouti scheme becomes evident at higher values of R. In particular, the curves corresponding to $R = 4$ BPCU are almost identical and for $R = 32$ BPCU the curve corresponding to the Alamouti scheme lags that of MIMO by more than 40 dB.

Finally, we extend our analysis to MIMO-ARQ channels. In this setup, the transmitter starts by picking up a message from the transmission buffer. It then uses a space-time encoder to map the message to a sequence of blocks $\mathbf{X}_p \in \mathbb{C}^{m \times l}$, $L \geq p \geq 1$. During transmission-round p, the transmitter sends $\mathbf{X}_p$ one column at a time over its m antennas. The receiver then tries to decode the message. If successful, it sends back a positive acknowledgement signal (ACK), which causes the transmitter to start sending the next message. However, if the receiver detects an error, it requests another round of transmission by feeding back a negative acknowledgement signal (NACK). The only exception to this rule is when L rounds of transmission have already been sent, in which case the transmitter abandons sending the current message and goes to the next one. In this chapter, we address the long-term static channel scenario, where all of the transmission-rounds corresponding to a message take place over the same

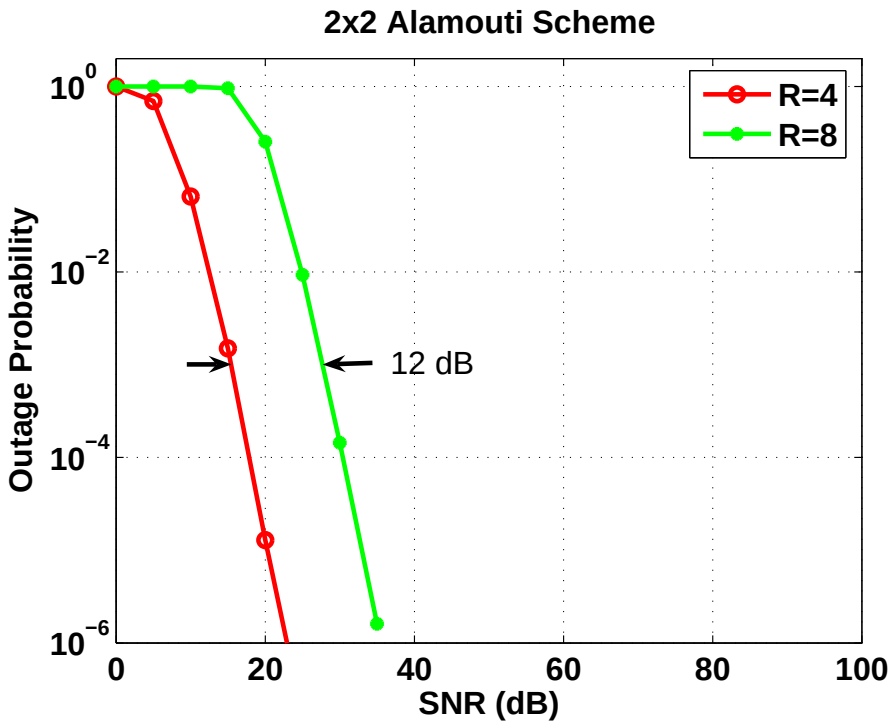

Figure 6.16: Outage curves corresponding to $R = 4, 8$ BPCU for the 2×2 Alamouti scheme.

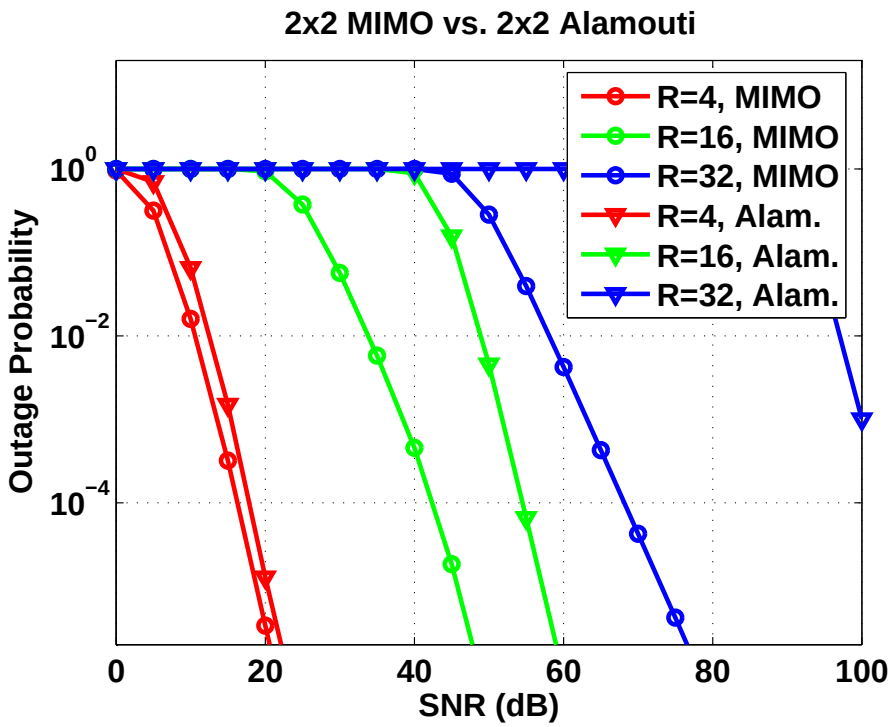

Figure 6.17: Comparison of outage curves corresponding to $R = 4, 16, 32$ BPCU for the 2×2 MIMO channel and the Alamouti scheme.

channel realization. We also impose a short term power constraint on the transmitter, such that power-control is not possible [29]. At this point, we need to distinguish between two closely related parameters, namely, the first-round transmission rate and the long-term average throughput. Assume that each message consists of b information bits which means that the first-round transmission rate is $R_1 = b/l$ BPCU. Since some of the messages take more than one transmission-round to be sent, the long-term average throughput η is strictly less than R_1. The gap between the two quantities, however, diminishes as the SNR grows. This is due to the fact that at high SNRs, most of the messages are decoded error-free after the first round of transmission and the ARQ retransmission-rounds are used only for those *rare* events in which the message does not get through with only one round of transmission. Recognizing the operational significance of η, in the following we state the TRT for ARQ channels in terms of η, rather than R_1.

Theorem 20 *The optimal throughput-reliability tradeoff for the coherent block-fading MIMO ARQ channel with m transmit antennas, n receive antennas, L maximum number of transmission-rounds, under the long-term static channel and short-term power constraint assumptions is given by*[10]

$$\lim_{\substack{\rho \to \infty \\ \eta \in \mathcal{R}_{ls}(k)}} \frac{\log P_e(\eta, \rho) - c_{ls}(k)\eta}{\log \rho} = -g_{ls}(k), \tag{6.23}$$

where η denotes the long-term average data rate. In (6.23)*, $\mathcal{R}_{ls}(k)$, $c_{ls}(k)$ and $g_{ls}(k)$ are defined by*

$$\mathcal{R}_{ls}(k) \triangleq \begin{cases} \{\eta | (k+1)L > \frac{\eta}{\log \rho} > kL\} & \text{for} \quad k \in \mathbb{Z}, \lfloor \frac{\min\{m,n\}}{L} \rfloor > k \geq 0 \\ \{\eta | \min\{m,n\} > \frac{\eta}{\log \rho} > \lfloor \frac{\min\{m,n\}}{L} \rfloor L\} & \text{for} \quad k = \lfloor \frac{\min\{m,n\}}{L} \rfloor \end{cases},$$

[10]The subscript "ls" stands for long-term static.

$$c_{ls}(k) \triangleq \frac{c(k)}{L} \quad \textit{and} \quad g_{ls}(k) \triangleq g(k),$$

respectively. $c(k)$ *and* $g(k)$ *are given by* (6.6) *and* (6.7).

Proof: (Sketch) The proof follows the same lines as that of Theorem 5 in [29]. In particular, the converse is obtained by lower-bounding the error probability of the ARQ protocol with that of a ML decoder that operates on the whole codeword $\{\mathbf{X_p}\}_{p=1}^{L}$. The achievability, on the other hand, is established through the use of an ensemble of Gaussian code-books, along with a bounded-distance decoder. The main idea here is to differentiate between the undetected-errors (i.e., the ones for which the receiver sends back an ACK signal) and those errors that the decoder makes after requesting L rounds of transmission. It can then be shown that, through judicious choice of decoder threshold-distance, the latter error type becomes dominant, and hence, the lower and upper bounds become tight. It is then straightforward to argue the existence of codes in the ensemble that perform at least as well as the ensemble average.

CHAPTER 7

CONCLUSIONS

We designed cooperation protocols for channels consisting of half-duplex nodes. We considered relay, CB, MAR and CMA scenarios and devised protocols in each case. We further demonstrated the gains and opportunities provided by ARQ techniques, through presenting DDF variants for the ARQ relay, MAR and CVMA channels. We evaluated the performance of the proposed protocols using DMT. We also presented a new formulation for the throughput-reliability tradeoff in MIMO channels that avoids the limitations imposed by the notion of multiplexing gain. In the sequel, we briefly review the main contributions of this work and suggest a few directions for possible future research.

7.1 Summary of Original Work

1. For the AF relay channel (single relay), we first established an upper bound on the achievable DMT. We then proposed a nonorthogonal amplify and forward (NAF) protocol that achieves this upper bound. Finally, we generalized the NAF protocol to the case of arbitrary number of relays and characterized its DMT. Notably, we showed that the NAF protocol outperforms the LW-STC protocol without requiring decoding/encoding at the relays.

2. For the DF relay channel (single relay), we proposed a dynamic decode and forward (DDF) protocol and proved its *optimality* (with respect to DMT), over the range of multiplexing gains $1/2 \geq r \geq 0$. Furthermore, we showed that the DDF protocol outperforms all AF protocols at any multiplexing gain. Finally, we extended the DDF protocol to the case of arbitrary number of relays and characterized its tradeoff curve.

3. We presented a novel variant of the (single relay) DDF protocol, which is particularly suitable for implementation. In this variant, the channel seen by the destination is reduced to a single-input single-output (SISO) time-selective channel. Furthermore, to reduce the complexity of the protocol, we confined the relay to start transmission only at the beginning of a *finite* number of segments. We gave the rule for determining these segments such that the variant is Pareto optimal with respect to DMT.

4. For the CB channel, we presented a variant of the DDF protocol that allows for efficient transmission of common information. We then established the superiority of this protocol, over AF protocols, by characterizing its tradeoff curve. We also argued that the gain offered by the DDF protocol is more significant in this scenario, as compared to the relay channel.

5. For the symmetric MAR channel (two users), we first derived an upper bound on the achievable DMT. We then modified the DDF protocol to match it to the MAR channel and derived its tradeoff curve. This characterization showed that the DDF protocol is DMT optimal over the range of multiplexing gains $3/4 \geq r \geq 0$. It also revealed that in the MAR channel, a *single* relay can be

utilized by *several* users to simultaneously improve the diversity gain achieved by *all* of them.

6. For the symmetric CMA channel, we proposed a novel AF protocol where an *artificial* inter-symbol interference (ISI) channel is created. We proved the optimality (in terms of DMT) of this protocol by showing that, for all multiplexing gains (i.e., $1 \geq r \geq 0$), it achieves the DMT of the corresponding $N \times 1$ MIMO channel. We then used this result to argue that the sub-optimality of the protocols proposed in [3] is a consequence of using orthogonal sub-channels, rather than the half-duplex constraint. We further utilized this result to elucidate the fundamental difference between the relay and CMA channels.

7. For the ARQ relay and MAR (two-user) channels, we first modified the corresponding DDF protocols to incorporate ACK/NACK feedback signals. We then characterized their achieved tradeoff curves, which proved their optimality.

8. For the ARQ CVMA channel (two users and two receiving antennas), we developed a new variant of the DDF protocol where the users are *purposefully* instructed not to cooperate in the first round of transmission. Lower and upper bounds on the achievable DMT were then derived. These bounds were shown to converge to the optimal tradeoff as the number of transmission rounds increases.

9. We identified the limitation imposed by the notion of multiplexing gain and develop a new formulation for the throughput-reliability tradeoff that avoids this limitation. In this formulation, the multiplexing gain notion is replaced by the more general concept of *operating* regions.

10. We used the proposed TRT formulation to elucidate the asymptotic trends exhibited by MIMO channels. In particular, we devised a piecewise linear approximation to the outage probability of MIMO channels (at *fixed* rates), which becomes progressively more accurate as rate and SNR grow.

11. We characterized the TRT, along with the corresponding piecewise linear outage approximation, for the V-BLAST, orthogonal constellations and ARQ MIMO channels.

7.2 Possible Future Work

In the sequel we give a few directions for possible future research.

- Proving (or disproving) the optimality of the DDF protocol for the single relay channel and $r > 0.5$ is an open problem.

- Generalizations of the proposed protocols to multi-antenna nodes, scenarios with different QoS constraints, and asymmetric channels are of definite interest.

- Extension of the proposed protocols to the non-coherent scenario, where all or part of CSI is not available to cooperating nodes and/or destination is very important.

- Characterization of the performance degredation resulting from asynchronicities among the cooperating partners is a rather valuable direction to follow.

- Design of practical coding/decoding strategies that approach the fundamental limits achievable through Gaussian codes and maximum likelihood decoding is an important venue to pursue [30, 31].

- The piecewaise-linear outage-probability approximation suggested by TRT gives accurate predictions when the operating point is well within certain regions of the rate-SNR plane. Characterization of the outage-probability in the transitional regions remains an open problem.

- For MIMO channels, we established the correspondence between the DMT and TRT formulations (refer to (6.8) and (6.9)). It remains to be seen if such a correspondence exits for general channels or not.

APPENDIX A

PROOF OF THE THEOREMS

A.1 Proof of Lemma 1

For simplicity, we prove the lemma for the real AWGN channel (the complex channel is a straightforward extension), i.e.

$$y_k = hx_k + n_k \quad \text{for} \quad l \geq k \geq 1,$$

where y_k is the received signal over the k^{th} symbol-interval, h is the channel gain and n_k is the i.i.d. Gaussian noise sample during symbol-interval k. Next we characterize the average error probability $P_{E|h}$, where averaging is invoked with respected to the ensemble of Gaussian codes, conditioned on a particular h (which is assumed to be known at the decoder). Towards this end, we divide a codeword into N segments, each consisting of T symbol-intervals, i.e. $l = NT$. In our analysis, we consider the asymptotic scenario where N and T grow to infinity. The ML decoder waits for N' segments before it starts decoding, where N' is given by

$$N' = \left\lfloor \frac{NR}{I(x_k; y_k)} \right\rfloor + 1, \tag{A.1}$$

In (A.1), $I(x_k; y_k)$ denotes the mutual information between x_k and y_k. Notice that assuming the data rate to be less than the channel capacity guarantees $N' \leq N$. Also,

observe that the fraction of codeword that the decoder has to wait before decoding satisfies

$$\lim_{N\to\infty} \frac{N'}{N} = \frac{R}{I(x_k; y_k)}.$$

Our analysis is based on *exactly* the same techniques used to establish the achievability of the AWGN channel capacity, i.e., we first upper-bound the ML error probability with that of a typical set decoder and then take the average of the latter error probability, over the ensemble of random Gaussian code-books. In particular, following *exactly* the same argument as in Theorem 10.1.1 in [25], we get

$$\begin{aligned} P_{E|h} &\leq 2\epsilon + 2^{3N'T\epsilon} 2^{-T(N'I(x_k;y_k)-NR)}, \\ &\leq 3\epsilon, \end{aligned}$$

for sufficiently large N and T. This follows from the choice of N' in (A.1). It is important to note that the transmission rate R is constant, i.e., independent of h, and N' is *not* known to the transmitter.

A.2 Proof of Theorem 2

Due to the source average energy constraint, setting A_1 and A_2 to anything other than the identity matrix will reduce the mutual information between $\mathbf{x}$ and $\mathbf{y}$. Since we are interested in obtaining an upper bound, we will choose $A_1 = I_{l'}$ and $A_2 = I_{l-l'}$, in which case (3.1) reduces to

$$\mathbf{y} = \begin{bmatrix} g_1 I_{l'} & 0 \\ g_2 hB & g_1 I_{l-l'} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ g_2 B \end{bmatrix} \mathbf{w} + \mathbf{v}. \tag{A.2}$$

Using singular value decomposition (SVD), the matrix B can be factored as

$$B = UDV^H,$$

where $U \in \mathbb{C}^{(l-l')\times(l-l')}$ and $V \in \mathbb{C}^{l'\times l'}$ are unitary and where $D \in \mathbb{C}^{(l-l')\times l'}$ is non-negative diagonal with the diagonal elements in decreasing order. Using these matrices, we define $\tilde{\mathbf{y}} \triangleq T\mathbf{y}$, $\tilde{\mathbf{x}} \triangleq T\mathbf{x}$, $\tilde{\mathbf{v}} \triangleq T\mathbf{v}$, and $\tilde{\mathbf{w}} \triangleq V^H\mathbf{w}$, for unitary transformation

$$T \triangleq \begin{bmatrix} V^H & 0 \\ 0 & U^H \end{bmatrix}.$$

The unitary property of V and T implies that $\Sigma_{\tilde{\mathbf{w}}} = \sigma_w^2 I_l$ and $\Sigma_{\tilde{\mathbf{v}}} = \sigma_v^2 I_l$, as well as

$$I(\mathbf{x};\mathbf{y}) = I(\tilde{\mathbf{x}};\tilde{\mathbf{y}}). \tag{A.3}$$

In terms of the new variables, (A.2) becomes

$$\begin{aligned} \tilde{\mathbf{y}} &= \begin{bmatrix} g_1 I_{l'} & 0 \\ g_2 hD & g_1 I_{l-l'} \end{bmatrix} \tilde{\mathbf{x}} + \begin{bmatrix} 0 \\ g_2 D \end{bmatrix} \tilde{\mathbf{w}} + \tilde{\mathbf{v}} \\ &= \begin{bmatrix} g_1 I_{l'} & 0 \\ g_2 hD & g_1 I_{l-l'} \end{bmatrix} \tilde{\mathbf{x}} + \tilde{\mathbf{n}} \end{aligned} \tag{A.4}$$

with

$$\Sigma_{\tilde{\mathbf{n}}} = \begin{bmatrix} \sigma_v^2 I_{l'} & 0 \\ 0 & \sigma_v^2 I_{l-l'} + |g_2|^2 \sigma_w^2 DD^H \end{bmatrix}.$$

If we denote the non-zero diagonal elements of D as $\{d_i\}_{i=1}^m$, then (A.4) can be written as

$$\begin{aligned} \tilde{\mathbf{y}}_i &= G_i \tilde{\mathbf{x}}_i + \tilde{\mathbf{n}}_i, \quad i = 1, \ldots, m \\ \tilde{y}_i &= g_1 \tilde{x}_i + \tilde{n}_i, \quad i = m+1, \ldots, l' \text{ and } i = l'+m+1, \ldots, l, \end{aligned}$$

where $\tilde{y}_i$, $\tilde{x}_i$ and $\tilde{n}_i$ represent the i^{th} element of $\tilde{\mathbf{y}}$, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{n}}$, respectively, and where $\tilde{\mathbf{y}}_i \triangleq [\tilde{y}_i, \tilde{y}_{l'+i}]^t$, $\tilde{\mathbf{x}}_i \triangleq [\tilde{x}_i, \tilde{x}_{l'+i}]^t$, $\tilde{\mathbf{n}}_i \triangleq [\tilde{n}_i, \tilde{n}_{l'+i}]^t$, and

$$G_i \triangleq \begin{bmatrix} g_1 & 0 \\ g_2 h d_i & g_1 \end{bmatrix}, \tag{A.5}$$

$$\Sigma_{\tilde{\mathbf{n}}_i} = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_v^2 + |g_2|^2 d_i^2 \sigma_w^2 \end{bmatrix}. \tag{A.6}$$

Note that, according to the SVD theorem,

$$m \leq \min\{l', l - l'\}. \tag{A.7}$$

Because $\Sigma_{\tilde{\mathbf{n}}}$ is diagonal, $I(\tilde{\mathbf{x}}; \tilde{\mathbf{y}})$ (and therefore $I(\mathbf{x}; \mathbf{y})$) is maximized when $\{\tilde{\mathbf{x}}_i\}_{i=1}^{m} \cup \{\tilde{x}_i\}_{i=m+1}^{l'} \cup \{\tilde{x}_i\}_{i=l'+m+1}^{l}$ are mutually independent, in which case we would have

$$\max_{\Sigma_{\tilde{\mathbf{x}}}} I(\tilde{\mathbf{x}}; \tilde{\mathbf{y}}) = \sum_{i=1}^{m} \max_{\Sigma_{\tilde{\mathbf{x}}_i}} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i) + \sum_{i=m+1}^{l'} \max I(\tilde{x}_i; \tilde{y}_i) + \sum_{i=l'+m+1}^{l} \max I(\tilde{x}_i; \tilde{y}_i). \tag{A.8}$$

The mutual information between $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{y}}_i$ is given by

$$I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i) = \log(\det\,(I_2 + \Sigma_{\tilde{\mathbf{n}}_i}^{-\frac{1}{2}} G_i \Sigma_{\tilde{\mathbf{x}}_i} G_i^H \Sigma_{\tilde{\mathbf{n}}_i}^{-\frac{1}{2}})).$$

A lower-bound on $\max_{\Sigma_{\tilde{\mathbf{x}}_i}} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i)$ is easily obtained by replacing $\Sigma_{\tilde{\mathbf{x}}_i}$ by EI_2:

$$\log(\det\,(I_2 + EG_i G_i^H \Sigma_{\tilde{\mathbf{n}}_i}^{-1})) \leq \max_{\Sigma_{\tilde{\mathbf{x}}_i}} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i). \tag{A.9}$$

Since $\log(\det\,(.))$ is an increasing function on the cone of positive-definite Hermitian matrices and since $\lambda_{\max} I_2 - \Sigma_{\tilde{\mathbf{x}}_i} \geq 0$ (where $\lambda_{\max}$ represents the largest eigenvalue of $\Sigma_{\tilde{\mathbf{x}}}$), we get the following upper-bound on $\max_{\Sigma_{\tilde{\mathbf{x}}_i}} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i)$:

$$\max_{\Sigma_{\tilde{\mathbf{x}}_i}} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i) \leq \log(\det\,(I_2 + \lambda_{\max} G_i G_i^H \Sigma_{\tilde{\mathbf{n}}_i}^{-1})). \tag{A.10}$$

From (A.9) and (A.10), we conclude that

$$\frac{\log(\det\,(I_2 + EG_i G_i^H \Sigma_{\tilde{\mathbf{n}}_i}^{-1}))}{\log(\rho)} \leq \frac{\max_{\Sigma_{\tilde{\mathbf{x}}_i}} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i)}{\log(\rho)} \leq \frac{\log(\det\,(I_2 + \lambda_{\max} G_i G_i^H \Sigma_{\tilde{\mathbf{n}}_i}^{-1}))}{\log(\rho)}.$$

Now, since $\lambda_{\max}$ is of the same exponential order as E, the bounds converge as ρ grows to infinity. That is

$$\lim_{\rho \to \infty} \frac{\max_{\Sigma_{\tilde{\mathbf{x}}_i}, d_i} I(\tilde{\mathbf{x}}_i; \tilde{\mathbf{y}}_i)}{\log(\rho)} = \lim_{\rho \to \infty} \frac{\log(\det\,(I_2 + EG_i G_i^H \Sigma_{\tilde{\mathbf{n}}_i}^{-1}))}{\log(\rho)}.$$

Plugging in for G_i and $\Sigma_{\tilde{\mathbf{n}}_i}$ from (A.5) and (A.6), respectively, we get

$$\lim_{\rho\to\infty} \frac{\max_{\Sigma_{\tilde{\mathbf{x}}_i},d_i} I(\tilde{\mathbf{x}}_i;\tilde{\mathbf{y}}_i)}{\log(\rho)} = \lim_{\rho\to\infty} \frac{1}{\log(\rho)} \log(1+\frac{|g_1|^2E}{\sigma_v^2}+\cdots \\ \frac{(|g_1|^2+|g_2|^2|h|^2|d_i|^2)E}{\sigma_v^2+|g_2|^2d_i^2\sigma_w^2}+\frac{|g_1|^4E^2}{\sigma_v^2(\sigma_v^2+|g_2|^2d_i^2\sigma_w^2)}).$$

It is then straightforward to see that

$$\lim_{\rho\to\infty} \frac{\max_{\Sigma_{\tilde{\mathbf{x}}_i},d_i} I(\tilde{\mathbf{x}}_i;\tilde{\mathbf{y}}_i)}{\log(\rho)} = (\max\{2(1-v_1),1-(v_2+u)\})^+, \tag{A.11}$$

where v_1, v_2 and u are the exponential orders of $1/|g_1|^2$, $1/|g_2|^2$ and $1/|h|^2$, respectively. In deriving this expression, we have assumed that $(v_1,v_2,u) \in \mathbb{R}^{3+}$; as explained earlier, we do not need to consider realizations in which v_1, v_2 or u are negative. Similarly,

$$\lim_{\rho\to\infty} \frac{\max I(\tilde{x}_i;\tilde{y}_i)}{\log(\rho)} = (1-v_1)^+,$$

which, together with (A.11), (A.3) and (A.8), results in:

$$\lim_{\rho\to\infty} \frac{\max_{\Sigma_{\mathbf{x}}} I(\mathbf{x};\mathbf{y})}{\log(\rho)} = (l-2m)(1-v_1)^+ + m(\max\{2(1-v_1),1-(v_2+u)\})^+. \tag{A.12}$$

For the quasi-static fading setup, the outage event is defined as the set of channel realizations for which the instantaneous capacity falls below the target data rate. Thus, our outage event O becomes

$$O = \{(v_1,v_2,u) | \max_{\Sigma_{\mathbf{x}}} I(\mathbf{x},\mathbf{y}) < lR\}.$$

Letting R grow with ρ according to

$$R = r\log(\rho),$$

and using (A.12), we conclude that, for large ρ,

$$O^+ = \{(v_1, v_2, u) \in \mathbb{R}^{3+} | (l-2m)(1-v_1)^+ + \cdots \\ m(\max\{2(1-v_1), 1-(v_2+u)\})^+ < rl\}, \tag{A.13}$$

and thus

$$P_O(R) \doteq \rho^{-d_o(r)} \quad \text{for} \quad d_o(r) = \inf_{(v_1,v_2,u)\in O^+} (v_1 + v_2 + u). \tag{A.14}$$

As Zheng and Tse have shown in Lemma 5 of [17], $d_o(r)$ provides an upper-bound on $d^*(r)$ (i.e., the optimal diversity gain at multiplexing gain r):

$$d^*(r) \leq d_o(r). \tag{A.15}$$

From (A.13) and (A.14), it is easy to see that the right hand side of (A.15) is maximized when m is set to its maximum, which, according to (A.7), is $\min\{l', l-l'\}$. This is the case when B is full-rank. On the other hand, $\min\{l', l-l'\}$ itself is maximized when $l' = l/2$ (assuming an even codeword length l), which corresponds to B being a square matrix. For this B, $d_o(r)$ can be shown to take the value of the right hand side of (3.3). This completes the proof.

A.3 Proof of Theorem 3

The proof closely follows that for the MIMO point-to point communication system in [17]. In particular, we assume that the source uses a Gaussian random code-book of codeword length l, where l is taken to be even, and data rate R, where R increases with ρ according to

$$R = r \log(\rho).$$

The error probability of the ML decoder, $P_E(\rho)$, can be upper bounded using Bayes' rule:

$$P_E(\rho) = P_O(R)P_{E|O} + P_{E,O^c}$$

$$P_E(\rho) \leq P_O(R) + P_{E,O^c},$$

where O denotes the outage event. The outage event O is chosen such that $P_O(R)$ dominates P_{E,O^c}, i.e.,

$$P_{E,O^c} \dot{\leq} P_O(R), \tag{A.16}$$

in which case

$$P_E(\rho) \dot{\leq} P_O(R). \tag{A.17}$$

In order to characterize O, we note that, since the destination observations during different frames are independent, the upper-bound on the ML conditional PEP [recalling (2.7)], assuming l to be even, changes to

$$P_{PE|g_1,g_2,h} \leq \det\left(I_2 + \frac{1}{2}\Sigma_s\Sigma_n^{-1}\right)^{-l/2}, \tag{A.18}$$

where $\Sigma_{\mathbf{s}}$ and $\Sigma_{\mathbf{n}}$ denote the covariance matrices of destination observation's signal and noise components during a single frame:

$$\Sigma_{\mathbf{s}} = \begin{bmatrix} |g_1|^2 & g_1 g_2^* b^* h^* \\ g_1^* g_2 bh & |g_1|^2 + |g_2|^2|bh|^2 \end{bmatrix} E \tag{A.19}$$

$$\Sigma_{\mathbf{n}} = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_v^2 + |g_2|^2|b|^2\sigma_w^2 \end{bmatrix}. \tag{A.20}$$

Let us define v_1, v_2, u, and w as the exponential orders of $1/|g_1|^2$, $1/|g_2|^2$, $1/|h|^2$, and $|b|^2$, respectively. Then the constraint on b given in (3.4) implies the following constraint on w:

$$w \leq \min\{u, 1\} \tag{A.21}$$

We assume b is chosen such that the exponential order w becomes

$$w \triangleq (u)^{-}.$$

which satisfies the constraint given by (A.21). Interestingly, if we consider $(v_1, v_2, u) \in \mathbb{R}^{3+}$, then w becomes zero and vanishes in the expressions. Plugging (A.19)-(A.20) into (A.18), we obtain

$$P_{PE|v_1,v_2,u} \dot{\leq} \rho^{-\frac{l}{2}(\max\{2(1-v_1),1-(v_2+u)\})^{+}} \quad \text{for} \quad (v_1, v_2, u) \in \mathbb{R}^{3+}.$$

With rate $R = r \log \rho$ BPCU and codeword length l, we have a total of ρ^{rl} codewords. Thus,

$$P_{E|v_1,v_2,u} \dot{\leq} \rho^{-\frac{l}{2}[(\max\{2(1-v_1),1-(v_2+u)\})^{+}-2r]} \quad \text{for} \quad (v_1, v_2, u) \in \mathbb{R}^{3+}.$$

P_{E,O^c} is the average of $P_{E|v_1,v_2,u}$ over the set of channel realizations that do not cause an outage (i.e., O^c). Using (2.4), one can see that

$$P_{E,O^c} \dot{\leq} \int_{O^{c+}} \rho^{-d_e(r,v_1,v_2,u)} dv_1 dv_2 du.$$

for

$$d_e(r, v_1, v_2, u) = \frac{l}{2}[(\max\{2(1-v_1), 1-(v_2+u)\})^{+} - 2r] + (v_1 + v_2 + u).$$

Now, P_{E,O^c} is dominated by the term corresponding to the minimum value of $d_e(r, v_1, v_2, u)$ over O^{c+}:

$$P_{E,O^c} \dot{\leq} \rho^{-d_e(r)} \quad \text{for} \quad d_e(r) = \inf_{v_1,v_2,u \in O^{c+}} d_e(r, v_1, v_2, u). \tag{A.22}$$

Using (2.5), $P_O(R)$ can be expressed

$$P_O \dot{=} \rho^{-d_o(r)} \quad \text{for} \quad d_o(r) = \inf_{(v_1,v_2,u) \in O^{+}} (v_1 + v_2 + u). \tag{A.23}$$

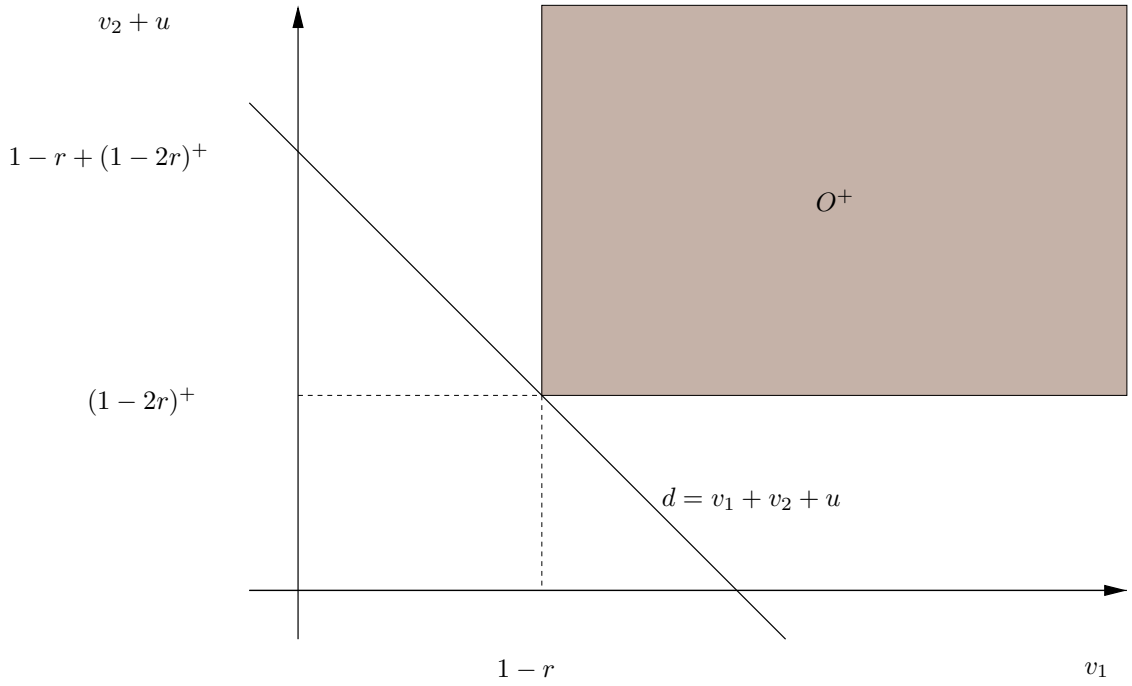

Figure A.1: Outage Region for the NAF protocol with a single relay.

Comparing (A.22) and (A.23), we realize that for (A.16) to be met, O^+ should be defined as

$$O^+ = \{(v_1, v_2, u) \in \mathbb{R}^{3+} | (\max\{2(1 - v_1), 1 - (v_2 + u)\})^+ \leq 2r\}.$$

Then, for any $(v_1, v_2, u) \in O^{c+}$, it is possible to choose l to make $d_e(r, v_1, v_2, u)$ arbitrarily large, ensuring (A.16). Note that, because of (A.17), $d_o(r)$ provides a lower-bound on the diversity gain achieved by the protocol. But $d_o(r)$, as given by (A.23), turns out to be identical to right hand side of (3.3) (refer to Fig. A.1). Thus the optimal DMT for this scenario is indeed given by (3.5) and the NAF protocol achieves it.

A.4 Proof of Theorem 5

Instead of considering specific codes, in the following we upper bound the average probability of error over random Gaussian ensemble of code-books (employed by both the source and relay). Therefore, averaging is invoked with respect to both, the fading channel distribution and the random code-books. It is then straightforward to see that there is at least one code-book in this ensemble whose average performance, now with respect only to the fading channel distribution, is better than the predictions of our upper bounds. For the single relay DDF protocol, the error probability of the ML decoder, averaged over the ensemble of Gaussian code-books and conditioned on a certain channel realization, can be upper bounded using Bayes' rule to give

$$P_{E|g_1,g_2,h} = P_{E,E_r^c|g_1,g_2,h} + P_{E,E_r|g_1,g_2,h}$$

$$P_{E|g_1,g_2,h} \leq P_{E|E_r^c,g_1,g_2,h} + P_{E_r|g_1,g_2,h},$$

where E_r and E_r^c denote the events that the relay decodes source's message erroneously and its complement, respectively. The first step in the proof follows from Lemma 1 by observing that if (3.6) is met, i.e., if the mutual information between the signal transmitted by the source and the signal received by the relay exceeds lR, then $P_{E_r|g_1,g_2,h}$ can be made arbitrarily small, provided that the code-length is sufficiently large. This means that for any $\epsilon > 0$ and for a sufficiently large code-length,

$$P_{E|g_1,g_2,h} < P_{E|E_r^c,g_1,g_2,h} + \epsilon.$$

Taking the average over the ensemble of channel realizations gives

$$P_E < P_{E|E_r^c} + \epsilon,$$

$$P_E \dot{\leq} P_{E|E_r^c}.$$

This means that the exponential order of $P_{E|E_r^c}$, i.e., destination's ML error probability assuming *error-free* decoding at the relay, provides a lower-bound on the diversity gain achieved by the protocol. Therefore, we only need to characterize $P_{E|E_r^c}$, which for the sake of notational simplicity, we will denote by P_E in the sequel. To characterize P_E, we note that the corresponding PEP [recalling (2.7)] is given by

$$P_{PE|g_1,g_2,h} \leq \left(1+|g_1|^2\frac{E}{2\sigma_v^2}\right)^{-l'}\left(1+\left(|g_1|^2+|g_2|^2\right)\frac{E}{2\sigma_v^2}\right)^{-(l-l')}. \tag{A.24}$$

Defining v_1, v_2, and u as the exponential orders of $1/|g_1|^2$, $1/|g_2|^2$, and $1/|h|^2$, respectively, gives

$$P_{PE|v_1,v_2,u}\dot{\leq}\rho^{-l[f(1-v_1)^+ +(1-f)(1-\min\{v_1,v_2\})^+]} \text{ for } (v_1,v_2,u)\in\mathbb{R}^{3+},$$

where $f \triangleq l'/l$. At a rate of $R = r\log\rho$ BPCU and a codeword length of l, there are a total of ρ^{rl} codewords. Thus,

$$P_{E,O^c}\dot{\leq}\rho^{-d_e(r)}$$

for

$$d_e(r) = \inf_{(v_1,v_2,u)\in O^{c+}} l[f(1-v_1)^+ + (1-f)(1-\min\{v_1,v_2\})^+ - r] + (v_1+v_2+u) \tag{A.25}$$

Examining (A.25), we realize that for (A.16) to hold, O^+ should be defined as

$$O^+ = \{(v_1,v_2,u)\in\mathbb{R}^{3+} | f(1-v_1)^+ + (1-f)(1-\min\{v_1,v_2\})^+ \leq r\} \tag{A.26}$$

so that it is possible to choose l to make $d_e(r)$ arbitrarily large, ensuring (A.16). As before, $P_O(R)$ is given by (A.23), which turns out to be identical to $d(r)$ given by (3.7). To see this, one needs to consider four different categories of channel realizations. The

first category is when both, v_1 and v_2 are greater than one. For this category,

$$\inf_{\substack{(v_1,v_2,u)\in O^+,\\ v_1>1,v_2>1}} (v_1+v_2+u) = 2. \tag{A.27}$$

The second category is when $1 \geq v_1 \geq 0$ and $v_2 > 1$. It is easy to see from (A.26) that for this category,

$$\inf_{\substack{(v_1,v_2,u)\in O^+,\\ 1\geq v_1\geq 0,v_2>1}} (v_1+v_2+u) = 2-r. \tag{A.28}$$

The third category to be considered is when $v_1 > 1$ and $1 \geq v_2 \geq 0$. Before proceeding further, note that from (3.6), one can show that

$$u = 1 - \frac{r}{f}. \tag{A.29}$$

This implies that $f \geq r$, since u is nonnegative. Returning back to (A.26), it is easy to verify that for this category

$$v_2 \geq 1 - \frac{r}{1-f}. \tag{A.30}$$

Now, if $f \geq \max\{r, 1-r\}$, then from (A.30) and (A.29) we get

$$\inf_{\substack{(v_1,v_2,u)\in O^+,\\ v_1>1,1\geq v_2\geq 0,\\ f\geq\max\{r,1-r\}}} (v_1+v_2+u) = \inf_{f\geq\max\{r,1-r\}} 2 - \frac{r}{f},$$

or

$$\inf_{\substack{(v_1,v_2,u)\in O^+,\\ v_1>1,1\geq v_2\geq 0,\\ f\geq\max\{r,1-r\}}} (v_1+v_2+u) = \begin{cases} 1+\frac{1-2r}{1-r}, & \frac{1}{2}\geq r\geq 0, \\ 1, & 1\geq r\geq \frac{1}{2}. \end{cases} \tag{A.31}$$

On the other hand, if $1 - r > f \geq r$, then

$$\inf_{\substack{(v_1,v_2,u)\in O^+,\\ v_1>1,1\geq v_2\geq 0,\\ 1-r>f\geq r}} (v_1+v_2+u) = \inf_{1-r>f\geq r} 3 - \frac{r}{1-f} - \frac{r}{f},$$

or

$$\inf_{\substack{(v_1,v_2,u)\in O^+,\\ v_1>1,1\geq v_2\geq 0,\\ 1-r>f\geq r}} (v_1+v_2+u) = 1+\frac{1-2r}{1-r} \quad \text{for} \quad \frac{1}{2} > r \geq 0.$$

This means that $\inf_{O^+}(v_1+v_2+u)$, for the third category, is indeed given by (A.31). It is noteworthy that the tradeoff curves given by (A.27), (A.28) and (A.31), are all better than the genie-aided tradeoff. In other words, the diversity gain achieved by this protocol is determined by the fourth category, where, both v_1 and v_2 are less than or equal to one. For this category, one needs to consider two cases (Note that (A.29) is still valid, implying $f \geq r$). The first case, when $0.5 \geq f \geq r$, is very easy. Referring to Fig. A.2 reveals that, in this case, $\inf_{(v_1,v_2)\in O^+} v_1+v_2$ and therefore $d_o(r)$ is equal to $2(1-r)$ (the genie aided tradeoff). The second case, when $f > \max\{r, 0.5\}$, is a little bit more difficult. As can be seen from Fig. A.3, in this case

$$\inf_{(v_1,v_2)\in O^+} (v_1+v_2) = \frac{1-r}{f}. \tag{A.32}$$

From (A.29) and (A.32), we conclude that

$$d_o(r) = \inf_{f>\max\{r,0.5\}} 1+\frac{1-2r}{f},$$

which gives (3.7). Again, according to (A.17), $d_o(r)$ provides a lower-bound on the diversity gain achieved by the protocol. On the other hand, $d_o(r)$ is also an upper bound on the achieved diversity since: 1) for $0 \leq r \leq 0.5$ $d_o(r)$ is the genie-aided diversity and 2) for $0.5 \leq r \leq 1$ it is easy to see that $v_1 = \frac{1-r}{r}+\epsilon$, $v_2 = 0$ and $u = 0$ correspond to a *channel* outage for any $\epsilon > 0$. Thus (3.7) is the diversity achieved by the DDF protocol and the proof is complete.

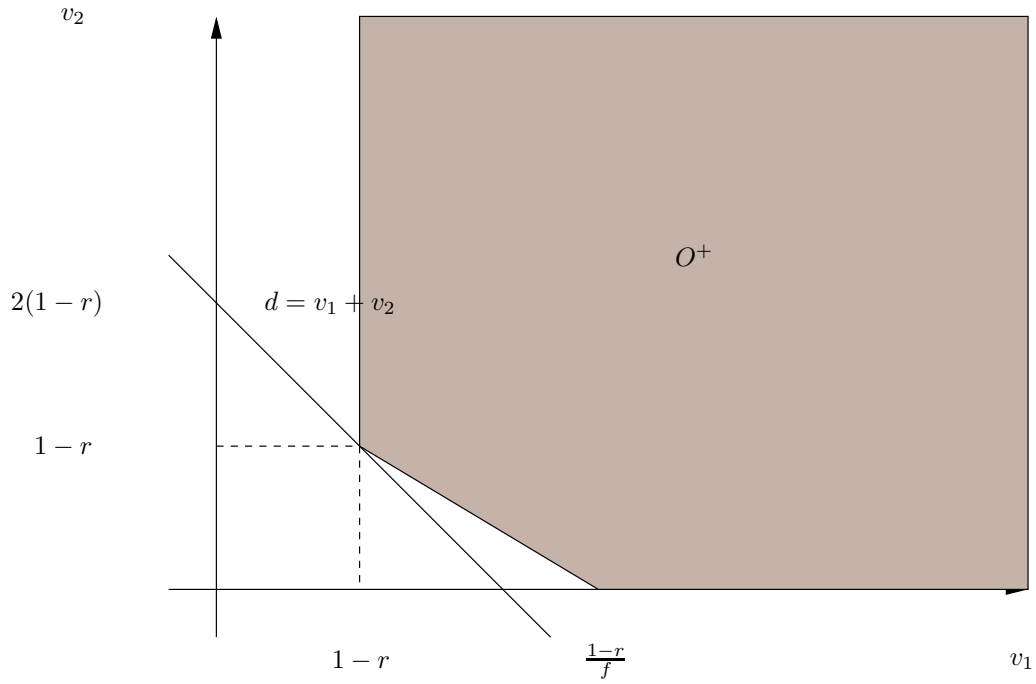

Figure A.2: Outage Region for the DDF protocol with a single relay ($f \leq 0.5$).

A.5 Proof of Theorem 6

Inspired by the single-relay case, we use ensembles of Gaussian code-books at the source and all the relays. To characterize the DMT achieved by the DDF protocol with $N-1$ relays, we first label the nodes according to the order in which they start transmission. That is, the source is labelled as node 1, the first relay that starts transmission as node 2, and so on. We then use Bayes' rule to upper bound the error probability of the ML decoder, averaged over the ensemble of Gaussian code-books and conditioned on a certain channel realization, to get

$$P_{E|g_j,h_{ji}} \leq P_{E|\{E_p^c\}_{p=2}^N,g_j,h_{ji}} + \sum_{n=2}^{N} P_{E_n|\{E_p^c\}_{p<n},g_j,h_{ji}}, \tag{A.33}$$

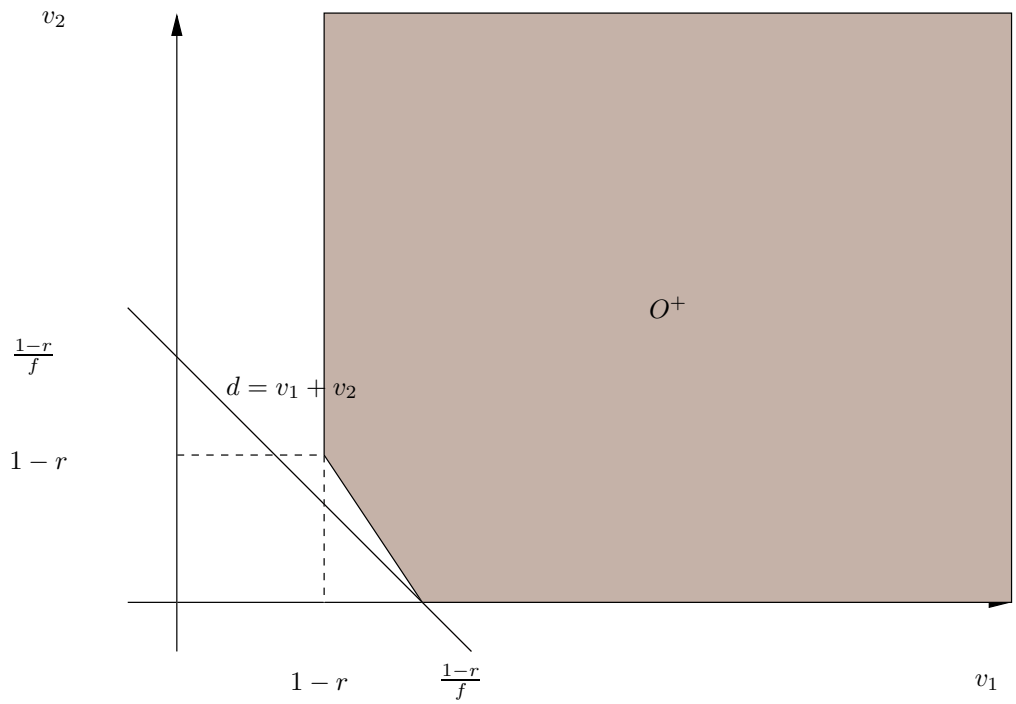

Figure A.3: Outage Region for the DDF protocol with a single relay ($f > 0.5$).

where $E_n, n \in \{2, \cdots, N\}$ denotes the event that node n decodes the source message in error, while E_n^c denotes its complement. Let us now examine $P_{E_n|\{E_p^c\}_{p<n},g_j,h_{ji}}$, i.e., the probability that node $n \in \{2, \cdots, N\}$ makes an error in decoding the source message, assuming error-free decoding at all previous nodes. It follows from Lemma 1, that if the mutual information between the signals transmitted by the source and active relays and the signal received by node n exceeds lR, then $P_{E_n|\{E_p^c\}_{p<n},g_j,h_{ji}}$ can be made arbitrarily small, provided that the code-length is sufficiently large. This means that for any $\epsilon > 0$ and for sufficiently large code-lengths,

$$P_{E_n|\{E_p^c\}_{p<n},g_j,h_{ji}} < \epsilon, \quad n \in \{2, \cdots, N\}. \tag{A.34}$$

Using (A.34), (A.33) can be written as

$$P_{E|g_j,h_{ji}} \leq P_{E|\{E_p^c\}_{p=2}^N,g_j,h_{ji}} + (N-1)\epsilon.$$

Taking the average over the ensemble of channel realizations gives

$$P_E < P_{E|\{E_p^c\}_{p=2}^N} + (N-1)\epsilon,$$

$$P_E \dot{\leq} P_{E|\{E_p^c\}_{p=2}^N}.$$

This means that the exponential order of $P_{E|\{E_p^c\}_{p=2}^N}$, i.e., destination's ML error probability assuming *error-free* decoding at all of the relays, provides a lower-bound on the diversity gain achieved by the protocol. Therefore, we only need to characterize $P_{E|\{E_p^c\}_{p=2}^N}$, which for the sake of notational simplicity, we will denote by P_E in the sequel. To characterize P_E, we note that the corresponding PEP, is upper-bounded by

$$P_{PE|g_j,h_{ji}} \leq \prod_{j=1}^{N} \left[1 + \left(\sum_{i=1}^{j} |g_j|^2\right) \frac{E}{2\sigma_v^2}\right]^{-l_j}.$$

As before, the gain of the channel that connects the j^{th} node to the destination is denoted by g_j, while the gain of the channel that connects nodes i and j is denoted by h_{ji}. We use l_j to denote the number of symbol-intervals in the codeword during which a total of j nodes are transmitting, so that $\sum_{j=1}^{N} l_j = l$, with l denoting the total codeword length. Note that $\sum_{j=1}^{p} l_j$ is the number of symbol-intervals that relay $p+1$ has to wait, before the mutual information between its received signal and the signals that the source and other relays transmit exceeds lR. Thus

$$\sum_{j=1}^{p} l_j \leq \ \min\{l, \lceil \frac{lR}{\log(1+|h_{p+1,1}|^2 c\rho)} \rceil\}, \quad \text{for} \quad N-1 \geq p \geq 1. \tag{A.35}$$

Defining v_j and u_{ji} as the exponential orders of g_j and h_{ji}, respectively, we have

$$P_{PE|v_j,u_{ji}} \dot{\leq} \rho^{-\sum_{j=1}^{N} l_j (1-\min\{v_1,\ldots,v_j\})^+}.$$

Choosing $R = r\log(\rho)$ for a total of ρ^{rl} codewords, the following expression for the conditional error probability can be derived.

$$P_{E|v_j,u_{ji}} \dot{\leq} \rho^{-l\left[\sum_{j=1}^{N} \frac{l_j}{l}(1-\min\{v_1,\ldots,v_j\})^+ - r\right]}.$$

Thus, O^+ is the set of channel realizations that satisfy

$$\sum_{j=1}^{N} \frac{l_j}{l}(1-\min\{v_1,\ldots,v_j\})^+ \leq r,$$

which can be simplified to

$$1-r \leq \sum_{j=1}^{N} \frac{l_j}{l} \min\{1, v_1, \ldots, v_j\}. \tag{A.36}$$

As before, $P_O(R)$ is characterized by

$$P_O(R) \doteq \rho^{-d_o(r)} \quad \text{for} \quad d_o(r) = \inf_{O^+} \sum_{j=1}^{N} \left(v_j + \sum_{i<j} u_{ji} \right). \tag{A.37}$$

Defining $\tilde{v}_j \triangleq \min\{v_1, \cdots, v_j\}, j = 1, \cdots, N$ lets us simplify (A.36) and (A.37) to

$$1 - r \leq \sum_{j=1}^{N} \frac{l_j}{l} \min\{1, \tilde{v_j}\} \tag{A.38}$$

$$d_o(r) \geq \inf_{O^+} \sum_{j=1}^{N} \left(\tilde{v}_j + \sum_{i<j} u_{ji} \right).$$

From the definition of $\tilde{v_j}$, it follows that

$$\tilde{v_1} \geq \tilde{v_2} \geq \cdots \geq \tilde{v_N} \geq 0.$$

Note that (A.35) can also be simplified to

$$\sum_{j=1}^{p} \frac{l_j}{l} \leq \min\Big\{1, \frac{r}{(1 - u_{p+1,1})^+}\Big\}, \quad \text{for} \quad N - 1 \geq p \geq 1,$$

or

$$1 - \frac{r}{\sum_{k=1}^{p} \frac{l_k}{l}} \leq u_{j1}, \qquad \text{for} \quad j > p. \tag{A.39}$$

In order to characterize $d_o(r)$, we need to consider three cases. The first case is when $1 \geq \tilde{v}_1$. In this case, (A.38) simplifies to

$$1 - r \leq \sum_{j=1}^{N} \frac{l_j}{l} \tilde{v_j}.$$

Let us define $x_j \triangleq j(\tilde{v}_j - \tilde{v}_{j+1}), j = 1, \cdots, N-1$ and $x_N \triangleq N\tilde{v}_N$. It immediately follows that $x_j \geq 0, j = 1, \ldots, N$. It is also easy to verify that

$$\sum_{j=1}^{N} \tilde{v}_j = \sum_{j=1}^{N} x_j \quad \text{and} \quad 1 - r \leq \sum_{j=1}^{N} \frac{f_j}{j} x_j, \tag{A.40}$$

where $f_j \triangleq \sum_{k=1}^{j} l_k/l$. From (A.40), it can be seen that

$$\inf_{\substack{O^+ \\ 1 \geq \tilde{v}_1}} \sum_{j=1}^{N} \tilde{v}_j = p\Big(\frac{1-r}{f_p}\Big), \quad \text{where} \quad p = \arg \max_{N \geq j \geq 1} \Big\{\frac{f_j}{j}\Big\}. \tag{A.41}$$

The infimum value corresponds to $x_p = p(1-r)/f_p$ and $x_j = 0, j \neq p$ or $\tilde{v}_j = (1-r)/f_p, p \geq j \geq 1$ and $\tilde{v}_j = 0, j > p$. But we assumed $1 \geq \tilde{v}_1$, so

$$f_p \geq 1 - r. \tag{A.42}$$

From (A.39), it follows that,

$$\inf_{\substack{O^+ \\ 1 \geq \tilde{v}_1}} \sum_{j>p} u_{j1} = (N-p)(1 - \frac{r}{f_p}), \quad 1 \geq f_p \geq r. \tag{A.43}$$

Now, from (A.41) and (A.43) we conclude that

$$\inf_{\substack{O^+, \\ 1 \geq \tilde{v}_1}} \sum_{j=1}^{N} \left(\tilde{v}_j + \sum_{i<j} u_{ji} \right) \geq \inf_{\substack{N \geq p \geq 1, \\ 1 \geq f_p \geq max\{r, 1-r\}}} d_o(r, p, f_p),$$

where,

$$d_o(r, p, f_p) \triangleq p(\frac{1-r}{f_p}) + (N-p)(1 - \frac{r}{f_p}). \tag{A.44}$$

It turns out that, (A.44) is an increasing function of p. Therefore, its infimum corresponds to $p = 1$. Now, examining $d_o(r, 1, f_1)$, i.e.,

$$d_o(r, 1, f_1) = (\frac{1-r}{f_1}) + (N-1)(1 - \frac{r}{f_1}),$$

we realize that, for $1/N \geq r \geq 0$, it decreases with f_1, thus its infimum corresponds to $f_1 = 1$. On the other hand, for $1 \geq r \geq 1/N$, $d_o(r, 1, f_1)$ becomes an increasing function of f_1, which means that its infimum corresponds to $f_1 = \max\{r, 1-r\}$, i.e.

$$\inf_{\substack{O^+, \\ 1 \geq \tilde{v}_1}} \sum_{j=1}^{N} \left(\tilde{v}_j + \sum_{i<j} u_{ji} \right) \geq \begin{cases} N(1-r), & \frac{1}{N} \geq r \geq 0, \\ 1 + \frac{(N-1)(1-2r)}{1-r}, & \frac{1}{2} \geq r \geq \frac{1}{N}, \\ \frac{1-r}{r}, & 1 \geq r \geq \frac{1}{2}. \end{cases} \tag{A.45}$$

The second case to be considered is when $\tilde{v}_i > 1 \geq \tilde{v}_{i+1}$, $N-1 \geq i \geq 1$. It immediately follows that

$$\inf_{\substack{O^+ \\ \tilde{v}_i > 1 \geq \tilde{v}_{i+1}}} \sum_{j=1}^{i} \tilde{v}_j = i. \tag{A.46}$$

In this case, (A.38) can be written as

$$1 - r - f_i \leq \sum_{j=i+1}^{N} \frac{l_j}{l} \tilde{v}_j. \tag{A.47}$$

If $f_i \geq 1 - r$, then from (A.47), we get

$$\inf_{\substack{O^+ \\ \tilde{v}_i > 1 \geq \tilde{v}_{i+1}, \\ f_i \geq \max\{r, 1-r\}}} \sum_{j=i+1}^{N} \tilde{v}_j = 0. \tag{A.48}$$

On the other hand, from (A.39), it follows that,

$$\inf_{u_{j1} \geq 1 - \frac{r}{f_i}, j > i} \sum_{j=i+1}^{N} u_{j1} = (N-i)(1 - \frac{r}{f_i}), \quad 1 \geq f_i \geq r. \tag{A.49}$$

Now, from (A.46), (A.48) and (A.49) one can see that

$$\inf_{\substack{O^+, \\ \tilde{v}_i > 1 \geq \tilde{v}_{i+1}, \\ f_i \geq \max\{r, 1-r\}}} \sum_{j=1}^{N} \left(\tilde{v}_j + \sum_{i<j} u_{ji} \right) \geq \inf_{\substack{N-1 \geq i \geq 1, \\ 1 \geq f_i \geq max\{r, 1-r\}}} d_o(r, i, f_i),$$

with

$$d_o(r, i, f_i) \triangleq i + (N-i)(1 - \frac{r}{f_i}).$$

The infimum of $d_o(r, i, f_i)$ corresponds to $i = 1$ and $f_i = \max\{r, 1-r\}$, i.e.,

$$\inf_{\substack{O^+, \\ \tilde{v}_i > 1 \geq \tilde{v}_{i+1}, \\ f_i \geq \max\{r, 1-r\}}} \sum_{j=1}^{N} \left(\tilde{v}_j + \sum_{i<j} u_{ji} \right) \geq \begin{cases} 1 + \frac{(N-1)(1-2r)}{1-r}, & \frac{1}{2} \geq r \geq 0, \\ 1, & 1 \geq r \geq \frac{1}{2}. \end{cases} \tag{A.50}$$

If $f_i < 1 - r$, then the problem of finding $\inf \sum_{j=i+1}^{N} \tilde{v}_j$ reduces to the first case (i.e., $1 \geq \tilde{v}_1$). Specifically, $\inf \sum_{j=i+1}^{N} \tilde{v}_j$ is given by (A.41), with $N - i$, $f_p - f_i$, $r + f_i$ and $p - i$ substituting N, f_p, r and p. Thus,

$$\inf_{\substack{O^+, \\ \tilde{v}_i > 1 \geq \tilde{v}_{i+1}, \\ 1-r > f_i \geq r}} \Big(\sum_{j=i+1}^{N} \tilde{v}_j \Big) = (p-i) \Big(\frac{1 - r - f_i}{f_p - f_i} \Big), \qquad \text{where } \; p = \arg \max_{N \geq j \geq i+1} \Big\{ \frac{f_j - f_i}{j - i} \Big\}. \tag{A.51}$$

Note that (A.42) still holds. Derivation of $\inf \sum_{j=1}^{N} \sum_{i<j} u_{ji}$ follows from (A.39),

$$\inf_{u_{j1} \geq 1-\frac{r}{f_k}, j>k} \sum_{j=1}^{N} \sum_{i<j} u_{ji} \geq (p-i)(1-\frac{r}{f_i}) + (N-p)(1-\frac{r}{f_p}), \quad \text{with} \quad f_p > f_i \geq r. \tag{A.52}$$

From (A.46), (A.51) and (A.52), we conclude that

$$\inf_{\substack{O^+, \\ \tilde{v}_i > 1 \geq \tilde{v}_{i+1}, \\ 1-r > f_i \geq r}} \sum_{j=1}^{N} \left(\tilde{v}_j + \sum_{i<j} u_{ji} \right) \geq \inf_{\substack{N \geq p > i \geq 1, \\ 1 \geq f_p \geq 1-r > f_i \geq r}} d_o(r, i, p, f_i, f_p), \tag{A.53}$$

where,

$$d_o(r, i, p, f_i, f_p) \triangleq i + (p-i)(\frac{1-r-f_i}{f_p - f_i}) + (p-i)(1-\frac{r}{f_i}) + (N-p)(1-\frac{r}{f_p}). \tag{A.54}$$

As can be seen from (A.54), $d_o(r, i, p, f_i, f_p)$ is a linear, and therefore monotonic, function of p. Thus, its infimum corresponds to either $p = i+1$ or $p = N$. Now if the infimum indeed corresponds to $p = i+1$, by plugging in $p = i$ into (A.54), we derive a lower-bound on it. That is,

$$\inf_{\substack{N \geq p > i \geq 1, \\ 1 \geq f_p \geq 1-r > f_i \geq r}} d_o(r, i, p, f_i, f_p) \geq \inf_{\substack{N > i \geq 1 \\ 1 \geq f_p \geq 1-r}} i + (N-i)(1-\frac{r}{f_p}).$$

or

$$\inf_{\substack{N \geq p > i \geq 1, \\ 1 \geq f_p \geq 1-r > f_i \geq r}} d_o(r, i, p, f_i, f_p) \geq 1 + (N-1)\frac{1-2r}{1-r}, \quad \text{for} \quad \frac{1}{2} > r \geq 0. \tag{A.55}$$

Choosing $p = N$, on the other hand, gives

$$d_o(r, i, N, f_i, 1) = i + (N-i)(2 - \frac{r}{1-f_i} - \frac{r}{f_i}),$$

which has an infimum value, corresponding to $i = 1$ and $f_i = r$ or $f_i = 1-r$, identical to the right-hand side of (A.55). This means that

$$\inf_{\substack{N \geq p > i \geq 1, \\ 1 \geq f_p \geq 1-r > f_i \geq r}} d_o(r, i, p, f_i, f_p) = 1 + (N-1)\frac{1-2r}{1-r}, \quad \text{for} \quad \frac{1}{2} > r \geq 0. \tag{A.56}$$

Now, from (A.56), (A.53) and (A.50), we conclude that

$$\inf_{\substack{O^+,\\ \tilde{v}_i>1\geq\tilde{v}_{i+1}}} \sum_{j=1}^{N}\left(\tilde{v}_j+\sum_{i<j}u_{ji}\right)\geq\begin{cases}1+\frac{(N-1)(1-2r)}{1-r}, & \frac{1}{2}\geq r\geq 0,\\ 1, & 1\geq r\geq\frac{1}{2}.\end{cases} \tag{A.57}$$

The third case (i.e., $\tilde{v}_N > 1$), is trivial

$$\inf_{\substack{O^+,\\ \tilde{v}_N>1}} \sum_{j=1}^{N}\left(\tilde{v}_j+\sum_{i<j}u_{ji}\right)\geq N. \tag{A.58}$$

From (A.45), (A.57) and (A.58) we conclude that (3.8) provides a lower-bound on the diversity gain achieved by the protocol. On the other hand, $d_o(r)$ is also an upper bound on the diversity since: 1) for $1/N \geq r \geq 0$, $d_o(r)$ is the genie-aided diversity, 2) for $0.5 \geq r \geq 1/N$, it can be shown that the realization, where $v_1 = 1+\epsilon$, $\{v_j\}_{j=2}^N = 0$, $\{u_{j1}\}_{j=2}^N = \frac{1-2r}{1-r}$ and $\{u_{ji}\}_{i\neq j} = 0$ corresponds to a *channel* outage for any $\epsilon > 0$, and 3) for $1 \geq r \geq 0.5$, realization $v_1 = \frac{1-r}{r}+\epsilon$, $\{v_j\}_{j=2}^N = 0$ and $\{u_{ji}\} = 0$ also corresponds to a channel outage for any $\epsilon > 0$. Thus (3.8) is the diversity achieved by the $N-1$ relay DDF protocol and the proof is complete.

A.6 Proof of Lemma 7

To prove the first part of the lemma, let us denote the signals received at the destination by $\{y_k\}_{k=1}^l$. Then

$$y_k=\begin{cases} g_1x_k+v_k & \text{for} \quad k=1,\cdots,l'\\ g_1x_k+g_2x_{k+1}^*+v_k & \text{for} \quad k=l'+1,l'+3,\cdots\\ g_1x_k-g_2x_{k-1}^*+v_k & \text{for} \quad k=l'+2,l'+4,\cdots\end{cases}$$

Now, through linear processing of $\{y_k\}_{k=1}^l$, the destination derives $\{\tilde{y_k}\}_{k=1}^l$ such that

$$\tilde{y}_k=\begin{cases} g_1x_k+\tilde{v}_k & \text{for} \quad k=1,\cdots,l'\\ \sqrt{|g_1|^2+|g_2|^2}x_k+\tilde{v}_k & \text{for} \quad k=l'+1,\cdots,l\end{cases}, \tag{A.59}$$

with $\tilde{v_k}$ being statistically identical to v_k. Using (A.59), it is straightforward to see that destination pairwise error probability, averaged over the ensemble of Gaussian

codes (used by the source) and conditioned on a certain channel realization, is still given by (A.24). This means that modifying the DDF protocol according to (3.9) does not effect the DMT achieved by the protocol.

To prove the second part, we notice that the effect of restricting the relay to start transmission only after the codeword is halfway through, i.e., adopting (3.10) in place of (3.6), is to replace (A.29) with

$$\inf_{O^+,f} u = \begin{cases} 0 & \text{if} \quad f = \frac{1}{2} \\ (1-\frac{r}{f})^+ & \text{if} \quad 1 \geq f > \frac{1}{2} \end{cases}, \tag{A.60}$$

where u denotes the exponential order of $1/|h|^2$. Now, since (A.60) is different from (A.29) only for $f = 1/2$, for which $\inf(v_1 + v_2)$ is already equal to the optimal value $2(1-r)$, we conclude that this restriction does not affect the DMT achieved by the protocol.

A.7 Proof of Lemma 8

We notice that the outage set for the DDF protocol with a general set of waiting fractions, $\{f_j\}_{j=1}^N$, is still given by (A.26), with the only difference that f is now given by (compare with (A.29))

$$f = \begin{cases} f_1 & \text{if} \quad 1-\frac{r}{f_1} > u \geq 0 \\ f_j & \text{if} \quad 1-\frac{r}{f_j} > u \geq (1-\frac{r}{f_{j-1}})^+ \\ 1 & \text{if} \quad u \geq (1-\frac{r}{f_N})^+ \end{cases}. \tag{A.61}$$

Next, we split O^+ such that

$$O^+ = \cup_{j=1}^{N+1} O_j^+, \quad \text{where} \quad O_j^+ = \{(v_1, v_2, u) \in O^+ | f = f_j\}. \tag{A.62}$$

Now, from (A.61) and (A.62) we get

$$\inf_{O_j^+} u = \begin{cases} 0 & \text{for} \quad j = 1 \\ (1-\frac{r}{f_{j-1}})^+ & \text{for} \quad N+1 \geq j > 2 \end{cases}.$$

Also, since $f_j \geq \frac{1}{2}$, we have $\inf_{(v_1,v_2)\in O_j^+}(v_1+v_2) = (1-r)/f_j$ (recall (A.32)). Thus

$$\begin{aligned} d_j(r) &\triangleq \inf_{O_j^+}(v_1+v_2+u), \\ d_j(r) &= \frac{1-r}{f_j} + (1-\frac{r}{f_{j-1}})^+. \end{aligned} \tag{A.63}$$

But, (A.62) along with (A.63) results in

$$\begin{aligned} d(r) &= \min_{f_j \geq r} d_j(r), \\ d(r) &= \min_{f_j \geq r} \frac{1-r}{f_j} + (1-\frac{r}{f_{j-1}})^+. \end{aligned} \tag{A.64}$$

Now let us assume that the set of waiting fractions $\{f_j^u\}_{j=1}^N$ is uniformly optimal. Pick $\{f_j\}_{j=1}^N$ such that $f_N^u < f_N < 1$. Then from (A.64) we conclude that for any $f_N^u < r < f_N$, $d^u(r) = 1-r$ and $d(r) = 1-r+1-r/f_N$. Thus $d^u(r) < d(r)$, which is in contradiction with the uniform optimality assumption of $\{f_j^u\}_{j=1}^N$. To prove the second part of the lemma, we observe that for $N \geq j \geq 1$, $\{f_j^p\}_{j=1}^N$ as given by (3.11) results in

$$d_{N+1}^p(r) < d_j^p(r), \quad \text{for} \quad f_j \geq r \geq 0, r \neq f_{j-1} \tag{A.65}$$

and

$$\begin{aligned} d_{N+1}^p(f_{j-1}) &= d_j^p(f_{j-1}), \quad \text{or} \\ d_{N+1}^p(f_{j-1}) &= \frac{1-f_{j-1}}{f_j}. \end{aligned} \tag{A.66}$$

Now (A.65), along with (A.64) and (A.63) proves that $\{f_j^p\}_{j=1}^N$ achieves (3.12). The only thing left is to show that $d^p(r)$ is indeed Pareto optimal, i.e., no other set $\{f_j\}_{j=1}^N$ dominates $\{f_j^p\}_{j=1}^N$, in the Pareto sense. To do so, we assume such a set exists and observe that since $f_0 = f_0^p = 0$ and $f_{N+1} = f_{N+1}^p = 1$, there should be $N+1 \geq i \geq 1$

and $N+1 \geq \ell \geq 1$ such that

$$f_{i-1} \leq f_{\ell-1}^p < f_\ell^p < f_i, \quad \text{or} \tag{A.67}$$

$$f_{i-1} < f_{\ell-1}^p < f_\ell^p \leq f_i. \tag{A.68}$$

Now, if (A.67) is true, then we observe from (A.64) that

$$d(f_{\ell-1}^p) \leq d_i(f_{\ell-1}^p) = \frac{1 - f_{\ell-1}^p}{f_i} < \frac{1 - f_{\ell-1}^p}{f_\ell^p} = d^p(f_{\ell-1}^p),$$

where we have used (A.66) in deriving the last step. This, however is in contradiction with the Pareto dominance of $\{f_j\}_{j=1}^N$, since

$$d(f_{\ell-1}^p) < d^p(f_{\ell-1}^p).$$

On the other hand, if (A.68) is true, then

$$d(f_{i-1}) \leq d_i(f_{i-1}) = \frac{1 - f_{i-1}}{f_i} < 2 - (1 + \frac{1}{f_N^p}) f_{i-1} = d^p(f_{i-1}),$$

or

$$d(f_{i-1}) < d^p(f_{i-1}),$$

which again is in contradiction with Pareto dominance of $\{f_j\}_{j=1}^N$. This completes the proof of the second part.

A.8 Proof of Theorem 9

To characterize the DMT achieved by the CB-DDF protocol, we first label the N destinations according to the order in which they start transmission. That is, the first destination that starts transmission is denoted as destination 1, the next destination

as destination 2, and so on. Note that the error probability of destination j can be written as

$$P_{E_j} = P_{E_j|S_j^c} P_{S_j^c} + P_{E_j|S_j} P_{S_j}, \tag{A.69}$$

where S_j denotes the event that destination j decodes the message and starts re-transmission before the end of the codeword and S_j^c is its complement. Now, since both P_{S_j} and $P_{S_j^c}$ are less than one, (A.69) gives

$$P_{E_j} \leq P_{E_j|S_j^c} + P_{E_j|S_j}. \tag{A.70}$$

In order to characterize $P_{E_j|S_j}$, we need to characterize $P_{E_j|S_j,g,h}$, i.e., destination j's ML error probability, averaged over the ensemble of Gaussian code-books and conditioned on a certain channel realization, under the assumption that it started transmission before the end of the codeword. Towards this end and through using Bayes' rule, one can upper bound $P_{E_j|S_j,g,h}$ to get

$$P_{E_j|S_j,g,h} \leq \sum_{i=1}^{j} P_{E_i|\{E_p^c\}_{p<i},S_j,g,h}. \tag{A.71}$$

Now, let us examine $P_{E_i|\{E_p^c\}_{p<i},S_j,g,h}$, i.e., the probability that destination i ($i \leq j$), makes an error in decoding the source message, conditioned on S_j (which ensures that destination i has indeed started re-transmission) and assuming *error-free* decoding at all of the active destinations. It follows from Lemma 1, that if the mutual information between the signals transmitted by the source and active destinations and the signal received by destination i exceeds lR (which is implied by S_j), then $P_{E_i|\{E_p^c\}_{p<i},S_j,g,h}$ can be made arbitrarily small, provided that the code-length is sufficiently large. This means that for any $\epsilon > 0$ and for sufficiently large code-lengths,

$$P_{E_i|\{E_p^c\}_{p<i},S_j,g,h} < \epsilon, \quad i \leq j. \tag{A.72}$$

Using (A.72), (A.71) can be written as

$$P_{E_j|S_j,g,h} \leq j\epsilon.$$

Taking the average over the ensemble of channel realizations gives

$$P_{E_j|S_j} < j\epsilon.$$

This together with (A.70), yields

$$\begin{aligned} P_{E_j} &< P_{E_j|S_j^c} + j\epsilon, \\ P_{E_j} &\dot{\leq} P_{E_j|S_j^c}. \end{aligned} \tag{A.73}$$

This means that the exponential order of $P_{E_j|S_j^c}$, provides a lower-bound on the diversity gain achieved by the protocol. Now, examining $P_{E_j|S_j^c}$, it is easy to realize that the event in which the j^{th} destination (out of N destinations), spends the entire codeword listening, i.e. S_j^c, is identical to the DDF relay protocol with the rest of the destinations taking the role of the $N-1$ relays. Thus, from (A.73), we see that communication to the j^{th} destination achieves the same diversity order as does the DDF relay protocol with $N-1$ relays, namely, (4.1). This completes the proof.

A.9 Proof of Theorem 10

A simple max-cut min-flow examination reveals that the optimal diversity gain for this channel is upper bounded by

$$d_{\text{MAR}}(r) \leq \min\{d_{3\times1}(r), d_{2\times2}(r), d_{2\times1}(\frac{r}{2}), d_{1\times2}(\frac{r}{2})\}, \tag{A.74}$$

where $d_{m\times n}(\cdot)$ denotes the optimal diversity gain for an $m \times n$ MIMO channel. Now, (A.74) results in (4.2) and the proof of the converse part is complete.

In order to derive a lower bound on the diversity gain achieved by the DDF MAR protocol, we upper bound the *source-specific* ML error probabilities, with that of the *joint* ML decoder. Furthermore, instead of characterizing the latter probability for specific codes, in the sequel, we characterize its average, P_E, over the ensemble of random Gaussian codes. It is then straightforward to see that there exists a code in the ensemble, whose error probability is better than P_E. To characterize P_E, we use the Bayes' rule to derive the following upper bound

$$P_E \leq P_{E|\overline{E}_r} + P_{E_r},$$

where E_r and $\overline{E}_r$ denote the events that the relay makes errors in decoding the messages, and its complement, respectively. Next, we note that if we denote the signals transmitted by the two sources and the relay by $\{x_{j,k}\}_{k=1}^{l}$ and $\{x_{r,k}\}_{k=l'+1}^{l}$, respectively, and the signals received by the relay and the destination by $\{y_{r,k}\}_{k=1}^{l'}$ and $\{y_k\}_{k=1}^{l}$, then the number of symbol intervals l' that the relay waits before decoding the messages satisfies

$$\frac{lR}{2} \leq I(\{x_{1,k}\}_{k=1}^{l'}; \{y_{r,k}\}_{k=1}^{l'} | \{x_{2,k}\}_{k=1}^{l'}), \tag{A.75}$$

$$\frac{lR}{2} \leq I(\{x_{2,k}\}_{k=1}^{l'}; \{y_{r,k}\}_{k=1}^{l'} | \{x_{1,k}\}_{k=1}^{l'}), \tag{A.76}$$

$$lR \leq I(\{x_{1,k}\}_{k=1}^{l'}, \{x_{2,k}\}_{k=1}^{l'}; \{y_{r,k}\}_{k=1}^{l'}). \tag{A.77}$$

In these expressions, R is the *total* data rate (in BPCU) at the destination and $I(.;.)$ denotes the mutual information function. Now, observing that (A.75), (A.76) and (A.77) guarantee that (e.g., refer to section 14.3.1 of [25])

$$P_{E_r} \leq \epsilon, \quad \text{for any} \quad \epsilon > 0,$$

we conclude

$$P_E \dot{\leq} P_{E|\overline{E}_r}.$$

For the sake of notational simplicity, in the sequel, $P_{E|\overline{E}_r}$ is denoted by P_E. In characterizing P_E, we follow the approach of Tse, Viswanath and Zheng [18], by partitioning the error event E into the set of partial error events E_I, i.e.,

$$E = \bigcup_I E_I.$$

where I denotes any *nonempty* subset of $\{1, 2\}$ and E_I (referred to as type-I error) is the event that the joint ML decoder incorrectly decodes the messages from sources whose indices belong to I while correctly decoding all other messages. Because the partial error events are mutually exclusive, we have

$$P_E = \sum_I P_{E_I}. \tag{A.78}$$

To characterize P_{E_I}, we use the Bayes' rule to derive the following upper-bound

$$P_{E_I} = P_{O_I} P_{E_I|O_I} + P_{E_I,\overline{O}_I}$$

$$P_{E_I} \leq P_{O_I} + P_{E_I,\overline{O}_I},$$

where O_I and $\overline{O}_I$ denote the type-I outage event and its complement, respectively. The type-I outage event is defined such that P_{O_I} dominates $P_{E_I,\overline{O}_I}$, i.e.

$$P_{E_I,\overline{O}_I} \dot{\leq} P_{O_I}. \tag{A.79}$$

Thus,

$$P_{E_I} \dot{\leq} P_{O_I}. \tag{A.80}$$

Characterization of O_I, however, requires derivation of $P_{PE_I|g,h}$, i.e., the joint ML decoder's type-I pairwise error probability (PEP), conditioned on a particular channel realization and averaged over the ensemble of random Gaussian codes. For this purpose, let us denote the gain of the channels connecting the two sources to the relay by h_1 and h_2 and those of the channels connecting the two sources and the relay to the destination by g_1, g_2 and g_r, respectively. It is then straightforward to see that (recall (2.7))

$$P_{PE_{\{1\}}|g,h} \leq (1+\frac{1}{2}\rho|g_1|^2)^{-l'}(1+\frac{1}{2}\rho(|g_1|^2+|g_r|^2))^{-(l-l')},$$
$$P_{PE_{\{1,2\}}|g,h} \leq (1+\frac{1}{2}\rho(|g_1|^2+|g_2|^2))^{-l'}(1+\frac{1}{2}\rho(|g_1|^2+|g_2|^2+|g_r|^2))^{-(l-l')}.$$

Notice that since the channel is symmetric, $P_{E_{\{1\}}} \doteq P_{E_{\{2\}}}$, i.e., we do not need to characterize $P_{PE_{\{2\}}|g,h}$. Let us denote the exponential orders of $\{1/|g_j|^2\}_{j=1}^2$ and $1/|g_r|^2$ by $\{v_j\}_{j=1}^2$ and v_r, respectively, and those of $\{1/|h_j|^2\}_{j=1}^2$ by $\{u_j\}_{j=1}^2$. Realizing that, at a rate of $R = r\log\rho$ BPCU and a codeword length of l, there are a total of ρ^{lr} codewords in the code, we get

$$P_{E_{\{1\}}|v,u} \dot{\leq} \rho^{-l[f(1-v_1)^+ + (1-f)(1-\min\{v_1,v_r\})^+ - \frac{r}{2}]}, \tag{A.81}$$

$$P_{E_{\{1,2\}}|v,u} \dot{\leq} \rho^{-l[f(1-\min\{v_1,v_2\})^+ + (1-f)(1-\min\{v_1,v_2,v_r\})^+ - r]}, \tag{A.82}$$

where $f \triangleq l'/l$. Careful examination of (A.81) reveals that defining $O_{\{1\}}^+$ as

$$O_{\{1\}}^+ \triangleq \{(v_1,\cdots,u_2) \in \mathbb{R}^{5+} | f(1-v_1)^+ + (1-f)(1-\min\{v_1,v_r\})^+ \leq \frac{r}{2}\}, \tag{A.83}$$

satisfies (A.79). In this expression, $O_{\{1\}}^+$ denotes $O_{\{1\}} \cap \mathbb{R}^{5+}$, where $\mathbb{R}^{n+}$ represents the set of all nonnegative real n-tuples. This is because, if (A.83) is satisfied, then through choosing a large enough l, $P_{E_{\{1\}},\overline{O}_{\{1\}}}$ can be made arbitrarily small. Likewise,

defining $O^+_{\{1,2\}}$ as

$$O^+_{\{1,2\}} \triangleq$$
$$\{(v_1, \cdots, u_2) \in \mathbb{R}^{5+} | f(1-\min\{v_1, v_2\})^+ + (1-f)(1-\min\{v_1, v_2, v_r\})^+ \leq r\}, \tag{A.84}$$

satisfies (A.79). With the type-I outage events specified, the only thing left is to characterize P_{O_I}, i.e.

$$P_{O_I} \doteq \rho^{-d_I(r)} \text{ where } d_I(r) \triangleq \inf_{O_I^+} \{v_1 + v_2 + v_r + u_1 + u_2\}. \tag{A.85}$$

Toward this end, we use (A.83) to derive $\inf_{(v_1,v_2,v_r)\in O^+_{\{1\}}} \{v_1 + v_2 + v_r\}$, as a *function* of f, i.e.

$$\inf_{(v_1,v_2,v_r)\in O^+_{\{1\}}} \{v_1 + v_2 + v_r\}(f) = \lambda_{\{1\}}(f),$$

where

$$\lambda_{\{1\}}(f) \triangleq \begin{cases} 2-r, & \frac{1}{2} > f \geq 0 \\ 2-\frac{r}{2(1-f)}, & 1-\frac{r}{2} > f \geq \frac{1}{2} \\ \frac{2-r}{2f}, & 1 \geq f \geq 1-\frac{r}{2} \end{cases} \tag{A.86}$$

Likewise, one can use (A.84) to derive

$$\inf_{(v_1,v_2,v_r)\in O^+_{\{1,2\}}} \{v_1 + v_2 + v_r\}(f) = \lambda_{\{1,2\}}(f),$$

where

$$\lambda_{\{1,2\}}(f) \triangleq \begin{cases} 3(1-r), & \frac{2}{3} > f \geq 0 \\ 3-\frac{r}{1-f}, & 1-r > f \geq \frac{2}{3} \\ 2\frac{1-r}{f}, & 1 \geq f \geq 1-r \end{cases}, \text{ if } \frac{1}{3} > r \geq 0 \tag{A.87}$$

or

$$\lambda_{\{1,2\}}(f) \triangleq \begin{cases} 3(1-r), & \frac{2}{3} > f \geq 0 \\ 2\frac{1-r}{f}, & 1 \geq f \geq \frac{2}{3} \end{cases}, \text{ if } 1 \geq r \geq \frac{1}{3} \tag{A.88}$$

Now, to complete the derivation of $d_I(r)$, we need to characterize $\inf_{(u_1,u_2)\in\mathbb{R}^{2+}}\{u_1+u_2\}(f)$. For this purpose, we use (A.75), (A.76) and (A.77) to derive

$$l' = \min\{l, \max\{\lceil \frac{lR}{2\log_2(1+\min\{|h_1|^2,|h_2|^2\}c\rho)}\rceil, \lceil \frac{lR}{\log_2(1+(|h_1|^2+|h_2|^2)c\rho)}\rceil\}\},$$

where c is the ratio of destination noise variance to that of the relay and $\lceil x\rceil$ denotes the closest integer to x towards plus infinity. In terms of channel exponential orders, this last expression can be rewritten as

$$f = \min\{1, \max\{\frac{r}{2(1-\max\{u_1,u_2\})^+}, \frac{r}{(1-\min\{u_1,u_2\})^+}\}\}, \quad (u_1,u_2)\in\mathbb{R}^{2+}. \tag{A.89}$$

Now, using (A.89), one can show that

$$\inf_{(u_1,u_2)\in\mathbb{R}^{2+}}\{u_1+u_2\}(f) = \lambda(f),$$

where

$$\lambda(f) \triangleq \begin{cases} 2(1-\frac{r}{f}), & \frac{3r}{2} > f \geq r \\ 1-\frac{r}{2f}, & 1 \geq f \geq \frac{3r}{2} \end{cases}. \tag{A.90}$$

Using (A.85) we conclude

$$d_{\{1\}}(r) = \inf_f\{\lambda(f)+\lambda_{\{1\}}(f)\},$$

which can be characterized using (A.86) and (A.90),

$$d_{\{1\}}(r) = \begin{cases} 2-r, & \frac{1}{2} > r \geq 0 \\ \frac{4-5r}{2(1-r)}, & \frac{2}{3} \geq r \geq \frac{1}{2} \\ \frac{2-r}{2r}, & 1 \geq r \geq \frac{2}{3} \end{cases}. \tag{A.91}$$

Similarly,

$$d_{\{1,2\}}(r) = \inf_f\{\lambda(f)+\lambda_{\{1,2\}}(f)\},$$

which can be derived using (A.87), (A.88) and (A.90),

$$d_{\{1,2\}}(r) = \begin{cases} 3(1-r), & \frac{2}{3} > r \geq 0 \\ 2\frac{1-r}{r}, & 1 \geq r \geq \frac{2}{3} \end{cases} . \tag{A.92}$$

Now, (A.91) and (A.92), together with (A.85), (A.80) and (A.78), result in (4.3) and thus complete the proof of the achievability part.

A.10 Proof of Theorem 11

Realizing that (4.7) also corresponds to the optimal DMT for a MIMO point-to-point communication system with N transmit antennas and a single receive antenna (i.e., the case of "genie-aided" cooperation between N sources), we only need to show that the CMA-NAF protocol achieves this tradeoff. To achieve this goal, we assume that each of the sources uses a Gaussian random code with codeword length l and data rate R, where l is chosen as in (4.5) and R grows with ρ according to (4.6). We then characterize the joint ML decoder's error probability $P_E(\rho)$. Note that the error probability of the joint ML decoder upper-bounds the error probabilities of the source-specific ML decoders and thus provides a lower-bound on the achievable overall diversity gain (as a function of r). In characterizing $P_E(\rho)$, we follow the approach of Tse *et al.* [18] by partitioning the error event E into the set of partial error events $\{E^I\}$, i.e.,

$$E = \bigcup_I E^I,$$

where I denotes any *nonempty* subset of $\{1, ..., N\}$ and E^I (referred to as a "type-I error") is the event that the joint ML decoder incorrectly decodes the messages from sources whose indices belong to I while correctly decoding all other messages.

Because the partial error events are mutually exclusive,

$$P_E(\rho) = \sum_I P_{E^I}(\rho). \tag{A.93}$$

Using Bayes' rule, one can upper-bound $P_{E^I}(\rho)$ as

$$P_{E^I}(\rho) = P_O(R)P_{E^I|O} + P_{E^I,O^c}$$

$$P_{E^I}(\rho) \leq P_O(R) + P_{E^I,O^c},$$

where, as before, O and O^c denote the outage event and its complement, respectively. The outage event is defined such that $P_O(R)$ dominates P_{E^I,O^c} for all I:

$$P_{E^I,O^c} \dot{\leq} P_O(R). \tag{A.94}$$

Thus,

$$P_{E^I}(\rho) \dot{\leq} P_O(R),$$

which, together with (A.93), results in

$$P_E(\rho) \dot{\leq} P_O(R). \tag{A.95}$$

This means that $P_O(R)$, as defined by (A.94), provides an upper-bound to the joint ML decoder's error probability and therefore a lower-bound to the achievable diversity gain $d^*(r)$. The derivation of $P_O(R)$, however, requires the characterization of $P_{PE^I|g_j,h_{ji}}$ (i.e., the joint ML decoder's type-I PEP, conditioned on a particular channel realization and averaged over the ensemble of Gaussian random codes). Here, we upper-bound $P_{PE^I|g_j,h_{ji}}$, for each I, by the PEP of a suboptimal joint ML decoder that uses only a subset of the destination's observations (referred to as the *type-I decoder*):

$$P_{PE^I|g_j,h_{ji}} \leq \det(I_m + \frac{1}{2}\Sigma_{\mathbf{s}^I}\Sigma_{\mathbf{n}^I}^{-1})^{-1} \tag{A.96}$$

In (A.96), $\Sigma_{\mathbf{s}^I}$ and $\Sigma_{\mathbf{n}^I}$ represent the $m \times m$ covariance matrices corresponding to the signal and noise components, respectively, of the *partial* observation vector used by the type-I decoder, provided that the symbols of the sources that are not in set I are set to zero. The size m will be characterized in the sequel.

Before going into more detail on the type-I decoder, we note that, since $\Sigma_{\mathbf{s}^I}$ and $\Sigma_{\mathbf{n}^I}$ are both positive definite matrices, the right-hand side of (A.96) can be upper-bounded as

$$P_{PE^I|g_j,h_{ji}} \dot{\leq} \det(\Sigma_{\mathbf{s}^I})^{-1} \det(\Sigma_{\mathbf{n}^I}). \tag{A.97}$$

The discussion is simplified if we define v_j and u_{ji} as the exponential orders of $1/|g_j|^2$ and $1/|h_{ji}|^2$, respectively. Note that the exponential orders of $\{|b_j|^2\}_{j=1}^N$ do not appear in the following expressions for the reasons outlined in the proof of Theorem 3. We also note that the exponential orders of the broadcast gains $\{|a_j|^2\}_{j=1}^N$ are zero. Furthermore, recalling (2.4), the PDFs of negative v_j and u_{ji} are effectively zero for large values of ρ, allowing us to concern ourselves only with their non-negative realizations. With this ideas in mind, we return to (A.97) and claim that

$$\det(\Sigma_{\mathbf{n}^I}) \dot{\leq} 1. \tag{A.98}$$

To understand (A.98), recall that the noise component of the destination observation is a linear combination of the noise originating at the sources (i.e., $\{w_{j,k}\}_{j=1}^N$) and the noise originating at the destination (i.e., $v_{j,k}$). Furthermore, the coefficients of this linear combination are the products of some channel, broadcast, and repetition gains. Then, because these noise variances and magnitude-squared gains can be written as non-positive powers of ρ, equation (A.98) must hold. Combining (A.98) and (A.97)

yields

$$P_{PE^I|v_j,u_{ji}} \dot{\leq} \det(\Sigma_{\mathbf{s}^I})^{-1} \quad \text{for} \quad v_j \geq 0, u_{ji} \geq 0. \tag{A.99}$$

As mentioned earlier, $\Sigma_{\mathbf{s}^I}$ represents the covariance matrix of the signal component of the partial observation used by the type-I decoder, provided that the symbols of the sources that are not in I are set to zero. To fully characterize $\Sigma_{\mathbf{s}^I}$, though, we must know which observations are used by the type-I decoder and which are discarded. The type-I decoder picks *one* observation for every source in set I, for a total of $m = |I|$ observations per frame (where $|I|$ denotes the size of I and therefore $1 \leq |I| \leq N$). Provided that frame k is not the last frame in its super-frame and assuming that during this super-frame, source i is helping source $j \in I$, the destination observation component corresponding to source j will be either the $y_{j,k}$ that corresponds to source j's broadcast of $x_{j,k}$ or the $y_{i,k'}$ that corresponds to helper i's re-broadcast of $x_{j,k}$ (where $k' \in \{k, k+1\}$). As an example, consider the case when $N = 4$ and assume that during a certain super-frame, source 3 is helping source $2 \in I$ (i.e., $j = 2$, $i = 3$). In this case, the type-I decoder picks either $y_{2,k}$ or $y_{3,k}$ in correspondence to $x_{2,k}$. However, if instead of source 3, source 1 is helping source 2 (i.e., $j = 2$, $i = 1$), then the type-I decoder has to choose between $y_{2,k}$ or $y_{1,k+1}$. Back to our description of the type-I decoder, if $i \in I$, then the decoder always picks $y_{j,k}$ over $y_{i,k'}$. On the other hand, if $i \notin I$, then the decoder chooses $y_{j,k}$ when $|g_j|^2 \geq |g_i|^2$ or $y_{i,k'}$ when $|g_j|^2 < |g_i|^2$ (i.e., the observation received through the better channel). The preceding discussion focused on the case where frame k is not the last frame of the super-frame. If frame k is indeed last, then the decoder always chooses $y_{j,k}$ over $y_{i,k'}$.

We define $\mathbf{s}_{j,k}^I$, where $j \in I$, as the vector (of dimension $ml \times 1$) of contributions of symbol $x_{j,k}$ to the destination observations picked by the type-I decoder. Clearly,

$$\mathbf{s}^I = \sum_{k=1}^{l} \sum_{j \in I} \mathbf{s}_{j,k}^I.$$

Taking into account the independence of the transmitted symbols (i.e., $x_{j,k}$), we have

$$\Sigma_{\mathbf{s}^I} = \sum_{k=1}^{l} \sum_{j \in I} \mathrm{E}\{\mathbf{s}_{j,k}^I (\mathbf{s}_{j,k}^I)^H\}. \tag{A.100}$$

In order to illuminate some of the properties of $\mathbf{s}_{j,k}^I$, assume that we sort the chosen observations in chronological order. From the description given, it is apparent that, associated with each chosen observation (i.e., $y_{j,k}$ or $y_{i,k'}$) there is *one* symbol $x_{j,k}$ (with $j \in I$) which has contributions only from this observation forward. This means that if we define S^I as

$$S^I \triangleq [\mathbf{s}_{j_1,k_1}^I \mathbf{s}_{j_2,k_2}^I \dots \mathbf{s}_{j_{ml},k_{ml}}^I]_{ml \times ml},$$

where $j_p \in I$ and $k_p \in \{1, \dots, l\}$ are chosen such that the first non-zero elements of $\mathbf{s}_{j_p,k_p}^I, p = 1, \dots, ml$ are sorted in chronological order, then S^I will be lower-triangular and consequently $(S^I)^H$ will be upper-triangular. Furthermore, based on the choice between $y_{j,k}$ or $y_{i,k'}$ (corresponding to $x_{j,k}$), the first non-zero element of $\mathbf{s}_{j_p,k_p}^I$ (i.e., the p^{th} diagonal element of S^I) will be $g_j a_j x_{j,k}$ or $g_i b_i h_{ij} a_j x_{j,k}$, respectively. Next, we define $\psi_{j,k}^I$ as the signature of $x_{j,k}$, i.e.,

$$\psi_{j,k}^I \triangleq \frac{1}{x_{j,k}} \mathbf{s}_{j,k}^I \quad j \in I,$$

and Ψ^I as

$$\Psi^I \triangleq [\psi_{j_1,k_1}^I \psi_{j_2,k_2}^I \dots \psi_{j_{ml},k_{ml}}^I]_{ml \times ml}.$$

It follows then, that Ψ^I is also lower-triangular with the p^{th} diagonal element being equal to $g_j a_j$ or $g_i b_i h_{ij} a_j$. Using these definitions, (A.100) can be written as

$$\Sigma_{\mathbf{s}^I} = E \sum_{k=1}^{l} \sum_{j \in I} \psi_{j,k}^I (\psi_{j,k}^I)^H. \tag{A.101}$$

The significance of Ψ^I can now be seen from the fact that (A.101) can be written as

$$\Sigma_{\mathbf{s}^I} = E \Psi^I (\Psi^I)^H.$$

Now, as the determinant of triangular matrices is simply the product of their diagonal elements, from (A.99) we conclude that

$$P_{PE^I|v_j,u_{ji}} \dot{\leq} \rho^{-m(N-1)L+\sum_{j\in I}\left[(m-1)Lv_j+\sum_{i\notin I}\left(\min\{v_j,u_{ji}+v_i\}(L-1)+v_j\right)\right]}, \quad v_j \geq 0, u_{ji} \geq 0.$$

It is obvious that for large L's, the previous inequality can be rewritten as

$$P_{PE^I|v_j,u_{ji}} \dot{\leq} \rho^{-\left[-\sum_{j\in I}\left((m-1)v_j+\sum_{i\notin I}\min\{v_j,u_{ji}+v_i\}\right)+m(N-1)\right]L}, \quad v_j \geq 0, u_{ji} \geq 0. \tag{A.102}$$

At rate $R = r\log(\rho)$ and codeword length l, and when the symbols of the sources that are not in I are set to zero, there are a total of $\rho^{m(N-1)Lr}$ unique codewords. Thus,

$$P_{E^I|v_j,u_{ji}} \dot{\leq} \rho^{-\left[-\sum_{j\in I}\left((m-1)v_j+\sum_{i\notin I}\min\{v_j,u_{ji}+v_i\}\right)+m(N-1)(1-r)\right]L}, \quad v_j \geq 0, u_{ji} \geq 0. \tag{A.103}$$

This conditional type-I error probability leads to

$$P_{E^I,O^c} \dot{\leq} \rho^{-d_{e^I}(r)},$$

where

$$\begin{aligned} d_{e^I}(r) \triangleq & \min_{O_c^+} \sum_j \left(v_j + \sum_i u_{ji}\right) + \cdots \\ & \left[-\sum_{j\in I}\left((m-1)v_j + \sum_{i\notin I}\min\{v_j, u_{ji}+v_i\}\right) + m(N-1)(1-r)\right] L \end{aligned} \tag{A.104}$$

Examining (A.104), we realize that for (A.94) to be met, O^+ should be defined as the set of all real $\frac{N(N+1)}{2}$-tuples with nonnegative elements that satisfy the following condition for at least one nonempty $I \subseteq \{1, \ldots, N\}$:

$$\sum_{j \in I} \left((m-1)v_j + \sum_{i \notin I} \min\{v_j, v_i + u_{ji}\} \right) \geq m(N-1)(1-r) \tag{A.105}$$

This way, by choosing large enough l, $d_{e^I}(r)$ can be made arbitrary large and thus (A.94) is always met. From (A.105), it follows that

$$\sum_{j \in I} \left((m-1)v_j + \sum_{i \notin I} \min\{v_j, v_i + \max_{j \neq i}\{u_{ji}\}\} \right) \geq m(N-1)(1-r). \tag{A.106}$$

Substituting $\min\{v_j, v_i + \max_{j \neq i}\{u_{ji}\}\}$ in this expression by v_j gives

$$\sum_{j \in I} v_j \geq m(1-r). \tag{A.107}$$

On the other hand, replacing $\min\{v_j, v_i + \max_{j \neq i}\{u_{ji}\}\}$ in (A.106) by $v_i + \max_{j \neq i}\{u_{ji}\}$ results in

$$(m-1)\sum_{j \in I} v_j + m\sum_{i \notin I} (v_i + \max_{j \neq i}\{u_{ji}\}) \geq m(N-1)(1-r). \tag{A.108}$$

Under the constraints given by (A.107) and (A.108), it is easy to see that

$$\inf_{O^+} \left(\sum_{j \in I} v_j + \sum_{i \notin I} (v_i + \max_{j \neq i}\{u_{ji}\}) \right) \geq N(1-r). \tag{A.109}$$

Now, from (A.109) and (A.37), it follows that

$$d_o(r) \geq N(1-r).$$

Again, according to (A.95), $d_o(r)$ provides a lower-bound on the diversity gain achieved by the protocol. Thus the protocol achieves the diversity gain given by (4.7) and the proof is complete.

A.11 Proof of Lemma 12

A simple min-cut max-flow examination reveals that the optimal diversity gain for this channel is upper bounded by those of the 2×1 and 1×2 ARQ MIMO channels, thus

$$d_{\mathrm{R}}(r_e, L) \leq 2(1 - \frac{r_e}{L}) \quad \text{for} \quad 1 > r_e \geq 0. \tag{A.110}$$

Next we prove the achievability of this upper bound, by characterizing the DMT for the proposed ARQ-DDF relay protocol. We do this in two steps. First, we construct an ensemble of random Gaussian codes and characterize its average error probability P_E, and throughput η. We then show, through a simple expurgation argument, that there are codes in the ensemble that perform at least as well as these two averages, therefore achieving P_E and η *simultaneously*.

Let $C(\rho) = \{C_s(\rho), C_r(\rho)\}$ denote the random codes used by the source and the relay, respectively. These are codes of length Ll symbols, rate R_1/L BPCU and generated by an i.i.d complex Gaussian random process of mean zero and variance E. Let us denote the message to be sent by m_0. This message consists of $R_1 l$ information bits. We denote the source and relay codewords corresponding to m_0 by $\mathbf{x}_s(m_0)$ and $\mathbf{x}_r(m_0)$, respectively. We also denote the signature of message m_0 at the destination by $\mathbf{s}(m_0)$, i.e.,

$$\mathbf{y} = \mathbf{s}(m_0) + \mathbf{n}, \tag{A.111}$$

where $\mathbf{y}$ and $\mathbf{n}$ denote the destination received signal and additive noise, respectively. It is important to realize that $\mathbf{s}(m_0)$, not only depends on the message m_0, but also on the channel realization and the relay noise. Also, notice that $\mathbf{x}_r(m_0)$ is only partially

transmitted. This is because the half-duplex relay, itself, needs first to listen to the source to be able to decode the message. Finally, we use superscript ℓ to denote the portion of the signal that corresponds to the first ℓ rounds of transmission. The decoder $\{\varphi, \psi\}$, consists of two functions, $\varphi = \{\varphi^\ell\}_{\ell=1}^{L}$ and $\psi = \{\psi^\ell\}_{\ell=1}^{L-1}$.

- At round ℓ ($L \geq \ell \geq 1$), φ^ℓ outputs the message that minimizes $|\mathbf{y}^\ell - \mathbf{s}^\ell|^2$, i.e.

$$\varphi^\ell(\mathbf{y}^\ell) = \arg\min_m |\mathbf{y}^\ell - \mathbf{s}^\ell(m)|^2, \quad \text{for} \quad L \geq \ell \geq 1. \tag{A.112}$$

  We denote the event that $\varphi^\ell(\mathbf{y}^\ell)$ differs from m_0, with m_0 denoting the transmitted message, by E^ℓ.

- At round ℓ ($L-1 \geq \ell \geq 1$), ψ^ℓ outputs a one, if m is the *unique* message for which

$$|\mathbf{y}^\ell - \mathbf{s}^\ell(m)|^2 \leq \ell l(1+\delta)\sigma^2, \tag{A.113}$$

  where σ^2 denotes the destination noise variance and δ is some positive value. In any other case, ψ^ℓ outputs a zero. We denote the event that ψ^ℓ outputs a one, by A^ℓ.

The decoder uses φ and ψ to decode the message as follows:

1. At the end of round ℓ, ($L-1 \geq \ell \geq 1$), the decoder computes both, $\varphi^\ell(\mathbf{y}^\ell)$ and $\psi^\ell(\mathbf{y}^\ell)$. If $\psi^\ell(\mathbf{y}^\ell) = 1$, then the decoder declares $\varphi^\ell(\mathbf{y}^\ell)$ as the received message and sends back an ACK. Otherwise, it requests for another round of transmission by sending back a NACK signal.

2. At the end of the Lth round, though, the decoder outputs $\varphi^L(\mathbf{y}^L)$ as the received message.

To characterize the average error probability, P_E, we first use the Bayes' rule to write

$$P_E \leq P_{E|\overline{E}_r} + P_{E_r},$$

where E_r and $\overline{E}_r$ denote the events that the relay makes an error in decoding the message, and its complement, respectively. Since the relay starts transmission only after the mutual information between its received signal and the source signal exceeds $R_1 l$, we have (e.g., refer to Theorem 10.1.1 in [25])

$$P_{E_r} \leq \epsilon, \quad \text{for any} \quad \epsilon > 0.$$

This means that

$$P_E \dot{\leq} P_{E|\overline{E}_r}.$$

For the sake of notational simplicity, in the sequel, we denote $P_{E|\overline{E}_r}$ by P_E. In characterizing P_E, we take the approach of El Gamal, Caire and Damon in [29], i.e., we upper bound P_E by

$$P_E \leq \sum_{\ell=1}^{L-1} P_{E^\ell, A^\ell} + P_{E^L}. \tag{A.114}$$

Notice that P_{E^ℓ, A^ℓ} upper bounds the probability of *undetected* errors with $\ell < L$ rounds of transmission, while P_{E^L} upper bounds the probability of *decoding* errors at the end of L transmission rounds. The next step in characterizing P_E is to show that, for the decoder of interest, the undetected errors do not dominate the overall error event, i.e.,

$$P_E \dot{\leq} P_{E^L}. \tag{A.115}$$

Toward this end, we note that

$$P_{E^\ell, A^\ell} \leq \Pr\{|\mathbf{n}^\ell|^2 > \ell l(1+\delta)\sigma^2\}. \tag{A.116}$$

To understand (A.116), let us assume that $|\mathbf{n}^\ell|^2 \leq \ell l(1+\delta)\sigma^2$. One can then use (A.111) to conclude that $|\mathbf{y}^\ell - \mathbf{s}^\ell(m_0)|^2 \leq \ell l(1+\delta)\sigma^2$, where m_0 denotes the transmitted message. This, however, is in contradiction with $E^\ell \cap A^\ell$. This is because the latter event implies that some message m_1, other than m_0, is the *unique* message for which $|\mathbf{y}^\ell - \mathbf{s}^\ell(m_1)|^2 \leq \ell l(1+\delta)\sigma^2$. Thus $E^\ell \cap A^\ell \subseteq \{|\mathbf{n}^\ell|^2 > \ell l(1+\delta)\sigma^2\}$, which means that (A.116) is indeed true. Now, $|\mathbf{n}^\ell|^2$ has a central Chi-squared distribution with $2\ell l$ degrees of freedom. One can use the Chernoff bound to upper bound the tail of this distribution to get

$$\Pr\{|\mathbf{n}^\ell|^2 > \ell l(1+\delta)\sigma^2\} \leq (1+\delta)^{\ell l} e^{-\ell l \delta}. \tag{A.117}$$

This, however, in conjunction with (A.116) means that, for any $\delta > 0$, it is possible to choose T large enough, such that

$$P_{E^\ell, A^\ell} \leq \epsilon, \quad \text{for any} \quad \epsilon > 0. \tag{A.118}$$

Now, (A.115) follows from (A.118), together with (A.114). Examination of E^L reveals that P_{E^L} is the probability of error for the DDF relay protocol at a multiplexing gain of r_1/L, i.e.

$$P_{E^L} \doteq \rho^{-d_{\text{DDF-R}}(\frac{r_1}{L})}, \tag{A.119}$$

where $d_{\text{DDF-R}}(\cdot)$ denotes the diversity gain achieved by the DDF relay protocol (refer to (3.7)). Using (A.115), we conclude

$$P_E \dot{\leq} \rho^{-d_{\text{DDF-R}}(\frac{r_1}{L})}. \tag{A.120}$$

The final step is to show that for $1 > r_1 \geq 0$, we have $r_1 = r_e$. Towards this end, we use (5.1) and notice that for the scenario of interest,

$$p(\ell) \triangleq P_{\overline{A^1}, \cdots, \overline{A^\ell}}, \quad \text{for} \quad L > \ell > 0, \tag{A.121}$$

where $\overline{A^\ell}$ denotes the complement of A^ℓ. To characterize $p(\ell)$, we first upper bound it by

$$p(\ell) \leq P_{\overline{A^\ell}}. \tag{A.122}$$

Careful examination of $\overline{A^\ell}$ reveals that

$$\overline{A^\ell} = R_0^\ell \cup R_1^\ell, \tag{A.123}$$

where R_0^ℓ denotes the subset of destination signals, $\mathbf{y}^\ell$, not included in any of the spheres of squared radius $\ell l(1+\delta)\sigma^2$, centered at the signatures of all possible messages $\{\mathbf{s}^\ell(m)\}$, i.e.,

$$R_0^\ell \triangleq \cap_m \{|\mathbf{y}^\ell - \mathbf{s}^\ell(m)|^2 > \ell l(1+\delta)\sigma^2\}. \tag{A.124}$$

R_1^ℓ, on the other hand, represents the subset of destination signals, $\mathbf{y}^\ell$, included in more than one such spheres, i.e.,

$$R_1^\ell \triangleq \cup_m R_1^\ell(m), \tag{A.125}$$

where

$$R_1^\ell(m) \triangleq \cup_{m_1 \neq m} \{|\mathbf{y}^\ell - \mathbf{s}^\ell(m)|^2 \leq \ell l(1+\delta)\sigma^2, |\mathbf{y}^\ell - \mathbf{s}^\ell(m_1)|^2 \leq \ell l(1+\delta)\sigma^2\}. \tag{A.126}$$

Note that $R_1^\ell(m)$ consists of the intersections of the sphere corresponding to message m, with those of the other messages. Next, we assume that message m_0 is transmitted and characterize $P_{\overline{A^\ell}|m_0}$. Toward this end, we write

$$\begin{aligned} &\overline{A^\ell} = R_0^\ell \cup R_1^\ell, \quad \text{or} \\ &\overline{A^\ell} = R_0^\ell \cup \left(R_1^\ell \cap \overline{R_1^\ell(m_0)}\right) \cup R_1^\ell(m_0), \end{aligned} \tag{A.127}$$

where the last step follows from (A.125). Now, examining (A.124), we have

$$R_0^\ell \subseteq \{|\mathbf{y}^\ell - \mathbf{s}^\ell(m_0)|^2 > \ell l(1+\delta)\sigma^2\}.$$

On the other hand, realizing that $R_1^\ell \cap \overline{R_1^\ell(m_0)}$ consists of the intersections of all spheres excluding the one corresponding to m_0, gives

$$R_1^\ell \cap \overline{R_1^\ell(m_0)} \subseteq \{|\mathbf{y}^\ell - \mathbf{s}^\ell(m_0)|^2 > \ell l(1+\delta)\sigma^2\}.$$

Thus

$$\begin{aligned} R_0^\ell \cup \left(R_1^\ell \cap \overline{R_1^\ell(m_0)}\right) &\subseteq \{|\mathbf{y}^\ell - \mathbf{s}^\ell(m_0)|^2 > \ell l(1+\delta)\sigma^2\}, \\ &= \{|\mathbf{n}^\ell|^2 > \ell l(1+\delta)\sigma^2\}, \end{aligned}$$

where the last step follows from (A.111) and the assumption that m_0 is the transmitted message. Recalling (A.117), we conclude

$$\Pr\{R_0^\ell \cup \left(R_1^\ell \cap \overline{R_1^\ell(m_0)}\right)|m_0\} \leq \epsilon, \quad \text{for any} \quad \epsilon > 0.$$

This, together with (A.127), means that

$$P_{\overline{A^\ell}|m_0} \dot{\leq} \Pr\{R_1^\ell(m_0)|m_0\}. \tag{A.128}$$

Characterization of $\Pr\{R_1^\ell(m_0)|m_0\}$ is a little bit more involved. In particular, if we let $a \triangleq \mathbf{s}^\ell(m_0) - \mathbf{s}^\ell(m_1)$, $b \triangleq \mathbf{y}^\ell - \mathbf{s}^\ell(m_0)$ and $\Delta \triangleq \ell l(1+\delta)\sigma^2$, then

$$\begin{aligned} \{|b|^2 \leq \Delta, |a+b|^2 \leq \Delta\} =& \{|a|^2 \leq 4\Delta, |b|^2 \leq \Delta, |a+b|^2 \leq \Delta\} \cup \\ & \{|a|^2 > 4\Delta, |b|^2 \leq \Delta, |a+b|^2 \leq \Delta\}. \end{aligned}$$

Since the second set on the right-hand side of this expression is empty, we get

$$\begin{aligned} \{|b|^2 \leq \Delta, |a+b|^2 \leq \Delta\} &= \{|a|^2 \leq 4\Delta, |b|^2 \leq \Delta, |a+b|^2 \leq \Delta\} \\ &\subseteq \{|a|^2 \leq 4\Delta\}. \end{aligned}$$

This, along with (A.126), results in

$$R_1^{\ell}(m_0) \subseteq \cup_{m_1 \neq m_0} \{ |\frac{\mathbf{s}^{\ell}(m_0) - \mathbf{s}^{\ell}(m_1)}{2}|^2 \leq \ell l (1+\delta)\sigma^2 \},$$

which gives

$$\Pr\{R_1^{\ell}(m_0)|m_0\} \leq \sum_{m_1 \neq m_0} \Pr\{ |\frac{\mathbf{s}^{\ell}(m_0) - \mathbf{s}^{\ell}(m_1)}{2}|^2 \leq \ell l (1+\delta)\sigma^2 \}. \tag{A.129}$$

We identify the right hand side of (A.129), as the union bound on the ML error probability, conditioned on transmission of m_0 (refer to equation (17) in [17], for a very similar expression). Here, the noise variance is $(1+\delta)\sigma^2$, the code length is ℓl and the rate is R_1/ℓ BPCU. As a result

$$\Pr\{R_1^{\ell}(m_0)\} \dot{\leq} \rho^{-d_{\text{DDF-R}}(\frac{r_1}{\ell})}, \quad \text{for} \quad L > \ell > 0, \tag{A.130}$$

where $d_{\text{DDF-R}}(.)$ is the diversity gain achieved by the DDF relay protocol (recall (3.7)). Now (A.130), along with (A.128) and (A.122) gives

$$p(\ell) \dot{\leq} \rho^{-d_{\text{DDF-R}}(\frac{r_1}{\ell})}, \quad \text{for} \quad L > \ell > 0. \tag{A.131}$$

This means that over the range of $1 > r_1 \geq 0$, the probabilities $\{p(\ell)\}_{\ell=1}^{L-1}$, decay polynomially with ρ. As a result, over this range, r_e equals r_1, i.e.

$$r_e = r_1, \quad \text{for} \quad 1 > r_1 \geq 0. \tag{A.132}$$

Now (A.132), together with (A.120), and the fact that for $1 > r_e \geq 0$, $d_{\text{DDF-R}}(\frac{r_e}{L}) = 2(1 - \frac{r_e}{L})$, gives

$$P_E \dot{\leq} \rho^{-2(1-\frac{r_e}{L})}, \quad \text{for} \quad 1 > r_e \geq 0. \tag{A.133}$$

Note that (A.133) only characterizes the relation between the *average* error probability P_E and *average* throughput η. To complete the proof, we need to show that there

exists a code in the ensemble, that *simultaneously* achieves P_E and η, as characterized by (A.133). Toward this end, we use Lemma 11 of [29]. The application of this lemma to the case of interest is immediate and thus the proof is complete.

A.12 Proof of Theorem 13

A simple max-cut min-flow examination reveals that the optimal diversity gain for this channel is upper bounded by

$$d_{\mathrm{MAR}}(r_e, L) \leq \min\{d_{3\times1}(r_e, L), d_{2\times2}(r_e, L), d_{2\times1}(\frac{r_e}{2}, L), d_{1\times2}(\frac{r_e}{2}, L)\}, \qquad \text{(A.134)}$$

where $d_{m\times n}(\cdot,\cdot)$ denotes the optimal diversity gain for an $m \times n$ ARQ MIMO channel (refer to [29]). Now, (A.134) results in

$$d_{\mathrm{MAR}}(r_e, L) \leq 2 - \frac{r_e}{L}, \quad \text{for} \quad 1 > r_e \geq 0, \qquad \text{(A.135)}$$

which completes the proof of the converse.

Next, we prove that the proposed protocol achieves this upper bound. To do this, we only need to describe the encoder and the decoder. The rest of the proof then follows that of Lemma 12, line by line. Toward this end, let $C(\rho) = \{C_1(\rho), C_2(\rho), C_r(\rho)\}$ denote the random codes used by the two sources and the relay, respectively. These are codes of length Ll, generated by an i.i.d complex Gaussian random process of mean zero and variance E. The rates of these codes are different, though. While $C_1(\rho)$ and $C_2(\rho)$ are of rate $R_1/2L$ BPCU, $C_r(\rho)$ is of rate R_1/L BPCU. In other words, the relay code has twice the rate of the source codes. This means that, corresponding to each pair of source codewords $(\mathbf{x}_1(m_1), \mathbf{x_2}(m_2)) \in C_1(\rho) \times C_2(\rho)$, there exists a codeword $\mathbf{x}_r(\mathbf{m}) \in C_r(\rho)$, where $\mathbf{m} \triangleq (m_1, m_2)$. We call $\mathbf{m}$ the joint message. Note that since each of the two source messages, i.e. m_1 and m_2, consists of $R_1 l/2$

information bits, the joint message $\mathbf{m}$, has a total of $R_1 l$ information bits in it. As before, we denote the destination signature corresponding to the joint message $\mathbf{m}$, by $\mathbf{s}(\mathbf{m})$, i.e.,

$$\mathbf{y} = \mathbf{s}(\mathbf{m}) + \mathbf{n}.$$

In order to decode the source messages and produce the ACK/NACK signals, the destination uses a *joint* bounded distance decoder. This decoder is identical to the one devised for the ARQ-DDF relay protocol (refer to Lemma 12), with the only modification that the joint message $\mathbf{m}$ takes the role of m everywhere, e.g., $\varphi^{\ell}(\cdot)$ is now defined as (compare to (A.112))

$$\varphi^{\ell}(\mathbf{y}^{\ell}) \triangleq \arg\min_{\mathbf{m}} |\mathbf{y}^{\ell} - \mathbf{s}^{\ell}(\mathbf{m})|^2, \quad \text{for} \quad L \geq \ell \geq 1.$$

In the proposed decoder, the destination provides a total of one bit of feedback, for the two sources, per transmission round. Therefore, there is no need for defining *source-specific* $\varphi(\cdot)$ and $\psi(\cdot)$ functions. With the encoder and decoder defined, one can now follow the same steps taken in the proof of Lemma 12 to show that (compare to (A.120))

$$P_E \dot{\leq} \rho^{-d_{\text{DDF-MAR}}(\frac{r_1}{L})}, \tag{A.136}$$

and that (compare to (A.131))

$$p(\ell) \dot{\leq} \rho^{-d_{\text{DDF-MAR}}(\frac{r_1}{\ell})}, \quad \text{for} \quad L > \ell > 0. \tag{A.137}$$

Now, (A.136) and (A.137), together with the fact that for $1 > r_e \geq 0$, $d_{\text{DDF-MAR}}(\frac{r_e}{L}) = 2 - \frac{r_e}{L}$ (refer to (4.3)), result in (compare to (A.133))

$$P_E \dot{\leq} \rho^{-(2-\frac{r_e}{L})}, \quad \text{for} \quad 1 > r_e \geq 0. \tag{A.138}$$

It is then straightforward to use Lemma 11 of [29] to show that there are codes in the ensemble $\{C(\rho)\}$, that achieve (A.138). This proves that the upper bound (A.135) is achievable and thus, completes the proof.

A.13 Proof of Theorem 14

A simple min-cut max-flow examination reveals that the optimal diversity gain for this channel is upper bounded by

$$d_{\text{CVMA}}(r_e, L) \leq \min\{d_{2\times 2}(r_e, L), d_{1\times 3}(\frac{r_e}{2}, L)\}, \tag{A.139}$$

where $d_{m\times n}(\cdot,\cdot)$ denotes the optimal diversity gain for an $m \times n$ ARQ MIMO channel. Now, (A.139) results in (5.5) and the proof of the converse part is complete.

To prove the achievability part, let $C(\rho) = \{C_1(\rho), C_2(\rho)\}$ denote the random codes used by the two sources. These are codes of length $2l$ symbols, rate $R_1/4$ BPCU, and generated by an i.i.d complex Gaussian random process of mean zero and variance E. Let us also denote the two messages to be sent by m_1 and m_2. Note that each message consists of $R_1 l/2$ information bits, such that the joint message $\mathbf{m} \triangleq (m_1, m_2)$ consists of a total of $R_1 l$ bits. We denote the codewords corresponding to m_1 and m_2 by $\mathbf{x}_1(m_1)$ and $\mathbf{x}_2(m_2)$. As before, the destination signatures of m_1, m_2 and $\mathbf{m}$ are denoted by $\mathbf{S}(m_1)$, $\mathbf{S}(m_2)$ and $\mathbf{S}(\mathbf{m})$, respectively. Thus

$$\mathbf{Y} = \mathbf{S}(\mathbf{m}) + \mathbf{N}, \quad \text{or}$$

$$\mathbf{Y} = \mathbf{S}(m_1) + \mathbf{S}(m_2) + \mathbf{N},$$

where $\mathbf{Y} \in \mathbb{C}^{2\times 2l}$ and $\mathbf{N} \in \mathbb{C}^{2\times 2l}$ represent the destination received signal and additive noise, respectively. We denote the signal received through antenna $j \in \{1, 2\}$ by $\mathbf{y}_j$. Similarly, the contribution of message m_i $i \in \{1, 2\}$, to the signal received through

antenna j, is denoted by $\mathbf{s}_j(m_i)$. As before, we use the superscript ℓ to denote the portion of the signal that corresponds to the first ℓ rounds of transmission.

Next, we describe the decoder. Since the performance analysis for the optimal decoder seems intractable, in the sequel, we describe a *suboptimal* bounded distance decoder and analyze its performance. Obviously, this analysis provides a *lower bound* on the diversity gain achieved through the protocol. To describe the decoder, let us label the source and the receiving antenna that are connected through the channel with the highest signal to interference (due to the other source) and noise ratio by s (s stands for superior), while labeling the remaining source and receiving antenna by i (i stands for inferior). This means that,

$$\frac{|g_{ss}|^2\rho}{|g_{is}|^2\rho+\sigma^2} \geq \max\{\frac{|g_{si}|^2\rho}{|g_{ii}|^2\rho+\sigma^2}, \frac{|g_{is}|^2\rho}{|g_{ss}|^2\rho+\sigma^2}, \frac{|g_{ii}|^2\rho}{|g_{si}|^2\rho+\sigma^2}\}, \tag{A.140}$$

where, e.g., g_{si} denotes the gain of the channel connecting source s to receive antenna i. Now, the decoder $\{\varphi, \psi\}$, uses the two sets of functions, $\varphi = \{\varphi_j^1, \varphi_s^1, \varphi_j^2, \varphi_i^2\}$ and $\psi = \{\psi_j^1, \psi_s^1\}$ to decode the messages, and produce the ACK/NACK feedback bits, as follows:

1. At the end of the first round, the decoder uses φ_j^1 to *jointly* decode the two messages, i.e.

$$\varphi_j^1(\mathbf{Y}^1) \triangleq \arg\min_{\mathbf{m}} ||\mathbf{Y}^1 - \mathbf{S}^1(\mathbf{m})||^2. \tag{A.141}$$

   We denote the event that $\varphi_j^1(\mathbf{Y}^1)$ is different from the actual joint message sent, by E_j^1.

2. To decide whether it has correctly decoded the two messages or not, the decoder uses ψ_j^1, where ψ_j^1 outputs a one, if $\mathbf{m}$ is the *unique* joint message satisfying

$$||\mathbf{Y}^1 - \mathbf{S}^1(\mathbf{m})||^2 \leq 2l(1+\delta)\sigma^2. \tag{A.142}$$

In (A.142), δ is some positive value. In any other case, ψ_j^1 outputs a zero. Now, if $\psi_j^1(\mathbf{Y}^1) = 1$, then the decoder sends back ACK signals to *both* of the users, declaring $\varphi_j^1(\mathbf{Y}^1)$ as the decoded joint message. This causes the two sources to start transmission of their next messages. Otherwise, it proceeds to the next step as described below. We denote the event $\psi_j^1(\mathbf{Y}^1) = 1$, by A_j^1.

3. At the end of the first round and in the event of failure in jointly decoding the two messages, i.e., $\overline{A_j^1}$, the decoder uses φ_s^1 to decode the superior message, treating the inferior source's contribution as interference. In doing so, the decoder only utilizes the signal it has received through its s antenna, i.e.

$$\varphi_s^1(\mathbf{y}_s^1) \triangleq \arg\min_{m_s} |\mathbf{y}_s^1 - \mathbf{s}_s^1(m_s)|^2. \tag{A.143}$$

We denote the event that $\varphi_s^1(\mathbf{y}_s^1)$ is different from the actual superior message sent, by E_s^1.

4. To decide whether it has correctly decoded the superior message or not, the decoder uses ψ_s^1, where ψ_s^1 outputs a one, if m_s is the *unique* superior message satisfying

$$|\mathbf{y}_s^1 - \mathbf{s}_s^1(m_s)|^2 \leq l(1+\delta)(|g_{is}|^2\rho + \sigma^2). \tag{A.144}$$

In (A.144), δ is some positive value. In any other case, ψ_s^1 outputs a zero. Now, if $\psi_s^1(\mathbf{y}_s^1) = 0$, the decoder requests for a second round of transmission

by sending back NACK signals to *both* of the sources. However, if $\psi_s^1(\mathbf{y}_s^1) = 1$, then the decoder sends back an ACK signal to the superior source, declaring $\varphi_s^1(\mathbf{y}_s^1)$ as the decoded superior message, while requesting a second round of transmission for the inferior message by sending back a NACK signal to the inferior source. We denote the event $\psi_s^1(\mathbf{y}_s^1) = 1$, by A_s^1.

5. At the end of the second round and conditioned on successful decoding of the superior message in the first round, i.e., $A_s^1 \cap \overline{A_j^1}$, the decoder declares $\varphi_i^2(\mathbf{Y}^2, \varphi_s^1(\mathbf{y}_s^1))$ as the decoded inferior message where

$$\varphi_i^2(\mathbf{Y}^2, \varphi_s^1(\mathbf{y}_s^1)) \triangleq \arg\min_{m_i} ||\mathbf{Y}^2 - \mathbf{S}^2(\varphi_s^1(\mathbf{y}_s^1)) - \mathbf{S}^2(m_i)||^2. \tag{A.145}$$

   In (A.145), $\varphi_s^1(\mathbf{y}_s^1)$ is the decoded superior message as given by (A.143). We denote the event that $\varphi_i^2(\mathbf{Y}^2, \varphi_s^1(\mathbf{y}_s^1))$ is different from the actual inferior message sent, by E_i^2.

6. Finally, in case of failure in decoding the superior message at the end of the first round, i.e., $\overline{A_s^1} \cap \overline{A_j^1}$, the decoder declares $\varphi_j^2(\mathbf{Y}^2)$ as the decoded joint message where

$$\varphi_j^2(\mathbf{Y}^2) \triangleq \arg\min_{\mathbf{m}} ||\mathbf{Y}^2 - \mathbf{S}^2(\mathbf{m})||^2. \tag{A.146}$$

   We denote the event that $\varphi_j^2(\mathbf{Y}^2)$ is different from the actual joint message sent, by E_j^2.

Having described the decoder, we next characterize its average error probability P_E. Toward this end, we first notice that

$$P_E \leq P_{E|\overline{E_r}} + P_{E_r},$$

where E_r and $\overline{E_r}$ denote the event that the relaying source makes an error in decoding the inferior message, and its complement, respectively. Since the superior source only starts relaying after the mutual information between its received signal and inferior source's transmitted signal exceeds $R_1 l/2$, we have (e.g., refer to Theorem 10.1.1 in [25])

$$P_{E_r} \leq \epsilon, \quad \text{for any} \quad \epsilon > 0.$$

This means that

$$P_E \dot{\leq} P_{E|\overline{E_r}}.$$

For the sake of notational simplicity, in the sequel, we denote $P_{E|\overline{E_r}}$ by P_E. To characterize P_E, we write

$$P_E = P_{E_j^1, A_j^1} + P_{E_s^1, A_s^1, \overline{A_j^1}} + P_{E_i^2, \overline{E_s^1}, A_s^1, \overline{A_j^1}} + P_{E_j^2, \overline{A_s^1}, \overline{A_j^1}}. \tag{A.147}$$

To understand (A.147), note that the first two terms correspond to making an *undetected* error in decoding one or both of the messages at the end of the first round, while the last two terms correspond to making a decoding error after requesting for two rounds of transmission. Next, we upper bound (A.147) by

$$P_E \leq P_{E_j^1, A_j^1} + P_{E_s^1, A_s^1} + P_{E_i^2, \overline{E_s^1}} + P_{E_j^2, \overline{A_s^1}}. \tag{A.148}$$

We start evaluating (A.148) by characterizing $P_{E_j^1, A_j^1}$. By examining the definitions for events E_j^1 and A_j^1 (refer to (A.141) and (A.142)), and through an argument similar to the one given for (A.116), we get

$$P_{E_j^1, A_j^1 | g,h} \leq \Pr\{||\mathbf{N}^1||^2 > 2l(1+\delta)\sigma^2\},$$

which for large enough l gives (compare to (A.118))

$$P_{E_j^1,A_j^1|g,h} \leq \epsilon \quad \text{for any} \quad \epsilon > 0, \quad \text{or}$$
$$P_{E_j^1,A_j^1} \leq \epsilon \quad \text{for any} \quad \epsilon > 0. \tag{A.149}$$

Next, we characterize $P_{E_s^1,A_s^1}$. Toward this end, we first fix a channel realization. Then, through examining the definitions for events E_s^1 and A_s^1 (refer to (A.143) and (A.144)), and by pursuing the same steps which led to (A.116), we get

$$P_{E_s^1,A_s^1|g,h} \leq \Pr\{|\mathbf{s}_s^1(m_i) + \mathbf{n}_s^1|^2 > l(1+\delta)(|g_{is}|^2\rho + \sigma^2)\},$$

where m_i denotes the actual inferior message sent and $\mathbf{s}_s^1(m_i)$ represents its signature, at the end of the first round, at the superior antenna. Realizing that, conditioned on a certain channel realization, $|\mathbf{s}_s^1(m_i) + \mathbf{n}_s^1|^2$ has a central Chi-squared distribution with $2l$ degrees of freedom, we conclude that for large enough l, we have (compare to (A.118))

$$P_{E_s^1,A_s^1|g,h} \leq \epsilon \quad \text{for any} \quad \epsilon > 0, \quad \text{or}$$
$$P_{E_s^1,A_s^1} \leq \epsilon \quad \text{for any} \quad \epsilon > 0. \tag{A.150}$$

In other words, (A.149) and (A.150) mean that, through using long enough codes, one can make the probability of making undetected errors arbitrarily small. Note that for doing so, the bounded distance decoder does not employ any kind of cyclic redundancy check (CRC) techniques. Now, using (A.148), (A.149) and (A.150) we conclude

$$P_E \dot{\leq} P_{E_i^2,\overline{E_s^1}} + P_{E_j^2,\overline{A_s^1}}. \tag{A.151}$$

To characterize $P_{E_i^2,\overline{E_s^1}}$, we first fix a channel realization and then write

$$\begin{aligned} P_{E_i^2,\overline{E_s^1}|g,h} &= P_{\overline{E_s^1}|g,h} P_{E_i^2|\overline{E_s^1},g,h}, \\ &\leq P_{E_i^2|\overline{E_s^1},g,h}. \end{aligned} \tag{A.152}$$

Now, using (A.145), it is straightforward to verify that (refer to (2.7))

$$\begin{aligned} P_{PE_i^2|\overline{E_s^1},g,h} \leq & \big(1+\frac{1}{2}\rho(|g_{is}|^2+|g_{ii}|^2)\big)^{-(l+l')} \times \\ & \big(1+\frac{1}{2}\rho(|g_{ss}|^2+|g_{si}|^2+|g_{is}|^2+|g_{ii}|^2)+\frac{1}{4}\rho^2\det(GG^H)\big)^{-(l-l')}, \end{aligned} \tag{A.153}$$

where

$$G \triangleq \begin{bmatrix} g_{ss} & g_{is} \\ g_{si} & g_{ii} \end{bmatrix}.$$

In (A.153), l' is the number of symbol intervals, in the second round, that the superior source needs to listen to the inferior one, before decoding its message, i.e.

$$l' \triangleq \min\{l, \lceil \frac{lR_1}{2\log_2(1+|h|^2 c\rho)} \rceil\}. \tag{A.154}$$

Since GG^H is a positive semi-definite matrix, we have

$$\begin{aligned} P_{PE_i^2|\overline{E_s^1},g,h} \leq & \big(1+\frac{1}{2}\rho(|g_{is}|^2+|g_{ii}|^2)\big)^{-(l+l')} \times \\ & \big(1+\frac{1}{2}\rho(|g_{ss}|^2+|g_{si}|^2+|g_{is}|^2+|g_{ii}|^2)\big)^{-(l-l')}. \end{aligned} \tag{A.155}$$

Note that in deriving (A.155) from (A.153), we have ignored the term $\frac{1}{4}\rho^2\det(GG^H)$. As a consequence, the derived upper bound may be loose. However, this is indeed necessary, for the sake of analysis tractability. Now, let us define v_{kl}, where $k,l \in \{s,i\}$, as the exponential order of $1/|g_{kl}|^2$, and u, as the exponential order of $1/|h|^2$. Then, using (A.155) and realizing that there are a total of $\rho^{\frac{lr_1}{2}}$ codewords in the inferior source's code-book, we derive

$$P_{E_i^2|\overline{E_s^1},v,u} \dot{\leq} \rho^{-l[(1-\min\{v_{is},v_{ii}\})^+(1+f)+(1-\min\{v_{ss},v_{si},v_{is},v_{ii}\})^+(1-f)-\frac{r_1}{2}]}, \tag{A.156}$$

where $f \triangleq l'/l$. Using (A.154), we have

$$f = \min\{1, \frac{r_1}{2(1-u)^+}\}. \tag{A.157}$$

An argument similar to that given for (A.80), reveals that if we define the outage event O_i^{2+} as

$$\begin{aligned} O_i^{2+} \triangleq \{(v_{ss}, \cdots, u) \in \mathbb{R}^{5+} | (1 - \min\{v_{is}, v_{ii}\})^+ (1+f) + \\ (1 - \min\{v_{ss}, v_{si}, v_{is}, v_{ii}\})^+ (1-f) \leq \frac{r_1}{2}\}, \end{aligned} \tag{A.158}$$

then

$$P_{E_i^2 | \overline{E_s^1}} \dot{\leq} P_{O_i^2}, \tag{A.159}$$

where (recall (A.85))

$$P_{O_i^2} \doteq \rho^{-d_i(r_1)}, \quad \text{with} \quad d_i(r_1) \triangleq \inf_{O_i^{2+}} \{v_{ss} + v_{si} + v_{is} + v_{ii} + u\}. \tag{A.160}$$

In writing (A.160), we have used the fact that

$$v_{ss} + v_{si} + v_{is} + v_{ii} = v_{11} + v_{12} + v_{21} + v_{22}.$$

Obviously, $(v_{ss}, v_{si}, v_{is}, v_{ii})$ should also satisfy

$$\begin{aligned} 1 - v_{ss} - (1 - v_{is})^+ \geq \max\{1 - v_{si} - (1 - v_{ii})^+, 1 - v_{is} - (1 - v_{ss})^+, \\ 1 - v_{ii} - (1 - v_{si})^+\}, \end{aligned} \tag{A.161}$$

which is the counterpart of (A.140), stated in terms of the channel exponential orders. Now, to characterize $d_i(r_1)$, we first derive $\inf_{O_i^{2+}} \{v_{ss} + v_{si} + v_{is} + v_{ii}\}$, as a function of f, i.e.

$$\inf_{O_i^{2+}} \{v_{ss} + v_{si} + v_{is} + v_{ii}\}(f) = \lambda_i(f),$$

where

$$\lambda_i(f) \triangleq \begin{cases} 4 - \frac{r_1}{1-f}, & 1 - \frac{r_1}{2} > f \geq 0 \\ \frac{4-r_1}{1+f}, & 1 \geq f \geq 1 - \frac{r_1}{2} \end{cases} . \tag{A.162}$$

On the other hand, from (A.157) we get

$$\inf\{u\}(f) = \lambda(f), \quad \text{where} \quad \lambda(f) \triangleq 1 - \frac{r_1}{2f}, \quad f \geq \frac{r_1}{2}. \tag{A.163}$$

Then, from (A.160), we get

$$d_i(r_1) = \inf_{1 \geq f \geq \frac{r_1}{2}} \lambda_i(f) + \lambda(f).$$

The right hand side of this expression can be derived using (A.162) and (A.163), i.e.

$$d_i(r_1) = \begin{cases} 3 - r_1, & 1 > r_1 \geq 0 \\ \frac{2(4-r_1)}{2+r_1}, & 2 \geq r_1 \geq 1 \end{cases} . \tag{A.164}$$

This, together with (A.160), (A.159) and (A.152), completes the characterization of $P_{E_i^2, \overline{E_s^1}}$, i.e.

$$P_{E_i^2, \overline{E_s^1}} \dot{\leq} \rho^{-d_i(r_1)}. \tag{A.165}$$

To characterize the second term of (A.151), i.e., $P_{E_j^2, \overline{A_s^1}}$, we first split the event E_j^2 (refer to (A.146)) into

$$E_j^2 = E_{jsi}^2 \cup E_{ji}^2 \cup E_{js}^2,$$

where E_{jsi}^2, E_{ji}^2 and E_{js}^2 represent the events that, at the end of the second round, the joint decoder makes errors in decoding both of the messages, only the inferior message and only the superior message, respectively. Since these three events are mutually exclusive, we have

$$P_{E_j^2, \overline{A_s^1}} = P_{E_{jsi}^2, \overline{A_s^1}} + P_{E_{ji}^2, \overline{A_s^1}} + P_{E_{js}^2, \overline{A_s^1}}. \tag{A.166}$$

Characterization of the first term is very straightforward,

$$\begin{aligned} P_{E^2_{jsi},\overline{A^1_s}} &\leq P_{E^2_{jsi}}, \\ &\doteq \rho^{-d_{2\times 2}(\frac{r_1}{2})}. \end{aligned} \tag{A.167}$$

Characterization of $P_{E^2_{ji},\overline{A^1_s}}$, though, is a little bit more involved. In particular, notice that

$$P_{E^2_{ji},\overline{A^1_s}} \leq \min\{P_{E^2_{ji}}, P_{\overline{A^1_s}}\}. \tag{A.168}$$

Now, conditioned on a certain channel realization, $P_{PE^2_{ji}|g,h}$ can be upper bounded as (refer to (2.7))

$$P_{PE^2_{ji}|g,h} \leq \big(1+\frac{1}{2}\rho(|g_{is}|^2+|g_{ii}|^2)\big)^{-2l}.$$

Realizing that there are a total of $\rho^{\frac{lr_1}{2}}$ pairs of codewords in each source's code book, we get

$$P_{E^2_{ji}|v,u} \dot{\leq} \rho^{-2l[(1-\min\{v_{is},v_{ii}\})^+-\frac{r_1}{4}]}. \tag{A.169}$$

Now, examining (A.169) reveals that if we define O^{2+}_{ji} as

$$O^{2+}_{ji} \triangleq \{(v_{ss},\cdots,u)\in\mathbb{R}^{5+} | (1-\min\{v_{is},v_{ii}\})^+ \leq \frac{r_1}{4}\}, \tag{A.170}$$

then for all channel realizations $(v_{ss},\cdots,u)\in\mathbb{R}^{5+}$ *not* included in O^{2+}_{ji}, $P_{E^2_{ji}|v,u}$ can be made arbitrary small, provided that l is large enough, i.e.

$$P_{E^2_{ji}|v,u} \leq \epsilon, \quad \text{for any} \quad \epsilon>0 \quad \text{and} \quad (v_{ss},\cdots,u)\in\overline{O^2_{ji}}\cap\mathbb{R}^{5+}. \tag{A.171}$$

Next, we turn our attention to $P_{\overline{A^1_s}}$. To characterize this probability, we first fix a channel realization. Then, by comparing the definition of A^1_s (refer to (A.144)) to that

of A^{ℓ} (refer to (A.113)), and pursuing the exact same arguments which led (A.123) to (A.131), we conclude

$$P_{\overline{A_s^1}|g,h} \dot{\leq} P_{E_s^1|g,h}. \tag{A.172}$$

But $P_{PE_s^1|g,h}$ is given by (notice how the inferior source's contribution is treated as interference),

$$P_{PE_s^1|g,h} \leq \left(1+\frac{1}{2}\rho(\frac{|g_{ss}|^2}{1+\rho|g_{is}|^2})\right)^{-l},$$

which, in terms of channel exponentials, translates into

$$P_{E_s^1|v,u} \dot{\leq} \rho^{-l\left[\left(1-v_{ss}-(1-v_{is})^+\right)^+-\frac{r_1}{2}\right]}. \tag{A.173}$$

Now, from (A.172) and (A.173), we conclude

$$P_{\overline{A_s^1}|v,u} \dot{\leq} \rho^{-l\left[\left(1-v_{ss}-(1-v_{is})^+\right)^+-\frac{r_1}{2}\right]}. \tag{A.174}$$

This means that if

$$O_s^{1+} \triangleq \{(v_{ss},\cdots,u)\in\mathbb{R}^{5+}|\left(1-v_{ss}-(1-v_{is})^+\right)^+ \leq \frac{r_1}{2}\}, \tag{A.175}$$

then for all channel realizations $(v_{ss},\cdots,u)\in\mathbb{R}^{5+}$, *not* included in O_s^{1+}, $P_{\overline{A_s^1}|v,u}$ can be made arbitrary small, provided that l is large enough, i.e.

$$P_{\overline{A_s^1}|v,u} \leq \epsilon, \quad \text{for all} \quad \epsilon>0 \quad \text{and} \quad (v_{ss},\cdots,u)\in\overline{O_s^1}\cap\mathbb{R}^{5+}. \tag{A.176}$$

Now, using (A.168), (A.171) and (A.176) we conclude

$$P_{E_{ji}^2,\overline{A_s^1}|u,v} \leq \epsilon \quad \text{for all} \quad \epsilon>0 \quad \text{and} \quad (v_{ss},\cdots,u)\in\overline{O_{s,ji}^{1,2}}\cap\mathbb{R}^{5+}, \tag{A.177}$$

where

$$O_{s,ji}^{1,2+} \triangleq O_s^{1+}\cap O_{ji}^{2+}, \tag{A.178}$$

or

$$O_{s,ji}^{1,2+} = \{(v_{ss}, \cdots, u) \in \mathbb{R}^{5+} | (1 - \min\{v_{is}, v_{ii}\})^+ \leq \frac{r_1}{4}, \\ \left(1 - v_{ss} - (1 - v_{is})^+\right)^+ \leq \frac{r_1}{2}\}. \tag{A.179}$$

This means that

$$P_{E_{ji}^2, \overline{A_s^1}} \dot{\leq} P_{O_{s,ji}^{1,2}}, \tag{A.180}$$

where

$$P_{O_{s,ji}^{1,2}} \doteq \rho^{-d_{s,ji}(r_1)}, \quad \text{with} \quad d_{s,ji}(r_1) \triangleq \inf_{O_{s,ji}^{12+}} \{v_{ss} + \cdots + u\}. \tag{A.181}$$

Now, using (A.179), it is straightforward to show that

$$d_{s,ji}(r_1) = \begin{cases} 4 - 2r_1, & \frac{4}{3} > r_1 \geq 0 \\ 2 - \frac{r_1}{2}, & 2 \geq r_1 \geq \frac{4}{3} \end{cases}. \tag{A.182}$$

This, together with (A.181) and (A.180), completes the characterization of $P_{E_{ji}^2, \overline{A_s^1}}$, i.e.

$$P_{E_{ji}^2, \overline{A_s^1}} \dot{\leq} \rho^{-d_{s,ji}(r_1)}. \tag{A.183}$$

Characterization of $P_{E_{js}^2, \overline{A_s^1}}$, proceeds in a similar way. In particular, pursuing the exact same steps leading to (A.169), reveals that

$$P_{E_{js}^2 | v,u} \dot{\leq} \rho^{-2l[(1-\min\{v_{ss}, v_{si}\})^+ - \frac{r_1}{4}]}.$$

This means that defining O_{js}^{2+} as

$$O_{js}^{2+} \triangleq \{(v_{ss}, \cdots, u) \in \mathbb{R}^{5+} | (1 - \min\{v_{ss}, v_{si}\})^+ \leq \frac{r_1}{4}\},$$

results in

$$P_{E_{js}^2 | v,u} \leq \epsilon, \quad \text{for any} \quad \epsilon > 0 \quad \text{and} \quad (v_{ss}, \cdots, u) \in \overline{O_{js}^2} \cap \mathbb{R}^{5+}.$$

This, however, together with (A.176), gives

$$P_{E^2_{js},\overline{A^1_s}|u,v} \leq \epsilon \ \text{ for all } \ \epsilon > 0 \ \text{ and } \ (v_{ss}, \cdots, u) \in \overline{O^{1,2}_{s,js}} \cap \mathbb{R}^{5+},$$

where

$$O^{1,2+}_{s,js} \triangleq O^{1+}_{s} \cap O^{2+}_{js}.$$

or

$$O^{1,2+}_{s,js} = \{(v_{ss}, \cdots, u) \in \mathbb{R}^{5+} | (1 - \min\{v_{ss}, v_{si}\})^+ \leq \frac{r_1}{4}, \\ \left(1 - v_{ss} - (1 - v_{is})^+\right)^+ \leq \frac{r_1}{2}\},$$

or

$$O^{1,2+}_{s,js} = \{(v_{ss}, \cdots, u) \in \mathbb{R}^{5+} | (1 - \min\{v_{ss}, v_{si}\})^+ \leq \frac{r_1}{4}\}. \tag{A.184}$$

Thus

$$P_{E^2_{js},\overline{A^1_s}} \dot{\leq} \rho^{-d_{s,js}(r_1)}, \tag{A.185}$$

where

$$d_{s,js}(r_1) \triangleq \inf_{O^{1,2+}_{s,js}} \{v_{ss} + \cdots + u\}.$$

Now using (A.184), together with (A.161), it is a simple matter to show that

$$d_{s,js}(r_1) = 4 - r_1. \tag{A.186}$$

This completes the characterization of $P_{E^2_{js},\overline{A^1_s}}$. Next, we use (A.166), (A.167), (A.183) and (A.185) we conclude

$$P_{E^2_{j},\overline{A^1_s}} \dot{\leq} \rho^{-d_{s,j}(r_1)}, \tag{A.187}$$

where

$$d_{s,j}(r_1) = \min\{d_{2\times 2}(r_1), d_{s,ji}(r_1), d_{s,js}(r_1)\}.$$

Using (A.182) and (A.186), however, we get

$$d_{s,j}(r_1) = d_{s,ji}(r_1). \tag{A.188}$$

Finally, (A.151), together with (A.165) and (A.187), gives

$$d_{\text{DDF-CVMA}}(r_1, 2) \geq \min\{d_i(r_1), d_{s,j}(r_1)\},$$

where $d_{\text{DDF-CVMA}}(r_1, 2)$ denotes the diversity gain achieved by the protocol. Now, using (A.164), (A.188) and (A.182) we get

$$d_{\text{DDF-CVMA}}(r_1, 2) \geq \begin{cases} 3 - r_1, & 1 > r_1 \geq 0 \\ 4 - 2r_1, & \frac{4}{3} > r_1 \geq 1 \\ 2 - \frac{r_1}{2}, & 2 > r_1 \geq \frac{4}{3} \end{cases} \tag{A.189}$$

Next, we prove that for $2 > r_1 \geq 0$, $r_{\text{e}} = r_1$. To do this, we only need to characterize $p(1)$ (recall (5.1)). Toward this end, we observe that

$$\begin{aligned} p(1) =& P_{\overline{A_j^1}}, \\ \doteq& P_{E_j^1}. \end{aligned} \tag{A.190}$$

$P_{E_j^1}$, however, is the joint error probability, for a multiple access channel with two single-antenna users and one double-antenna destination, which is known to be (refer to [18])

$$P_{E_j^1} \doteq \rho^{-(2-r_1)}. \tag{A.191}$$

Now, (A.191), together with (A.190), proves that $p(1)$ decays polynomially with ρ over the range $2 > r_1 \geq 0$. Thus

$$r_e = r_1 \quad \text{for} \quad 2 > r_1 \geq 0. \tag{A.192}$$

From (A.192) and (A.189), we conclude

$$d_{\text{DDF-CVMA}}(r_e, 2) \geq \begin{cases} 3 - r_e, & 1 > r_e \geq 0 \\ 4 - 2r_e, & \frac{4}{3} > r_e \geq 1 \\ 2 - \frac{r_e}{2}, & 2 > r_e \geq \frac{4}{3} \end{cases} \tag{A.193}$$

Now, application of Lemma 11 of [29] shows that there are codes in the ensemble that achieve (A.190) and (A.193), *simultaneously*. This proves the achievability of (5.6).

Next, we prove the asymptotic optimality result. To do this, however, we first need to generalize the protocol. To extend the protocol to the case of L ARQ rounds, the random codes $C(\rho) = \{C_1(\rho), C_2(\rho)\}$ used by the two sources should be modified such that each one is of length Ll symbols and rate $R_1/2L$ BPCU. Notice that these choices leave the total number of information bits per joint message equal to $R_1 l$ bits. The decoder used by the destination is also a direct extension of the one described in the proof of the achievability part. In particular, at the end of the ℓth round ($L \geq \ell \geq 1$) and in the event that none of the two messages has yet been successfully decoded, the decoder uses φ_j^ℓ and ψ_j^ℓ to jointly decode the two messages (φ_j^ℓ and ψ_j^ℓ are defined in a way similar to (A.141) and (A.142)). If successful, the decoder sends back ACK signals to both of the users (i.e., event A_j^ℓ), otherwise it tries to decode the superior message, using φ_s^ℓ and ψ_s^ℓ (which are defined in a way similar to (A.143) and (A.144)). In the case that the decoder succeeds in decoding the superior message (i.e., event A_s^ℓ), an ACK signal is sent to the corresponding user. The description given so far pertains to the scenario where none of the messages is successfully decoded by the end of the $(\ell - 1)$th round. If the superior message is already decoded, the decoder uses φ_i^ℓ ($L \geq \ell > 1$) to decode the inferior message, i.e.

$$\varphi_i^\ell(\mathbf{Y}^\ell, m_s) \triangleq \arg\min_{m_i} ||\mathbf{Y}^\ell - \mathbf{S}^\ell(m_s) - \mathbf{S}^\ell(m_i)||^2, \tag{A.194}$$

where m_s denotes the decoded superior message. To decide whether it has decoded the inferior message error-free or not, the decoder uses ψ_i^ℓ, where ψ_i^ℓ outputs a one, if m_i is the *unique* message satisfying

$$||\mathbf{Y}^\ell - \mathbf{S}^\ell(m_s) - \mathbf{S}^\ell(m_i)||^2 \leq 2\ell l(1+\delta)\sigma^2.$$

In any other case, ψ_i^ℓ outputs a zero. Now, if $\psi_i^\ell(\mathbf{Y}^\ell, m_s) = 1$, then the decoder sends back an ACK signal to the inferior user (i.e., event A_i^ℓ), declaring $\varphi_i^\ell(\mathbf{Y}^\ell, m_s)$ as the decoded inferior message.

Having described the decoder, we next characterize its error probability, P_E, as L grows to infinity. Note that P_E can be written as

$$P_E = \Sigma_{\ell=1}^{L} P_{E^\ell}, \tag{A.195}$$

where E^ℓ denotes the event of making an error at the end of the ℓth round. It is important to realize that for $\ell < L$, E^ℓ corresponds to an *undetected* error, i.e., one for which an ACK signal is sent back. Now, an argument similar to those given for (A.118), (A.149) and (A.150) reveals that for large enough l,

$$P_{E^\ell} \leq \epsilon, \quad \text{for} \quad L > \ell \geq 1.$$

This together with (A.195) gives

$$P_E \dot{\leq} P_{E^L}. \tag{A.196}$$

But

$$P_{E^L} \leq \Sigma_{\ell=1}^{L-1} P_{E_i^L, \overline{E_s^\ell}, A_s^\ell, \overline{A_j^\ell}} + P_{E_j^L, \overline{A_s^{L-1}}, \overline{A_j^{L-1}}}, \tag{A.197}$$

where $P_{E_i^L, \overline{E_s^\ell}, A_s^\ell, \overline{A_j^\ell}}$ upper-bounds the probability that the decoder makes an error after L rounds of transmission, conditioned on successful decoding of the superior message

at the end of the ℓth round. $P_{E_j^L,\overline{A_s^{L-1}},\overline{A_j^{L-1}}}$, on the other hand, upper-bounds the decoder error probability after L rounds, given that it does not decode any of the messages by the $(L-1)$th round. To characterize $P_{E_i^L,\overline{E_s^\ell},A_s^\ell,\overline{A_j^\ell}}$, notice that

$$P_{E_i^L,\overline{E_s^\ell},A_s^\ell,\overline{A_j^\ell}} \leq \min\{P_{E_i^L,\overline{E_s^\ell},A_s^\ell}, P_{\overline{A_j^\ell}}\}, \quad \text{for} \quad L > \ell \geq 1. \tag{A.198}$$

Since we are only interested in the case where L grows to infinity, we can assume that L is even. Now for $\frac{L}{2} \geq \ell \geq 1$, we have

$$P_{E_i^L,\overline{E_s^\ell},A_s^\ell} \leq P_{E_i^L,\overline{E_s^{L/2}},A_s^{L/2}}. \tag{A.199}$$

This is because in the event $E_i^L\overline{E_s^\ell}A_s^\ell$, ℓ corresponds to the number of rounds during which the superior user is acting as a DDF relay for the inferior one. Likewise, for $\ell \geq \frac{L}{2}+1$

$$P_{\overline{A_j^\ell}} \leq P_{\overline{A_j^{L/2}}}. \tag{A.200}$$

This is because in the event $\overline{A_j^\ell}$, ℓ corresponds to the number of rounds during which the two users are simultaneously transmitting their messages. Clearly, increasing ℓ reduces $P_{\overline{A_j^\ell}}$ (notice that for exactly the same reason, $P_{E_j^L} \leq P_{E_j^{L/2}}$). Following the same steps leading to (A.131) reveals that

$$P_{\overline{A_j^{L/2}}} \dot{\leq} P_{E_j^{L/2}}, \tag{A.201}$$

which simply states that the probability of sending back a NACK signal is of the same exponential order as the probability of making a joint error. Now from equations (A.197) to (A.201), we conclude

$$P_{E^L} \dot{\leq} \frac{L}{2}(P_{E_i^L,\overline{E_s^{L/2}},A_s^{L/2}} + P_{\overline{E_j^{L/2}}}), \quad \text{for } L \text{ even}. \tag{A.202}$$

Comparing (A.202) with (A.151), we realize that P_{E^L} is of same exponential order as the error probability of the ARQ-DDF CVMA protocol, with two rounds of transmission and a destination first-round rate of $\frac{2R_1}{L}$ BPCU, i.e.

$$P_{E^L} \dot{\leq} \rho^{-d_{\text{DDF-CVMA}}(\frac{2r_1}{L},2)}, \quad \text{for} \quad 2 > r_1 \geq 0,$$

where $d_{\text{DDF-CVMA}}(r_1, 2)$ is given by (A.189). Now, letting L to grow to infinity, together with (A.196) and (A.189), gives

$$\lim_{L\to\infty} P_E \dot{\leq} \rho^{-3}, \quad \text{for} \quad 2 > r_1 \geq 0. \tag{A.203}$$

To prove that $r_e = r_1$, we notice that (recall (5.1))

$$\eta \leq \frac{R_1}{1 + p(1)},$$

where

$$\begin{aligned} p(1) =& P_{\overline{A_j^1}}, \\ \dot{=}& \rho^{-(2-r_1)}. \end{aligned} \tag{A.204}$$

The last step of (A.204) follows from the same argument given for (A.191). This shows that $r_e = r_1, 2 > r_1 \geq 0$, which together with (A.203) proves (5.7). Note that the proof for existence of codes that achieve P_E and η simultaneously results from application of Lemma 11 of [29].

A.14 Proof of Theorem 16

For the channel described by (6.1), an outage is defined as the event that the realized mutual information does not support the intended rate, i.e.

$$O_{p(\mathbf{x})} \triangleq \{H \in \mathbb{C}^{n\times m} | I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H) < R\}. \tag{A.205}$$

Notice that the mutual information depends on both, the realized channel H and the input distribution $p(\mathbf{x})$. For this channel, the outage probability $P_o(R,\rho)$ is defined as

$$P_o(R,\rho) = \inf_{p(\mathbf{x})} \Pr\{O_{p(\mathbf{x})}\}.$$

It is shown in [17] that

$$P_o(R,\rho) \leq \Pr\{\log\det(I_n + \frac{\rho}{m}HH^H) < R\} \text{ and}$$
$$P_o(R,\rho) \geq \Pr\{\log\det(I_n + \rho HH^H) < R\}.$$

These equations can be re-written as

$$P_o(R,\rho) \leq \Pr\{\log(\prod_{i=1}^{\min\{m,n\}}(1+\frac{\rho}{m}\mu_i)) < R\} \text{ and} \tag{A.206}$$
$$P_o(R,\rho) \geq \Pr\{\log(\prod_{i=1}^{\min\{m,n\}}(1+\rho\mu_i)) < R\}, \tag{A.207}$$

where $\mu_{\min\{m,n\}} \geq \cdots \geq \mu_1 \geq 0$ represent the ordered eigenvalues of HH^H. The joint Probability Density Function (PDF) of $(\mu_1, \cdots, \mu_{\min\{m,n\}})$ is given by the Wishart distribution, i.e.

$$p(\mu) = K_{m,n}^{-1} \prod_{i=1}^{\min\{m,n\}} \mu_i^{|m-n|} \prod_{i<j}(\mu_i - \mu_j)^2 e^{-\sum_i \mu_i}, \tag{A.208}$$

where $\mu \triangleq (\mu_1, \cdots, \mu_{\min\{m,n\}})$ and $K_{m,n}$ is a normalizing factor. Now, let us focus on (A.206) and introduce the change of variables

$$\alpha_i \triangleq \frac{\log(1+\frac{\rho}{m}\mu_i)}{R}. \tag{A.209}$$

This implies that $\alpha_{\min\{m,n\}} \geq \cdots \geq \alpha_1 \geq 0$. In terms of the new variables, (A.206) can be written as

$$P_o(R,\rho) \leq \Pr\{\mathcal{A}\}, \tag{A.210}$$

where

$$\mathcal{A} \triangleq \{\alpha | \alpha_{\min\{m,n\}} \geq \cdots \geq \alpha_1 \geq 0, 1 - \sum_i \alpha_i > 0\}. \tag{A.211}$$

In (A.211), $\alpha \triangleq (\alpha_1, \cdots, \alpha_{\min\{m,n\}})$. On the other hand, (A.208) becomes

$$\begin{aligned} p(\alpha) =& KR^{\min\{m,n\}} \rho^{-mn} 2^{R\sum_i \alpha_i} \times \\ & \prod_{i=1}^{\min\{m,n\}} (2^{\alpha_i R} - 1)^{|m-n|} \prod_{i<j} (2^{\alpha_i R} - 2^{\alpha_j R})^2 e^{-\sum_i \frac{m(2^{\alpha_i R}-1)}{\rho}}, \end{aligned} \tag{A.212}$$

where $K \triangleq K_{m,n}^{-1} (\ln 2)^{\min\{m,n\}} m^{mn}$. Next, we define $\mathcal{R}_\delta(k)$, for integer k's, as

$$\mathcal{R}_\delta(k) \triangleq \begin{cases} \{R | \frac{1}{\delta} > \frac{\log \rho}{R} > 1 + \delta\} & \text{if} \quad k = 0 \\ \{R | \frac{1}{k} - \delta > \frac{\log \rho}{R} > \frac{1}{k+1} + \delta\} & \text{if} \quad k \in \mathbb{Z}, \min\{m,n\} > k > 0 \end{cases}, \tag{A.213}$$

where δ denotes a small positive value. Notice that $\delta = 0$ reduces $\mathcal{R}_\delta(k)$ to $\mathcal{R}(k)$, as given by (6.5). Now, it follows from (A.210) that

$$P_o(R, \rho) 2^{-c(k)R} \leq 2^{-c(k)R} \int_{\mathcal{A}} p(\alpha) \mathrm{d}\alpha, \quad R \in \mathcal{R}_\delta(k)$$

Note that this expression holds, regardless of the choice for $c(k)$, i.e., at this point we regard $c(k)$ as an *arbitrary* function of k. This inequality can be written as

$$P_o(R, \rho) 2^{-c(k)R} \leq A_1(R, \rho, \epsilon) + A_2(R, \rho, \epsilon), \quad R \in \mathcal{R}_\delta(k)$$

where

$$A_1(R, \rho, \epsilon) \triangleq 2^{-c(k)R} \int_{\mathcal{A}_1} p(\alpha) \mathrm{d}\alpha, \qquad A_2(R, \rho, \epsilon) \triangleq 2^{-c(k)R} \int_{\mathcal{A}_2} p(\alpha) \mathrm{d}\alpha$$

and

$$\mathcal{A}_1 \triangleq \{\alpha \in \mathcal{A} | \alpha_{\min\{m,n\}} > \frac{\log \rho}{R} + \epsilon\}, \qquad \mathcal{A}_2 \triangleq \{\alpha \in \mathcal{A} | \frac{\log \rho}{R} + \epsilon \geq \alpha_{\min\{m,n\}}\}. \tag{A.214}$$

This means that

$$\limsup_{\substack{\rho\to\infty\\ R\in\mathcal{R}_\delta(k)}} \frac{\log P_o(R,\rho)-c(k)R}{\log\rho} \leq \limsup_{\substack{\rho\to\infty\\ R\in\mathcal{R}_\delta(k)}} \frac{\log\big(1+A_1(R,\rho,\epsilon)/A_2(R,\rho,\epsilon)\big)}{\log\rho} + \limsup_{\substack{\rho\to\infty\\ R\in\mathcal{R}_\delta(k)}} \frac{\log A_2(R,\rho,\epsilon)}{\log\rho}, \tag{A.215}$$

To characterize the first term in the right-hand side of (A.215), we notice that

$$\begin{aligned} A_1(R,\rho,\epsilon) \leq & K R^{\min\{m,n\}} e^{\frac{m}{\rho}} \rho^{-mn} 2^{-c(k)R} \times \\ & \int_{\mathcal{A}_1} 2^{R\sum_i \alpha_i} \prod_{i=1}^{\min\{m,n\}} 2^{\alpha_i |m-n| R} \prod_{i<j} 2^{2\alpha_j R} e^{-m2^{(\alpha_{\min\{m,n\}} - \frac{\log\rho}{R})R}} \mathrm{d}\alpha \\ \leq & K R^{\min\{m,n\}} e^{\frac{m}{\rho}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{A}_1} 2^{f(\alpha)R} e^{-m2^{\epsilon R}} \mathrm{d}\alpha, \end{aligned}$$

where

$$f(\alpha) \triangleq \sum_{i=1}^{\min\{m,n\}} (|m-n|+2i-1)\alpha_i. \tag{A.216}$$

Realizing that $\mathrm{Vol}\{\mathcal{A}_1\} \leq 1$, we conclude

$$A_1(R,\rho,\epsilon) \leq K R^{\min\{m,n\}} (2^{\epsilon R})^{\frac{f_1-c(k)}{\epsilon}} e^{-m2^{\epsilon R}} e^{\frac{m}{\rho}} \rho^{-mn}, \tag{A.217}$$

where

$$f_1 \triangleq \sup_{\mathcal{A}_1} f(\alpha).$$

On the other hand

$$\begin{aligned}A_2(R,\rho,\epsilon) \geq & KR^{\min\{m,n\}} e^{\frac{m\min\{m,n\}}{\rho}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{A}_2} e^{-\sum_i m 2^{-(\frac{\log\rho}{R}-\alpha_i)R}} 2^{R\sum_i \alpha_i} \times \\ & \prod_{i=1}^{\min\{m,n\}} (1-2^{-\alpha_i R})^{|m-n|} 2^{|m-n|\alpha_i R} \prod_{i<j} (1-2^{-(\alpha_j-\alpha_i)R})^2 2^{2\alpha_j R} \mathrm{d}\alpha, \\ \geq & KR^{\min\{m,n\}} e^{\frac{m\min\{m,n\}}{\rho}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{A}_{\epsilon_1}} e^{-m\min\{m,n\}2^{-\epsilon_1 R}} 2^{R\sum_i \alpha_i} \times \\ & \prod_{i=1}^{\min\{m,n\}} (1-2^{-\epsilon_1 R})^{|m-n|} 2^{|m-n|\alpha_i R} \prod_{i<j} (1-2^{-\epsilon_1 R})^2 2^{2\alpha_j R} \mathrm{d}\alpha, \qquad \text{(A.218)}\end{aligned}$$

where

$$\mathcal{A}_{\epsilon_1} \triangleq \{\alpha \in \mathcal{A}_2 | \frac{\log\rho}{R} - \alpha_{\min\{m,n\}} > \epsilon_1, \alpha_1 > \epsilon_1, |\alpha_j - \alpha_i| > \epsilon_1 \ \forall i \neq j\}.$$

Realizing that $e^{-2^{-\epsilon_1 R}} \geq (1-2^{-\epsilon_1 R})$, (A.218) yields

$$A_2(R,\rho,\epsilon) \geq KR^{\min\{m,n\}} e^{\frac{m\min\{m,n\}}{\rho}} (1-2^{-\epsilon_1 R})^{m(n+\min\{m,n\})} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{A}_{\epsilon_1}} 2^{f(\alpha)R} \mathrm{d}\alpha,$$

where, as before, $f(.)$ is given by (A.216). Let us define α^* as

$$\alpha^* \triangleq \arg \sup_{\alpha \in \mathcal{A}_{\epsilon_1}} f(\alpha).$$

Then it follows from the continuity of $f(.)$ that, for any $\epsilon_2 > 0$, there exists a neighborhood I_{ϵ_2} of α^*, within which

$$f(\alpha) \geq f(\alpha^*) - \epsilon_2.$$

This means that

$$\begin{aligned}A_2(R,\rho,\epsilon) \geq & KR^{\min\{m,n\}} e^{\frac{m\min\{m,n\}}{\rho}} (1-2^{-\epsilon_1 R})^{m(n+\min\{m,n\})} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{A}_{\epsilon_1}\cap I_{\epsilon_2}} 2^{f(\alpha)R} \mathrm{d}\alpha, \\ A_2(R,\rho,\epsilon) \geq & KR^{\min\{m,n\}} e^{\frac{m\min\{m,n\}}{\rho}} (1-2^{-\epsilon_1 R})^{m(n+\min\{m,n\})} \rho^{-mn} 2^{(f(\alpha^*)-c(k)-\epsilon_2)R} \times \\ & \mathrm{Vol}\{\mathcal{A}_{\epsilon_1} \cap I_{\epsilon_2}\}. \qquad \text{(A.219)}\end{aligned}$$

Now, from (A.217), (A.219) and the fact that $e^{-\frac{m(\min\{m,n\}-1)}{\rho}} \leq 1$, we conclude that

$$\frac{A_1(R,\rho,\epsilon)}{A_2(R,\rho,\epsilon)} \leq \big(1-(2^{\epsilon R})^{-\frac{\epsilon_1}{\epsilon}}\big)^{-m(n+\min\{m,n\})}(2^{\epsilon R})^{\frac{f_1-f(\alpha^*)+\epsilon_2}{\epsilon}} e^{-m(2^{\epsilon R})}\mathrm{Vol}^{-1}\{\mathcal{A}_{\epsilon_1}\cap I_{\epsilon_2}\}. \tag{A.220}$$

This means that

$$\limsup_{\substack{\rho\to\infty \\ R\in\mathcal{R}_\delta(k)}} \frac{\log\big(1+A_1(R,\rho,\epsilon)/A_2(R,\rho,\epsilon)\big)}{\log\rho} = 0 \tag{A.221}$$

Note that (A.221) holds, whether ρ growing to infinity and $R\in\mathcal{R}_\delta(k)$ result in R growing to infinity or not. This is because the right hand side of (A.220) decays *exponentially* with $2^{\epsilon R}$, while it only grows *polynomially* with the same variable. To characterize the second term on the right-hand side of (A.215), we note that

$$\begin{aligned} A_2(R,\rho,\epsilon) \leq & KR^{\min\{m,n\}}\rho^{-mn}2^{-c(k)R}\int_{\mathcal{A}_2} 2^{R\sum_i\alpha_i}\prod_{i=1}^{\min\{m,n\}} 2^{\alpha_i|m-n|R}\prod_{i<j}2^{2\alpha_j R}\mathrm{d}\alpha, \\ \leq & KR^{\min\{m,n\}}2^{(f_2-c(k))R}\rho^{-mn}. \end{aligned}$$

This means that

$$\frac{\log A_2(R,\rho,\epsilon)}{\log\rho} \leq \frac{\log(KR^{\min\{m,n\}})}{\log\rho} + (f_2-c(k))\frac{R}{\log\rho} - mn, \tag{A.222}$$

where

$$f_2 \triangleq \sup_{\mathcal{A}_2} f(\alpha).$$

To derive f_2, one needs to consider two different cases. The first case is when $R\in\mathcal{R}_\delta(0)$, in which case

$$f_2 = m+n-1, \quad R\in\mathcal{R}_\delta(0). \tag{A.223}$$

In this case, the supremum is achieved at $\alpha^* = (0, \cdots, 0, 1)$. The second case is when $R \in \mathcal{R}_\delta(k)$, for $k \in \mathbb{Z}$ and $\min\{m, n\} > k > 0$, where

$$f_2 = [m + n - (2k + 1)] + k(k + 1)(\frac{\log \rho}{R} + \epsilon), \quad R \in \mathcal{R}_\delta(k), \min\{m, n\} > k > 0. \tag{A.224}$$

The supremum happens at

$$\alpha^* = (0, \cdots, 1 - k(\frac{\log \rho}{R} + \epsilon), \underbrace{\frac{\log \rho}{R} + \epsilon, \cdots, \frac{\log \rho}{R} + \epsilon}_{k \text{ times}}).$$

Notice that, assuming $\epsilon \leq \delta$, (A.213) guarantees that $1 - k(\frac{\log \rho}{R} + \epsilon) > 0$. Plugging in for f_2 in (A.222), we conclude

$$\begin{aligned} \frac{\log A_2(R, \rho, \epsilon)}{\log \rho} \leq & \frac{\log(K R^{\min\{m,n\}})}{\log \rho} - g(k) + (mn - g(k)) \frac{R}{\log \rho} \epsilon + \\ & (\tilde{c}(k) - c(k)) \frac{R}{\log \rho}, \quad \text{for} \quad R \in \mathcal{R}_\delta(k). \end{aligned} \tag{A.225}$$

In this expression, $g(k)$ is given by (6.7) and $\tilde{c}(k)$ is defined as

$$\tilde{c}(k) \triangleq m + n - (2k + 1). \tag{A.226}$$

Now, from (A.215), (A.221) and (A.225), together with the fact that ϵ can be made arbitrarily small, one concludes

$$\begin{aligned} \limsup_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R, \rho) - c(k) R}{\log \rho} \leq & -g(k) + \\ & (\tilde{c}(k) - c(k)) \times \limsup_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{R}{\log \rho}. \end{aligned} \tag{A.227}$$

Next we turn our attention to (A.207) and introduce the following change of variables.

$$\beta_i \triangleq \frac{\log(1 + \rho \mu_i)}{R}. \tag{A.228}$$

This implies that $\beta_{\min\{m,n\}} \geq \cdots \geq \beta_1 \geq 0$. In terms of the new variables, (A.207) can be written as

$$P_o(R,\rho) \geq \Pr\{\mathcal{B}\}, \tag{A.229}$$

where

$$\mathcal{B} \triangleq \{\beta | \beta_{\min\{m,n\}} \geq \cdots \geq \beta_1 \geq 0, 1 - \sum_i \beta_i > 0\}. \tag{A.230}$$

In (A.230), $\beta \triangleq (\beta_1, \cdots, \beta_{\min\{m,n\}})$. On the other hand, (A.208) becomes

$$\begin{aligned} p(\beta) =& KR^{\min\{m,n\}} \rho^{-mn} 2^{R\sum_i \beta_i} \times \\ & \prod_{i=1}^{\min\{m,n\}} (2^{\beta_i R} - 1)^{|m-n|} \prod_{i<j} (2^{\beta_i R} - 2^{\beta_j R})^2 e^{-\sum_i \frac{2^{\beta_i R}-1}{\rho}}, \end{aligned} \tag{A.231}$$

where $K \triangleq K_{m,n}^{-1} (\ln 2)^{\min\{m,n\}}$. This means that

$$P_o(R,\rho) 2^{-c(k)R} \geq 2^{-c(k)R} \int_{\mathcal{B}} p(\beta) \mathrm{d}\beta, \quad R \in \mathcal{R}_\delta(k).$$

Thus

$$\begin{aligned} P_o(R,\rho) 2^{-c(k)R} \geq & KR^{\min\{m,n\}} e^{\frac{\min\{m,n\}}{\rho}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{B}} e^{-\sum_i 2^{-(\frac{\log\rho}{R} - \beta_i)R}} 2^{R\sum_i \beta_i} \times \\ & \prod_{i=1}^{\min\{m,n\}} (1 - 2^{-\beta_i R})^{|m-n|} 2^{|m-n|\beta_i R} \prod_{i<j} (1 - 2^{-(\beta_j - \beta_i)R})^2 2^{2\beta_j R} \mathrm{d}\beta, \\ \geq & KR^{\min\{m,n\}} e^{\frac{\min\{m,n\}}{\rho}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{B}_{\epsilon_1}} e^{-\min\{m,n\} 2^{-\epsilon_1 R}} 2^{R\sum_i \beta_i} \times \\ & \prod_{i=1}^{\min\{m,n\}} (1 - 2^{-\epsilon_1 R})^{|m-n|} 2^{|m-n|\beta_i R} \prod_{i<j} (1 - 2^{-\epsilon_1 R})^2 2^{2\beta_j R} \mathrm{d}\beta, \end{aligned} \tag{A.232}$$

where

$$\mathcal{B}_{\epsilon_1} \triangleq \{\beta \in \mathcal{B} | \frac{\log\rho}{R} - \epsilon_1 > \beta_{\min\{m,n\}}, \beta_1 > \epsilon_1, |\beta_j - \beta_i| > \epsilon_1 \ \forall i \neq j\}.$$

Realizing that $e^{-2^{-\epsilon_1 R}} \geq (1-2^{-\epsilon_1 R})$, (A.232) yields

$$P_o(R,\rho)2^{-c(k)R} \geq KR^{\min\{m,n\}} e^{\frac{\min\{m,n\}}{\rho}} (1-2^{-\epsilon_1 R})^{mn+\min\{m,n\}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{B}_{\epsilon_1}} 2^{f(\beta)R} \mathrm{d}\beta,$$

where, as before, $f(.)$ is given by (A.216). Let us define β^* as

$$\beta^* \triangleq \arg \sup_{\beta \in \mathcal{B}_{\epsilon_1}} f(\beta).$$

Again, it follows from the continuity of $f(.)$ that, for any $\epsilon_2 > 0$, there exists a neighborhood I_{ϵ_2} of β^*, within which

$$f(\beta) \geq f(\beta^*) - \epsilon_2.$$

This means that

$$P_o(R,\rho)2^{-c(k)} \geq KR^{\min\{m,n\}} e^{\frac{\min\{m,n\}}{\rho}} (1-2^{-\epsilon_1 R})^{mn+\min\{m,n\}} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{B}_{\epsilon_1} \cap I_{\epsilon_2}} 2^{f(\beta)R} \mathrm{d}\beta,$$

$$P_o(R,\rho)2^{-c(k)} \geq KR^{\min\{m,n\}} e^{\frac{\min\{m,n\}}{\rho}} (1-2^{-\epsilon_1 R})^{mn+\min\{m,n\}} \rho^{-mn} 2^{(f(\beta^*)-c(k)-\epsilon_2)R} \times$$
$$\mathrm{Vol}\{\mathcal{B}_{\epsilon_1} \cap I_{\epsilon_2}\}.$$

Thus

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho} \geq -mn + \liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} (f(\beta^*) - c(k) - \epsilon_2) \frac{R}{\log \rho}. \tag{A.233}$$

Since (A.233) is valid for arbitrarily small values of ϵ_1 and ϵ_2, we conclude

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho} \geq -g(k) + (\tilde{c}(k) - c(k)) \times \liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{R}{\log \rho}, \tag{A.234}$$

where we have used the fact that $\mathcal{B}_0 = \mathcal{A}_2|_{\epsilon=0}$, which means that $f(\beta^*)$ can be easily derived from (A.223) and (A.224) by simply plugging in $\epsilon = 0$. Examining (A.227) and (A.234) reveals that the choice

$$c(k) = \tilde{c}(k), \quad \forall k \tag{A.235}$$

guarantees the existence of

$$\lim_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho} = -g(k), \tag{A.236}$$

regardless of whether $\lim_{\rho\to\infty} \frac{R}{\log\rho}$ exits or not. Now, since (A.236) holds for arbitrarily small values of δ, we get (6.4). Note that (A.235), together with (A.226), result in (6.6) and thus complete the proof.

A.15 Proof of Theorem 17

The proof follows that of Theorem 2 in [17]. In particular, we prove (6.17) in two steps. The first step is to show that

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq -g(k). \tag{A.237}$$

Towards this end, let us denote the ML error probability, conditioned on a certain channel realization H, by $P_{E|H}$. It then follows that

$$\begin{aligned} P_e(R,\rho) &= \mathrm{E}_H\{P_{E|H}(R,\rho)\}, \\ &= \int P_{E|H}(R,\rho)p(H)\mathrm{d}H. \end{aligned}$$

Thus

$$P_e(R,\rho) \geq \int_{\mathcal{C}} P_{E|H}(R,\rho)p(H)\mathrm{d}H, \tag{A.238}$$

where $\mathcal{C}$ denotes any subset of the set of all channel realizations. Let us define $\mathcal{C}_\epsilon$ as the set of channel realizations for which the conditional ML error probability cannot be made smaller than ϵ, i.e.

$$\mathcal{C}_\epsilon \triangleq \{H | P_{E|H}(R,\rho) \geq \epsilon\}. \tag{A.239}$$

From (A.238) and (A.239) one concludes that

$$P_e(R,\rho) \geq \epsilon P_{\mathcal{C}_\epsilon}(R,\rho),$$

where

$$P_{\mathcal{C}_\epsilon}(R,\rho) = \Pr\{\mathcal{C}_\epsilon\}. \tag{A.240}$$

This means that

$$\frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq \frac{\log P_{\mathcal{C}_\epsilon}(R,\rho) - c(k)R}{\log \rho} + \frac{\log \epsilon}{\log \rho},$$

where $c(k)$ is given by (6.6), or

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq \liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_{\mathcal{C}_\epsilon}(R,\rho) - c(k)R}{\log \rho}, \tag{A.241}$$

where $\mathcal{R}_\delta(k)$ is given by (A.213). Now, application of Fano's inequality reveals that

$$P_{E|H}(R,\rho) \geq 1 - \frac{\mathrm{I}(\mathbf{x};\mathbf{y}|\mathbf{H}=H)}{R} - \frac{1}{Rl},$$

where l denotes the codeword length [17]. This, together with (A.239), means that

$$\{H | 1 - \frac{\mathrm{I}(\mathbf{x};\mathbf{y}|\mathbf{H}=H)}{R} - \frac{1}{Rl} \geq \epsilon\} \subseteq \mathcal{C}_\epsilon,$$

which using (A.205) results in

$$P_{\mathcal{C}_\epsilon}(R,\rho) \geq P_o((1-\epsilon-\frac{1}{Rl})R,\rho) \quad \text{or}$$

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_{\mathcal{C}_\epsilon}(R,\rho) - c(k)R}{\log \rho} \geq \liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}_\delta(k)}} \frac{\log P_o((1-\epsilon)R,\rho) - c(k)R}{\log \rho}. \tag{A.242}$$

Now, from (A.241) and (A.242), together with the fact that both, δ and ϵ can be made arbitrarily small, we conclude that

$$\liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq \liminf_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho}.$$

But, from Theorem 16 (refer to (6.4)), we know that the right-hand side equals $-g(k)$ (given by (6.7)). This proves (A.237) and thus completes the first step.

The second step in proving (6.17) is to show that

$$\limsup_{\substack{\rho \to \infty \\ R \in \mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \leq -g(k), \tag{A.243}$$

provided that the codeword length, l, satisfies $l \geq m+n-1$. To prove this, consider a Gaussian code-book with 2^{Rl} codewords of length l. It is straightforward to verify that the ML error probability, conditioned on a certain channel realization, is upper-bounded by

$$\begin{aligned} P_{E|H}(R,\rho) &\leq 2^{Rl} \det(I_n + \frac{\rho}{2m} HH^H)^{-l}, \\ &= 2^{Rl} \prod_{i=1}^{\min\{m,n\}} (1 + \frac{\rho}{2m}\mu_i)^{-l}, \end{aligned} \tag{A.244}$$

where $\mu_{\min\{m,n\}} \geq \cdots \geq \mu_1 \geq 0$ represent the ordered eigenvalues of HH^H. The joint PDF of $(\mu_1, \cdots, \mu_{\min\{m,n\}})$ is given by (A.208). The change of variables

$$\gamma_i \triangleq \frac{\log(1 + \frac{\rho}{2m}\mu_i)}{R},$$

changes (A.244) and (A.208), into

$$P_{E|\gamma}(R,\rho) \leq 2^{(1-\sum_{i=1}^{\min\{m,n\}} \gamma_i)Rl} \tag{A.245}$$

and

$$p(\gamma) = KR^{\min\{m,n\}}\rho^{-mn}2^{R\sum_i \gamma_i}\times \prod_{i=1}^{\min\{m,n\}}(2^{\gamma_i R}-1)^{|m-n|}\prod_{i<j}(2^{\gamma_i R}-2^{\gamma_j R})^2 e^{-\sum_i \frac{2m(2^{\gamma_i R}-1)}{\rho}},$$

where $\gamma \triangleq (\gamma_1, \cdots, \gamma_{\min\{m,n\}})$ and $K \triangleq K_{m,n}^{-1}(\ln 2)^{\min\{m,n\}}(2m)^{mn}$. Next we define $\mathcal{D}$ as

$$\mathcal{D} \triangleq \{\gamma | \gamma_{\min\{m,n\}} \geq \cdots \geq \gamma_i \geq 0, 1 - \sum_{i=1}^{\min\{m,n\}} \gamma_i \geq 0\}.$$

Referring to (A.245) reveals that $\mathcal{D}$ consists of those channel realizations for which the upper-bound on the ML error probability cannot be made arbitrarily small, even through the use of infinitely long codewords. For these channel realizations, we upper-bound $P_{E|\gamma}(R,\rho)$ by 1, i.e.

$$\begin{aligned} P_e(R,\rho) &= P_{E,\mathcal{D}^c}(R,\rho) + P_{E,\mathcal{D}}(R,\rho), \\ P_e(R,\rho) &\leq P_{E,\mathcal{D}^c}(R,\rho) + P_{\mathcal{D}}(R,\rho), \end{aligned} \tag{A.246}$$

where $\mathcal{D}^c$ denotes the complement of $\mathcal{D}$. Let us first focus on $P_{\mathcal{D}}(R,\rho)$. Realizing that $\mathcal{D}$ is precisely the same set as $\mathcal{A}$ (refer to (A.211)) and that, up to a scaling factor, $p(\gamma)$ is identical to $p(\alpha)$ (refer to (A.212)), it follows immediately that

$$\limsup_{\substack{\rho\to\infty \\ R\in\mathcal{R}(k)}} \frac{\log P_{\mathcal{D}}(R,\rho) - c(k)R}{\log \rho} \leq -g(k). \tag{A.247}$$

Now, turning our attention back to $P_{E,\mathcal{D}^c}(R,\rho)$, we realize that

$$\lim_{\substack{\rho\to\infty \\ R\in\mathcal{R}(k)}} P_{E,\mathcal{D}_1^c}(R,\rho)2^{-c(k)R} = 0, \tag{A.248}$$

which means that

$$\limsup_{\substack{\rho\to\infty \\ R\in\mathcal{R}(k)}} \frac{\log P_{E,\mathcal{D}^c}(R,\rho)-c(k)R}{\log\rho} = \limsup_{\substack{\rho\to\infty \\ R\in\mathcal{R}(k)}} \frac{\log P_{E,\mathcal{D}_2^c}(R,\rho)-c(k)R}{\log\rho}, \tag{A.249}$$

where

$$\mathcal{D}_1^c \triangleq \{\gamma\notin\mathcal{D}|\gamma_{\min\{m,n\}}>\frac{\log\rho}{R}\} \text{ and } \mathcal{D}_2^c \triangleq \{\gamma\notin\mathcal{D}|\gamma_{\min\{m,n\}}\le\frac{\log\rho}{R}\}.$$

Notice that (A.248) holds for exactly the same reason as (A.221) does (refer to the comment after (A.221)). Now, using (A.245), we have

$$\begin{aligned} P_{E,\mathcal{D}_2^c}(R,\rho)2^{-c(k)R} &= 2^{-c(k)R}\int_{\mathcal{D}_2^c} P_{E|\gamma}(R,\rho)p(\gamma)\mathrm{d}\gamma, \\ &\le KR^{\min\{m,n\}}\rho^{-mn}2^{-c(k)R}\int_{\mathcal{D}_2^c} 2^{[f(\gamma)+l(1-\sum_{i=1}^{\min\{m,n\}}\gamma_i)]R}\mathrm{d}\gamma, \end{aligned}$$

where

$$f(\gamma)\triangleq 2^{\sum_i(|m-n|+2i-1)\gamma_i R}.$$

Thus

$$P_{E,\mathcal{D}_2^c}(R,\rho)2^{-c(k)R} \le KR^{\min\{m,n\}}\rho^{-mn}2^{(f_2-c(k))R}\mathrm{Vol}\{\mathcal{D}_2^c\}, \tag{A.250}$$

where

$$f_2 \triangleq \sup_{\mathcal{D}_2^c} f(\gamma)+l(1-\sum_{i=1}^{\min\{m,n\}}\gamma_i).$$

Realizing that for $l\ge m+n-1$, the supremum occurs at $\gamma=\gamma^*$, such that $1-\sum\gamma_i^*=0$, f_2 can be easily derived from (A.223) and (A.224) by simply plugging in $\epsilon=0$. Therefore, (A.250) gives

$$\limsup_{\substack{\rho\to\infty \\ R\in\mathcal{R}(k)}} \frac{\log P_{E,\mathcal{D}_2^c}(R,\rho)-c(k)R}{\log\rho} \le -g(k). \tag{A.251}$$

Now, from (A.251), (A.249) and (A.247), we conclude (A.243), which together with (A.237), proves (6.17). Since we proved (6.17) using an *ensemble* of Gaussian codes, it follows that, for any code-length $l \geq m+n-1$, there exists at least a code, for which (6.17) holds. This completes the proof.

A.16 Proof of Theorem 18

Realizing that the V-BLAST protocol essentially transforms the $m \times m$ MIMO channel into a multiple-access channel with m single-antenna users and a destination with m receive antennas, we prove (6.18) by following the same lines as that of Theorem 2 in [18]. In particular, let E_S denote the event that a certain decoder makes errors in decoding the codewords transmitted by a subset S of the antennas. It then follows that

$$\sum_{S \neq \varnothing} \Pr\{E_S\} \geq P_e(R, \rho). \tag{A.252}$$

It is also clear that

$$P_e(R, \rho) \geq \Pr\{E_{S^*}\}, \tag{A.253}$$

where S^* denotes any non-empty subset of $\{1, \cdots, m\}$. Let us define S^* as the non-empty subset of $\{1, \cdots, m\}$, such that for all other non-empty subsets S, we have

$$\liminf_{\substack{\rho \to \infty \\ \frac{|S|}{m} R \in \mathcal{R}_S(k) \\ \frac{|S^*|}{m} R \in \mathcal{R}_{S^*}(k^*)}} \Big(g_S(k) - g_{S^*}(k^*) \Big) - \Big(\frac{|S|}{m} c_S(k) - \frac{|S^*|}{m} c_{S^*}(k^*) \Big) \frac{R}{\log \rho} \geq 0. \tag{A.254}$$

In (A.254), $|S|$ denotes the cardinality of set S. Also, $\mathcal{R}_S(k)$, $c_S(k)$ and $g_S(k)$ denote $\mathcal{R}(k)$, $c(k)$ and $g(k)$, as defined by (6.5), (6.6) and (6.7), for a MIMO channel with

$|S|$ transmit and m receive antennas. The proof of (6.18) then follows in three steps. First, we prove that for any decoder

$$\liminf_{\substack{\rho\to\infty \\ \frac{|S^*|}{m}R\in\mathcal{R}_{S^*}(k^*)}} \frac{P_e(R,\rho)-\frac{|S^*|}{m}c_{S^*}(k)R}{\log\rho} \geq -g_{S^*}(k^*). \tag{A.255}$$

This follows immediately from (A.253) and the fact that $\Pr\{E_{S^*}\}$ upper-bounds the ML error probability of a MIMO channel with $|S^*|$ transmit antennas, m receive antennas and rate $\frac{|S^*|}{m}R$. The second step in proving (6.18) is to show that there exists a code, along with a decoder, for which

$$\limsup_{\substack{\rho\to\infty \\ \frac{|S^*|}{m}R\in\mathcal{R}_{S^*}(k^*)}} \frac{P_e(R,\rho)-\frac{|S^*|}{m}c_{S^*}(k^*)R}{\log\rho} \leq -g_{S^*}(k^*). \tag{A.256}$$

We prove this by showing that the error probability of the joint ML decoder, averaged over the ensemble of Gaussian codes, satisfies (A.256). The existence of the desired code then follows from the fact that there exist codes in the ensemble that perform at least as well as the average. For this purpose, assume that each of the antennas uses a Gaussian code-book of code-word length $l=2m+1$ and size $2^{\frac{R}{m}l}$ codewords. It then follows from Theorem 17, (refer to (6.17)), that

$$\lim_{\substack{\rho\to\infty \\ \frac{|S|}{m}R\in\mathcal{R}_S(k)}} \frac{\log\Pr\{E_S\}-\frac{|S|}{m}c_S(k)R}{\log\rho} = -g_S(k), \quad \forall S\neq\varnothing, \tag{A.257}$$

which means

$$\lim_{\substack{\rho\to\infty \\ \frac{|S|}{m}R\in\mathcal{R}_S(k) \\ \frac{|S^*|}{m}R\in\mathcal{R}_{S^*}(k^*)}} \left[\frac{\log\Pr\{E_S\}/\Pr\{E_{S^*}\}}{\log\rho}+\Big(g_S(k)-g_{S^*}(k^*)\Big)- \Big(\frac{|S|}{m}c_S(k)-\frac{|S^*|}{m}c_{S^*}(k^*)\Big)\frac{R}{\log\rho}\right]=0. \tag{A.258}$$

Now, (A.258), together with (A.254), results in

$$\limsup_{\substack{\rho \to \infty \\ \frac{|S|}{m} R \in \mathcal{R}_S(k) \\ \frac{|S^*|}{m} R \in \mathcal{R}_{S^*}(k^*)}} \frac{\log \Pr\{E_S\}/\Pr\{E_{S^*}\}}{\log \rho} \leq 0, \quad \forall S \neq \varnothing. \tag{A.259}$$

Returning to (A.252), we have

$$\frac{\log(1 + \sum_{S \neq S^*, \varnothing} \frac{\Pr\{E_S\}}{\Pr\{E_{S^*}\}})}{\log \rho} + \frac{\log \Pr\{E_{S^*}\} - \frac{|S^*|}{m} c_{S^*}(k^*) R}{\log \rho} \geq \frac{\log P_e(R, \rho) - \frac{|S^*|}{m} c_{S^*}(k^*) R}{\log \rho}.$$

Taking the lim sup of both sides, together with (A.259) and (A.257) results in (A.256). Notice that (A.255) and (A.256) mean that

$$\lim_{\substack{\rho \to \infty \\ \frac{|S^*|}{m} R \in \mathcal{R}_{S^*}(k^*)}} \frac{P_e(R, \rho) - \frac{|S^*|}{m} c_{S^*}(k) R}{\log \rho} = -g_{S^*}(k^*). \tag{A.260}$$

The third and last step in proving (6.18) is to show that $|S^*| = 1$. One can prove this directly using the definitions of $\mathcal{R}(k)$, $c(k)$ and $g(k)$ (equations (6.5), (6.6) and (6.7)). However, we choose to do this using observations (6.8) and (6.9). In particular, notice that based on these observations, (A.254) reduces to finding the subset S^*, for which

$$d_S(\frac{|S|}{m} r) - d_{S^*}(\frac{|S^*|}{m} r) \geq 0, \quad \forall S \neq \varnothing,$$

where $d_S(r)$ represents the DMT for a MIMO system with $|S|$ transmit and m receive antennas. From the proof of Theorem 3 in [18], we know that $|S^*| = 1$. Now, this together with (A.260) results in (6.18) and thus completes the proof.

BIBLIOGRAPHY

[1] A. Sendonaris, E. Erkip and B. Aazhang, "User Cooperation Diversity. Part I. System Description," *IEEE Transactions on Communications*, page(s): 1927- 1938 Volume: 51, Issue: 11, Nov. 2003

[2] A. Sendonaris, E. Erkip and B. Aazhang, "User Cooperation Diversity. Part II. Implementation Aspects and Performance Analysis," *IEEE Transactions on Communications*, page(s): 1939-1948 Volume: 51, Issue: 11, Nov. 2003

[3] J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Transactions on Information Theory*, 2002, submitted.

[4] J. N. Laneman and G. W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Transactions on Information Theory*, Volume: 49 , Issue: 10, Oct. 2003 Pages: 2415 - 2425

[5] M. Janani, A. Hedayat, T. Hunter and A. Nosratinia, "Coded Cooperation in Wireless Communications: Space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing*, Volume: 52, Issue: 2 , Feb. 2004 Pages:362-371

[6] A. Stefanov and E. Erkip, "Cooperative Coding for Wireless Networks," *IEEE Transactions on Communications*, accepted for publication

[7] R. U. Nabar, H. Bolcskei, F. W. Kneubuhler, "Fading Relay Channels: Performance Limits and Space-Time Signal Design," *IEEE Journal on Selected Areas in Communications*, Volume: 22, Issue: 6, Aug. 2004 Pages: 1099-1109

[8] N. Prasad, M. K. Varanasi, "Diversity and Multiplexing Tradeoff Bounds for Cooperative Diversity Protocols," *Proc. International Symposium on Information Theory,* June 2004, Page: 268.

[9] G. Kramer, M. Gastpar P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *submitted to the IEEE Transactions on Information Theory*, Feb. 2004

[10] A. Host-Madsen, "On the Capacity of Wireless Relaying," *Proc. 2002 IEEE 56th Vehicular Technology Conference,* Volume: 3, Sep. 2002, Pages: 1333-1337.

[11] A. Host-Madsen, "On the Capacity of Cooperative Diversity in Slow Fading Channels, " *Proc. Allerton Conference on Communications, Control and Computing*, Monticello, IL, Oct. 2002.

[12] A. Host-Madsen, "A New Achievable Rate for Cooperative Diversity Based on Generalized Writing on Dirty Paper," *Proc. International Symposyuom on Information Theory*, Yokohama, Japan, 2003.

[13] M. A. Khojastepour, A. Sabharwal, B. Aazhang, "Lower Bounds on the Capacity of Gaussian Relay Channel," *Proc. 38th Annual Conference on Information Sciences and Systems*, Princeton NJ, March 2004.

[14] M. A. Khojastepour, A. Sabharwal, B. Aazhang, "On the Capacity of Gaussian Cheap Relay Channel," *Proc IEEE 2003 Global Communications Conference*, San Francisco, CA Dec. 2003.

[15] U. Mitra and A. Sabharwal, "On Achievable Rates for Complexity Constrained Relay Channels," *Proc 41st Allerton Conference on Communication, Control and Computing*, Monticello, IL, Oct. 2003.

[16] S. Zahedi, M. Mohseni and A. El Gamal, "On the Capacity of AWGN Relay Channel with Linear Relaying Functions," *IEEE International Symposyum on Information Theory*, Chicago, IL, June 2004.

[17] L. Zheng and D. N. C. Tse., "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna Channels," *IEEE Trans. Info. Theory*, 49:1073-1096, May 2003.

[18] D. N. C. Tse, P. Viswanath and L. Zheng, "Diversity-Multiplexing Tradeoff in Multiple Access Channels," *IEEE Trans. on Info. Theory*, 2003, submitted.

[19] K. Azarian Yazdi, H. El Gamal, and P. Schniter, "On the Design of Cooperative Transmission Schemes," *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Oct. 2003.

[20] K. Azarian, H. El Gamal, and P. Schniter, "On the Achievable Diversity-vs-Multiplexing Tradeoff in Cooperative Channels," *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2004.

[21] H. El Gamal, "The Diversity-Multiplexing Tradeoff in Half-Duplex Cooperative Channels: Achievable Curves and Optimal Strategies," *LIDS Colloquia*, EECS Dept., MIT, May 2004.

[22] K. Azarian, H. El Gamal, and P. Schniter, "On the Achievable Diversity-Multiplexing Tradeoff in Half Duplex Cooperative Channels," *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Oct. 2004.

[23] K. Azarian, H. El Gamal, and P. Schniter, "Achievable Diversity-vs-Multiplexing Tradeoffs in Half-Duplex Cooperative Channels," *Proc. IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004.

[24] T. M. Cover and A. A.El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Transactions on Information Theory*, Volume: 25, pp. 572-584, Sept. 1979.

[25] T. M. Cover, J. A. Thomas, "*Elements of Information Theory*," *John Wiley and Sons*, 1991.

[26] G. Foschini, G. Golden, R. Valenzuela and P. Wolniansky, "Simplified Processing for High Spectral Efficiency Wireless Communication Employing Multi-Element Arrayas," *IEEE Jour. Select. Areas on Comm.*, 17:1841-1852, Nov. 1999.

[27] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. Info. Theory*, 45:1456-1467, July 1999.

[28] S. Alamouti, "A Simple Transmitter Diversity Scheme for Wireless Communications," *IEEE Jour. Select. Areas on Comm.*, 16:1451-1458, Oct. 1998.

[29] H. El Gamal, G. Caire and M. O. Damen, "The MIMO ARQ Channel: Diversity-Multiplexing-Delay Tradeoff," *Submitted to the IEEE Trans. Info. Theory.*

[30] S. Yang and J. C. Belfiore, "Optimal Space-Time Codes for the MIMO Amplify and Forward Cooperative Channel," *submitted to IEEE Trans. Info. Theory*, Sep. 2005.

[31] A. Murugan, K. Azarian and H. El Gamal, "Cooperative Lattice Coding and Decoding," *IEEE JSAC Special Issue on Cooperative Communications and Networking*, submitted, Feb. 2006.