Asymptotic Analysis of Wireless Systems with Rayleigh Fading

Thesis by

Chaitanya Kumar Rao

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended 7 March 2007)

Acknowledgements

I am most grateful to my PhD advisor Babak Hassibi for his involvement in making this thesis possible. I have had many good experiences at Caltech, but one of the things I will miss the most is seeing his technical brilliance in action, whether it is his ability to comprehend and solve problems, or to explain concepts with such clarity. He has also been very supportive, approachable, and understanding to me over the years. I feel honored and privileged to have worked with him.

In addition, much of this thesis would not have come about without discussions with the other high-achieving past and present members of my research group. I am very fortunate to have encountered such a good-natured set of people. I would especially like to acknowledge Amir F. Dana, Maralle Fakhereddin, Radhika Gowaikar, Yindi Jing, Frédérique Oggier, Masoud Sharif, Mihailo Stojnic, and Haris Vikalo, whose ideas and advice have directly impacted this work. Finally, I would like to extend my heartfelt gratitude to my close friends and loving family for their strong moral support during my PhD.

Abstract

With wireless systems becoming increasingly complex and demanding, there is great importance in understanding ways of improving the reliability of transmitting information over the wireless medium. Rayleigh fading is a phenomenon that degrades the performance of many wireless systems. This thesis looks at ways to improve either the reliability or the rate at which we can successfully transmit information over Rayleigh-fading channels.

We study four wireless schemes, the first in the low signal-to-noise ratio (SNR) regime, the remaining three at high SNR. The analysis provides insights that can be applied to more general SNRs. At the low SNR extreme we are interested in maximizing capacity, while at high SNR we are concerned with the rate of decay of error probability with SNR. All wireless channels involved are modeled by independent Rayleigh fading and additive Gaussian noise for simpler analysis.

Firstly we investigate a point-to-point multiple antenna link at low SNR. At low SNR channel estimates can be unreliable, and therefore we assume the channel is unknown to both transmitter and receiver. We adopt a block-fading model and find the mutual information between transmitter and receiver up to second order in the SNR. This depends on the number of antennas in the system and the coherence interval of the channel. The expression is valid for input distributions with regular behavior of fourth- and sixth-order moments, in particular most practical schemes. This assumption is weaker than that of other works,

which require finite higher-order moments. In addition the approach used may be applied to other similar channel models. Subject to fourth-order moment and singular-value peak constraints, we determine the optimal signaling to maximize this mutual information.

We undertake high SNR analysis by finding the diversity-multiplexing gain trade-off of three further wireless systems with fading. This is a recent method for measuring the relationship between rate and reliability. Using techniques from existing works we find the optimal diversity-multiplexing gain trade-off for an $(M \times N)$ multiple antenna system with R single antenna relays. This uses a two-stage protocol in which the source first transmits to relays, then the relays multiply their received signal by a unitary matrix, before forwarding the result to the multiple-antenna receiver. The optimal trade-off is found to be equal to that of a multiple-input multiple-output (MIMO) link with R transmit and $\min\{M, N\}$ receive antennas.

Next we look at systems with interference, introduced by more than one disjoint source-destination pair competing for the wireless medium. We consider a four-node network with two source-destination pairs (an interference channel) and establish relationships between the rate and diversity achievable by certain schemes. With a view to increasing diversity, we then show through two more schemes how cooperation amongst the nodes achieves this, but at the cost of a reduced rate of the system. Through outage probability calculations, we find one scheme that increases diversity by a factor of three but reduces rate by a factor of four. Another scheme increases diversity by a factor of two but reduces rate by a factor of three. These schemes can easily be generalized from two to m source-destination pairs.

A final scheme is considered where n relay nodes are added to the m source-destination pairs, which act to cancel interference in an aim to increase diversity. The outage behavior of this scheme is analyzed and it is shown that a maximum diversity of $n - m^2 + 2$ can be obtained. This result can be extended to the case $n < m^2$, but a higher rate penalty is incurred in this case.

Contents

Notation and Abbreviations

A^*	Conjugate transpose of A
$\bar{A}$	Complex conjugate of A
A^T	Transpose of A
$:=$	Defined to be equal to
$\|\|\cdot\|\|$	Frobenius or Euclidean (l^2) norm of a vector or matrix: $\|\|A\|\|^2 = \mathrm{tr} AA^*$
$\mathbb{C}$	The set of complex numbers
$x \sim \mathbb{C}\mathcal{N}(0,1)$	x is a Complex Gaussian random variable with zero mean and unit variance. Its probability density function is given by $p(x) = e^{-\|x\|^2}/\pi$ for $x \in \mathbb{C}$. That is, a variable of the form $a + jb$ where $a, b \sim \mathcal{N}(0, 1/2)$.
cdf	Cumulative distribution function
δ_{ij}	Kronecker delta: 1 if $i = j$, 0 otherwise
$\mathrm{diag}(a_1, \ldots, a_n)$	The $n \times n$ matrix with diagonal entries $a_1, \ldots, a_n$, zeros elsewhere
$\mathbf{E}$	Expectation
$\exp(x)$	The exponential function e^x
$h(X)$	Differential entropy of the random variable X: $-\mathbf{E} p(X) \log p(X)$
$h(X\|Y)$	Conditional differential entropy of X given Y
I_n	$n \times n$ identity matrix (often used without subscript)
$I(X;Y)$	The mutual information between X and Y: $h(X) - h(X\|Y)$
i.i.d.	Independently and identically distributed

MAC	Multiple access channel
$(x)^+$	$\max\{x,0\}$, the larger of x and 0
MIMO	Multiple-input multiple-output, a multiple antenna system
P	Signal-to-noise ratio (high regime)
Pr	Probability
pdf	Probability density function
$p(x)$	Pdf evaluated at x
$p(x\|y)$	Conditional probability density function
QAM	Quadrature amplitude modulation
ρ	Signal-to-noise ratio (low regime)
$\sigma_{\min}(A)$	The smallest singular value of matrix A
SNR	Signal-to-noise ratio
$\sum_{i<j}$, $\prod_{i<j}$	Shorthand for the double sum or double product $\sum_j \sum_{i<j}$ or $\prod_j \prod_{i<j}$
$\mathrm{tr}(A)$	Trace: the sum of the eigenvalues or diagonal elements of a square matrix A
$\mathrm{vec}(A)$	The vector formed by stacking the columns of A
w.p.	With probability

Chapter 1

Introduction

1.1 Wireless Communications

Wireless technologies have experienced tremendous growth in recent times. While broadcast services such as radio, television and satellite communication have been present for decades, most of the recent development has occurred in applications where the information is more individualized. A large fraction of the developed world now has access to a cellular phone, which can transmit not only voice but also text, images and video. Computers can connect wirelessly to the internet and to other devices such as keyboards and printers. Also, sensor networks can seamlessly measure, process and distribute information about their local surroundings. Such applications are driven by our need to automate and to communicate on the move. We are becoming more reliant on them than ever before.

The wireless medium is more challenging to communicate over than wired media, even when there is a sole transmitter. Unlike in wires, electromagnetic signals weaken significantly as they propagate through three-dimensional space. Furthermore they can scatter or diffract in unpredictable fashion off physical objects and be corrupted by background noise.

Add to this the scenario of many mobile users transmitting simultaneously. Now interference is an issue, as signals intended for one user combine with those intended for another.

The motion of the users can further limit the speed at which data can be sent reliably. It is easy to take for granted these effects as our demand for higher data rates and better quality of service grows. The highly successful technologies we have today come about due to a variety of reasons including:

- advances in antenna, battery and radio frequency circuit design, and the miniaturization of devices;
- adoption of efficient wireless protocols and standards that control traffic flow, deciding what device transmits at what time and over what frequency;
- improved signal processing, modulation, coding, error correction and security methods.

A wireless network can be viewed as a collection of devices (nodes) in a region that communicate with each other via electromagnetic signals. The link between any two of these is a communication channel. Generally some nodes will be providers (sources) of information and some will want that information (destinations). Some nodes (known as relays) may be there to facilitate communication between sources and destinations, by amplifying signals or removing noise from them. The relays do not introduce any new information into the system. A node may play all three of these roles at some point. Some nodes may be active all the time, others very rarely.

In this thesis we restrict ourselves to systems with a number (m) of source-destination pairs (with or without additional relays). That is, there are m source nodes each with their own information to send and each with a single intended recipient, so there are m total destination nodes. There is no information that more that one destination needs, although sharing of information is allowed if that benefits the system as a whole. We call a system with $m = 1$ source-destination pair a *point-to-point* system.

There is highly active research in the theory of wireless communication [20, 68, 61], in an attempt to gain a better understanding of the random nature of the medium and how to communicate effectively over it. By adopting simple but realistic models, tools from information and communication theory can be used to set upper limits on data rates that can be achieved, as well as to predict reliability. In this work we are concerned with both of these issues in investigating several wireless communication problems. Communication theory has proved to be very successful in designing today's practical high-performance systems. Nevertheless many open problems exist, especially in the network setting, where it is still unclear how much better current systems can perform.

Before introducing the problems considered in the thesis, we will provide some of the fundamentals.

1.2 Communication Theory Preliminaries

The basic communications problem is to have information conveyed successfully across a *channel* that links a source (or transmitter) and a destination (or receiver). We will assume the same frequency is used for all transmissions. In this case we can represent the information as a sequence of complex numbers, which capture the magnitude and phase of the envelope of the electromagnetic signals [47]. An *envelope* can be thought of as the low-frequency waveform that bounds the high-frequency signals. Each complex number corresponds to a *symbol*, or group of bits of information. We will equate the time duration of a symbol transmission to one channel use, thus adopting a model discrete in time.

The set of symbols used for transmission constitute a *code*. This may be viewed, for example, as a collection of discrete points in the complex plane, but they need not be regularly spaced.

In this thesis we study Rayleigh fading channels with additive Gaussian noise. This means the channel alters the transmitted signal (s) by multiplying it by a random complex number (h, drawn from a particular distribution), and then adding another random complex number v to it, forming the received signal x:

$$x = hs + v. \tag{1.1}$$

The noise term v can arise from other electromagnetic signals unrelated to x. It is assumed to be a Gaussian random variable by the central limit theorem [21], since it is considered as the sum of a large number of identically distributed random variables.

The transmission is successful if the receiver can accurately determine s from x. Otherwise an error in transmission has occurred. The receiver has knowledge of the code and statistics of h and v, but not necessarily their values. If h is known to the receiver, it can decode by finding that symbol s which minimizes $|x - hs|$. For further explanation and an introduction to digital communications, see [47].

We are interested in analyzing particular wireless schemes, and measure performance by either *capacity* or *probability of error*. Since the communication channel we study is random in nature, error events are described by their probability. Capacity is the maximum rate at which information can be transmitted reliably, that is, with vanishing probability of error. Rate here refers to the number of bits of information that one can transmit per symbol, or channel use. It is related to the base 2 logarithm of the size of the code. Capacity is a function of the system's *signal-to-noise ratio* (SNR), which is defined as the average power of the received signal divided by the average noise power at the receiver.

If the average noise power is normalized to one, the SNR can be regarded as the power at which one is allowed to transmit. The greater the power, the further apart symbols can be spaced in the signal space, and the more reliable the system can be, leading to a higher capacity. The notion of capacity was introduced by Shannon in 1948 [58], and for the basic single transmitter and receiver case, was shown to be related to the maximum mutual information between sender and receiver. This is a quantity which describes how much information about the input is inferred from the output (or vice versa) and is related to the difference of two entropy functions, each being the expected logarithm of probability density functions of x and s [10, 39]:

$$I(x;s) = h(x) - h(x|s) = -\mathbf{E}\log p(x) + \mathbf{E}\log p(x|s).$$

This result will be used in chapter 2.

1.3 Rayleigh Fading and MIMO Communications

Fading is a phenomenon in wireless transmission whereby a signal has traveled along many paths and combines constructively or destructively at the receiver. As a result the net signal received varies in amplitude and phase over time in an unpredictable manner. The number of symbols over which this variation is considered minimal is known as the *coherence interval*. The received signal also varies over distance, as a second receiver placed a short distance away may experience a different sum of paths.

Suppose a message signal is modulated at some carrier frequency (i.e., multiplied by a sinusoidal signal) and transmitted over a channel that exhibits multipath fading. If there is no direct line of sight between the transmitter and receiver, the phases of the paths can be assumed to be uniformly distributed at the receiver and the magnitudes of the paths are independent. The received signal will be bounded by some envelope whose magnitude and phase can be represented by a complex number. By applying the central limit theorem, it can be shown [60] that the real and imaginary parts of this complex number tend to independent zero-mean Gaussian random variables each having the same variance, as the number of paths tends to infinity. This leads to the random number h in (1.1) representing the *fade* of the communication system. Normalizing h to have unit variance, h is drawn from the complex Gaussian distribution $\mathbb{C}\mathcal{N}(0, 1)$ and its probability density function is given by

$$p(h) = \frac{e^{-|h|^2}}{\pi}, \quad h \in \mathbb{C}. \tag{1.2}$$

The magnitude of h has a Rayleigh probability distribution [21], hence the name given to this type of multipath fading without direct line-of-sight transmission. This model best applies to indoor or urban environments, where signal attenuation due to distance is less significant. We assume h is fixed during a coherence interval before changing to a new independent value.

Also note that the distribution of $x = |h|^2$ is exponential:

$$p(x) = e^{-x}, \quad x > 0.$$

This property makes analysis of Rayleigh fading channels attractive, and so this model has been studied extensively. The review paper [5] describes much of the communication and information theory related to fading channels.

The type of fading described is known as a form of *flat fading* since the previous analysis assumes that the channel gain response is constant over the bandwidth of the signal. The signals arriving from multiple paths are not delayed (relative to a symbol period) in time. Since the bandwidth of the signal is narrow compared with that over which the channel behaves the same, flat fading systems are also known as *narrowband* systems.

Rayleigh fading severely degrades the error probability performance of wireless systems. This is because there is a significant probability that h in (1.1) is small in magnitude (known as an outage), resulting in the failure of two faded symbols hs_1 and hs_2 to be successfully distinguished when corrupted by noise. While error probability for a Gaussian channel without fading decays exponentially in the SNR, it decays only inversely in the SNR for a Gaussian channel with Rayleigh fading. This is shown in figure 1.1, where we plot error probability as a function of SNR with and without fading. This is done by simulating Equation (1.1) with v drawn from a $\mathbb{C}\mathcal{N}(0,1)$ distribution and setting s to $+\sqrt{\mathrm{SNR}}$ or $-\sqrt{\mathrm{SNR}}$ as SNR varies from 1 to 100. In the no fading case, we set h to 1; in the fading case h is drawn from a $\mathbb{C}\mathcal{N}(0,1)$ distribution. With this scheme an error occurs if $|x - hs| > |x + hs|$.

Since the scales of both axes are logarithmic, an inverse relationship is revealed by a line of slope –1. In figure 1.1 we see how much a system without fading outperforms one with fading: at 10 dB for example a system with fading has an error probability more than 0.02, while it is less than 0.0001 without fading. Similar effects are seen for more complicated coding strategies.

One way in which this fading phenomenon can be combated is by the use of multiple antennas at the transmitter and receiver. As mentioned at the beginning of this section, the strength of a received signal varies over distance, and if antennas are appropriately spaced, independent fades can be realized. Instead of a scalar the channel is now described by an $M \times N$ matrix H having (i,j)th entry h_{ij}. We assume independent fades between each

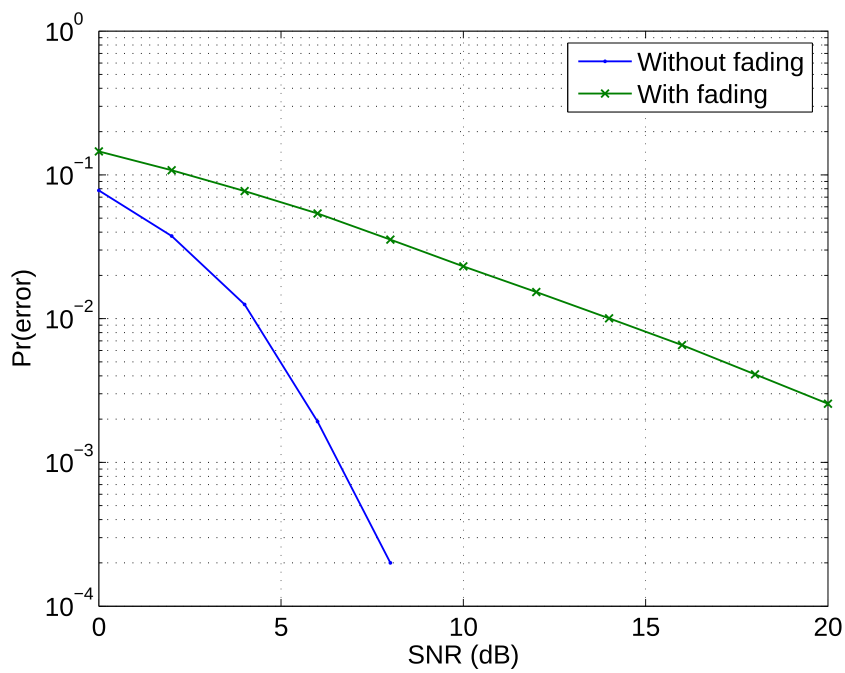

Figure 1.1: Error probability vs SNR for the system described by (1.1), with and without fading.

transmit-receive antenna pair so that from (1.2) the probability density function of H is given by

$$p(H) = \prod_{i=1}^{M}\prod_{j=1}^{N} \frac{e^{-|h_{ij}|^2}}{\pi} = \frac{e^{-\sum_{i=1}^{M}\sum_{j=1}^{N}|h_{ij}|^2}}{\pi^{MN}} = \frac{e^{-\mathrm{tr}HH^*}}{\pi^{MN}}.$$

Since the probability that all MN of these fades are bad is small, the reliability of the system can be improved, without having to boost the power. In the late 1990s Foschini [16] and Telatar [66] showed how such multiple antenna systems, also known as MIMO (multiple-input, multiple-output) systems, take advantage of the fading process to deliver higher data rates. Independent fading is now seen as a benefit since it can provide independent copies of the same signal. Codes which apply to these matrix channels are known as *space-time codes* [64], since numbers are now assigned to different antennas (space) over different times. Decoding such codes requires more complexity, but is now possible in practice due to advances in signal-processing chips [20].

In the analysis of many wireless communication problems, a solution is intractable in the general SNR case. This is the case with the problems we consider in this thesis. In order to gain some insight we consider asymptotic cases of SNR, where it either tends to 0 (low SNR regime) or to infinity (high SNR regime). For example in the high SNR regime, for a Rayleigh fading MIMO system with M transmit and N receive antennas, there exist coding strategies which ensure an error probability which decays as SNR^{-MN} as $\mathrm{SNR} \to \infty$ [64]. A result such as this tells us how reliability greatly benefits from additional antennas and can be used to design systems which operate well at moderate SNR levels.

All asymptotic analysis performed in this thesis is in the limit of SNR, where effectively, first order behavior is evaluated and lower order effects can be neglected. Other parameters, such as the number of antennas or users in the system, are assumed to be fixed at finite values.

1.4 Thesis Outline and Contributions

In this thesis we look at four wireless setups, the first in the low SNR regime, the remaining three at high SNR [49, 50, 51, 52, 53, 54]. At the low SNR extreme we are interested in how capacity scales with SNR and optimal signaling strategies. At high SNR we are concerned with the rate of decay of error probability with SNR. All wireless channels involved are independent with Rayleigh fading. The models do not depend on the distances between nodes, an assumption that applies best to environments with rich signal scattering [61]. Much of the notation used is summarized at the beginning of the thesis.

One of the main conclusions of this work is that knowledge of the channel by the receiver greatly enhances the performance that one can achieve from a wireless system in a Rayleigh fading environment, even if the transmitter does not know the channel.

1.4.1 MIMO Link at Low SNR

In chapter 2 we study a point-to-point multiple antenna link at low SNR (see figure 1.2). Such systems are used in energy efficient devices such as sensor networks, so their study is warranted. In this regime, it is reasonable to assume the channel is unknown at both the transmitter and receiver, since noise can adversely affect estimates of the fading matrix H.

With this assumption we adopt the block-fading model of Marzetta and Hochwald [38] and find the mutual information between transmitter and receiver up to second order in the SNR. This is a function of the number of antennas in the system and the coherence interval of the channel. The expression is valid for input distributions with regular behavior of fourth- and sixth-order moments, in particular most practical schemes. This assumption

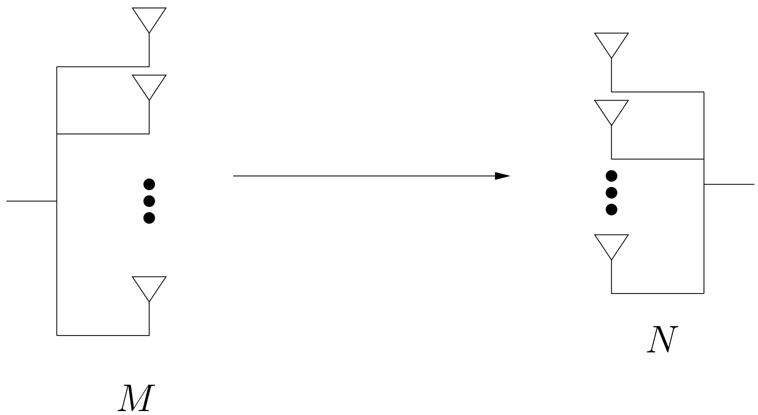

Figure 1.2: Chapter 2—MIMO link (low SNR).

is weaker than that of other works [46] and [23] which require finite higher order moments in similar mutual information computations. The approach we use is elementary and may be applied to other channel models.

We then optimize this expression subject to peak and fourth-order moment signal constraints to determine what signaling should be applied to the input and how many transmit antennas should be employed. We also study Gaussian modulation, unitary space-time modulation and training-based schemes. We find that in most cases one transmit antenna is optimal.

1.4.2 Diversity of Networks at High SNR

In chapter 3 we summarize the *diversity-multiplexing gain trade-off* for MIMO systems. First introduced by Zheng and Tse in [82], this is a way of comparing the rate versus probability of error of a system for different coding strategies, in the high SNR regime. Diversity here is a measure of reliability—the negative exponent of error probability when expressed as a function of SNR. The higher the diversity the better. A good introduction to this concept is given in chapter 9 of [68].

Typically in such analysis an outage calculation is performed, which requires knowledge of the distribution of the eigenvalues of the channel matrix. Here we also show how outage probability can be calculated in terms of the in terms of the distribution of the channel matrix itself, which may be easier to find in some instances.

The trade-off can also be applied to systems with more than one source-destination pair or with relays. This is done for the remaining setups in the thesis.

1.4.2.1 MIMO Link with Relays at High SNR

In chapter 4, applying results and techniques from existing work on the Rayleigh product channel [77], we find the optimal diversity-multiplexing gain trade-off for an $(M \times N)$ multiple-antenna system with R single antenna relays, as illustrated in figure 1.3. This setup uses a two-stage protocol [28] in which the source first transmits to relays, then the relays multiply their received signal by a unitary matrix, before forwarding the result to the multiple antenna receiver. The coding used is known as *distributed space-time coding* as it is spread over the relays in the network. The optimal trade-off and coding to achieve it has been previously studied in the $M = N = 1$ case [42, 14, 48]. Diversity is known to grow linearly in the number of relay nodes.

We find in our analysis that the optimal trade-off is the same as that of a MIMO link with R transmit and $\min\{M, N\}$ receive antennas. Previously only the maximum diversity $(R \min\{M, N\})$ and maximum multiplexing gain $(\min\{M, N\})$ were known. The optimal trade-off can be achieved by codes based on cyclic division algebras [42, 14, 48].

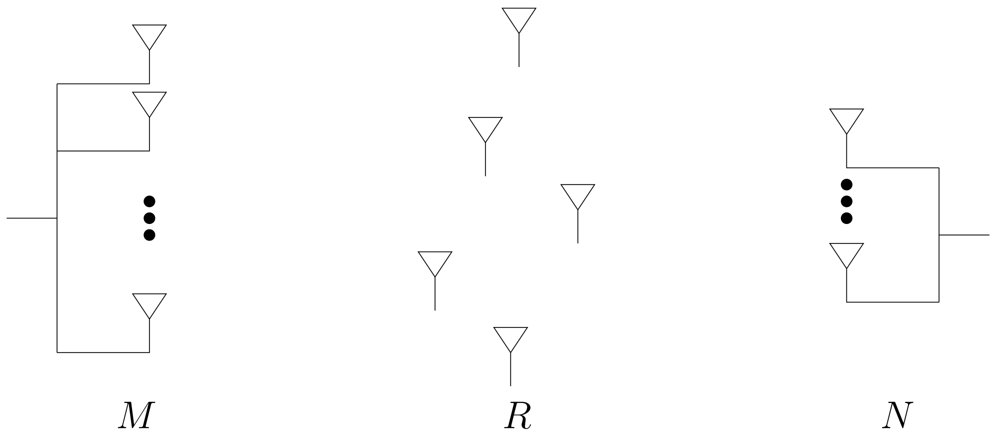

Figure 1.3: Chapter 4—MIMO link with single-antenna relays (high SNR).

1.4.2.2 More than One Source-Destination Pair at High SNR

In chapter 5 we look at systems with interference, introduced by more than one disjoint source-destination pair competing for the wireless medium. We are interested in the interplay between the rate at which one can transmit and the reliability that can be obtained at high SNR. To simplify the analysis we consider communication nodes with single antennas. In the first half of chapter 5 we consider a four-node network with two source-destination pairs. If the two sources (or two receivers) are not allowed to communicate with each other, this is known as an interference channel. The region of rates over which achievable communication over such a system is possible, has remained an open problem for over four decades [9, 24, 8, 15].

We consider several simple methods (based on [9]) for the two transmitters to communicate with their corresponding receivers and establish relationships between the rate (multiplexing gain) and diversity achieved. Then we show how cooperation amongst the nodes (allowing transmitters and receivers to share information) can be introduced to increase diversity, but at the expense of a reduced rate. This reduced rate arises from the assumption that nodes cannot transmit and receive information at the same time (a *half-duplex condition*). This work is in the area of *cooperative diversity*, where the idea is to gain system reliability by exploiting independent paths between transmitter and receiver [56, 41, 33, 3].

In our case, since there are three independent paths between transmitter and receiver that may potentially be used (see figure 1.4), diversity can be increased by a factor of up to three. Two cooperative diversity schemes are used and the trade-offs determined by outage probability calculations. Firstly the region of achievable rates is determined, then we compute the probability that a given rate pair is outside the achievability region due to fading of the channels. The results are summarized in table 1 of chapter 5.

Using nodes for cooperation in interference channels has been investigated by Høst-Madsen in [27], inheriting methods known for the interference and relay channels [32, 76]. However the rate equations there have difficult descriptions for outage analysis, and so here we consider protocols for the interference channel that allow for outage analysis, providing insight into what diversity is achievable. The two cooperative schemes also can be generalized from two to m source-destination pairs. One scheme can provide a diversity of at most $2m-1$ while reducing rate by $2m$. The other scheme has diversity m but cuts rate by a factor of $m+1$.

1.4.2.3 Source-Destination Pairs with Relays at High SNR

In the second half of chapter 5, we introduce n relay nodes into the m-user interference channel considered before (see figure 1.5), with a view to improving the diversity-multiplexing gain trade-off. Assuming the relays have channel knowledge, they may cooperate by multiplying their received signals by scalars chosen so that interference at each of the receivers

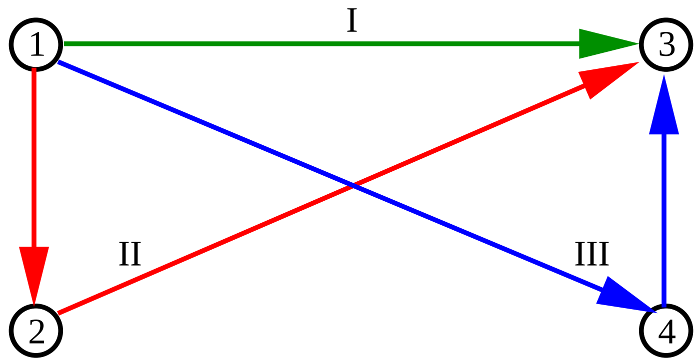

Figure 1.4: Chapter 5—four-node network consisting of two source-destination pairs labeled 1–3 and 2–4 (high SNR). Indicated are three independent transmission paths from node 1 to 3.

is canceled. This model was defined in [11] and its power efficiency studied. Depending on the present state of the channels, the choice of scalars may cancel interference, but it can lead to outage, whereby the codewords are no longer well separated. The contribution of this section of the thesis is the outage behavior of this scheme. It is shown that at full rate (two channel uses) a diversity of $n - m^2 + 2$ can be obtained. If $m^2 > n$, we can show that by using some of the transmitters and receivers as relays, high-diversity schemes are also possible, but they require more channel uses and therefore incur a rate penalty.

A natural progression from here is to understand what happens when all nodes are equipped with multiple antennas. There are also questions related to how robust these systems are to uncertainties or errors in the channels. Also, how do these results change when the nodes are heterogeneous, with different fades existing between channels? These are potential future directions of research.

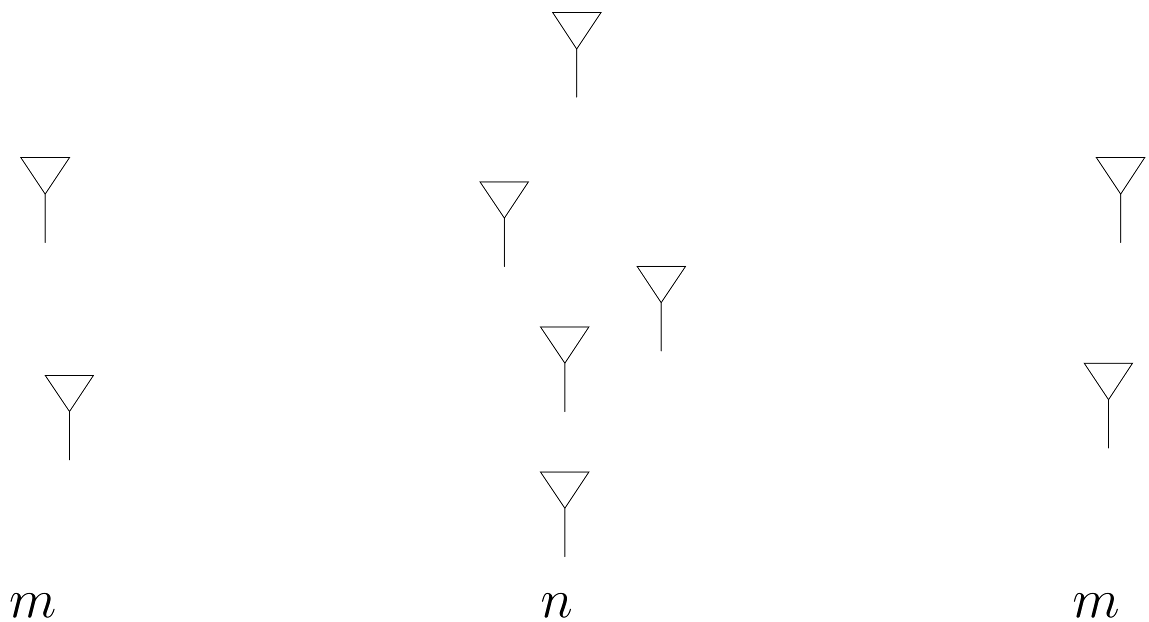

Figure 1.5: Chapter 5—m source-destination pairs with n relay nodes (high SNR).

Chapter 2

Multiple Antenna Systems at Low SNR

2.1 Introduction

Multiple antenna wireless systems have been shown to provide high capacity, exploiting the presence of fading in such channels. However, this is based on the premise that either the channel coefficients are known to the receiver, or that the signal-to-noise ratio (SNR) of the channel is high [16, 66, 81].

Wireless systems operating at low SNR (exhibiting weak signaling or in noisy environments) find increasing use in energy efficient devices such as sensor networks [72]. Recent work on analyzing the capacity of low SNR multiantenna links, assuming that the channel is known at the receiver, has appeared in [37]. However, at low SNR, channel estimates in some circumstances are unreliable and so it is sensible to assume that the channel is unknown. In the following analysis we therefore assume the channel is unknown to both transmitter and receiver. As we shall see, this leads to results qualitatively different from the known channel case.

The low SNR regime can be considered equivalent to the wideband regime, where the bandwidth of the system tends to infinity. In the literature, the wideband regime is mostly studied by adopting narrowband models and letting SNR tend to zero.

We use the block-fading model of a wireless multiple antenna system proposed by Marzetta and Hochwald in [38], expressing the mutual information between input and output as a function of the model parameter ρ (proportional to the SNR) up to second order. This model is described in detail in the next section. Maximizing this mutual information gives us insight about desired signaling at low SNR as well as the optimal number of antennas to be used at the transmitter and receiver. It has been shown by Abou-Faycal et al. in [1] that the optimum signaling at low SNR achieves the same minimum energy-per-bit as the known channel cases for single transmit antenna systems. We show that the on-off optimal signaling found in [1] also generalizes to the multiantenna setting (a result that also follows from Theorems 1 and 5 in [73]). However, this scheme requires increasingly peaky signals (indeed ones with unbounded fourth-order moment) and so may not be acceptable from a practical point of view in some situations. We therefore focus our attention on signaling schemes with bounded fourth-order moment.

Recent work by Verdú [73] has shown that knowledge of the first and second derivatives of capacity at low SNR also tells us about bandwidth and energy efficiency for signal transmission. If spectral efficiency (capacity per unit bandwidth) is expressed as a function of energy-per-bit, the minimum energy-per-bit for reliable communication is related to the first derivative of capacity, while the slope at this point (known as the *wideband slope*) is related to the second derivative [73]. More work on constrained signaling in the low-power (wideband) regime for Rayleigh fading channels is given in [67], [40] and [62], while [22] and [19] study the Rician case. In [46], amongst other things, the low-SNR mutual information for the same block-fading multiple antenna channel of [38] is also calculated. Similar results to ours have also been obtained by Hajek and Subramaniam [23], as a by-product of their study of the capacity of general communication channels under small-peak constraints. Our results differ from [46] and [23] in two ways. First, we require a weaker assumption on the input signals; essentially conditions on the fourth- and sixth-order moments, rather than an exponentially decaying input distribution as in [46], or a peak constraint on the singular values of the transmitted signal as in [23], both of which render all moments finite. Second, we study the optimal signaling structure derived in [38] and further optimize mutual information subject to various signaling constraints such as training.

There are two main parts to this chapter. In the first part we expand the mutual information of the wireless link to second order in the SNR ρ using an approach that may be applied to other channel models. Secondly we optimize this expression under both peak and fourth-order moment signal constraints to determine what signaling should be applied to the input and how many transmit and receive antennas should be employed. We also study Gaussian modulation, unitary space-time modulation and training-based schemes.

2.2 Model

We consider a discrete-time block-fading channel model due to Marzetta and Hochwald [38], in which there are M transmit and N receive antennas (see figure 2.1). The channel is described by a propagation matrix H that is assumed constant for a coherence interval of length T symbols (see [55] for further analysis of the dependence of coherence interval on capacity due to channel uncertainty). For the next T symbols the propagation matrix changes to a new independent value, and so on. Signals are represented by complex valued matrices with T rows, where the ith row is what is transmitted or received via each of the multiple antennas at the ith time slot.

For each coherence interval the $T \times N$ received matrix X is related to the $T \times M$

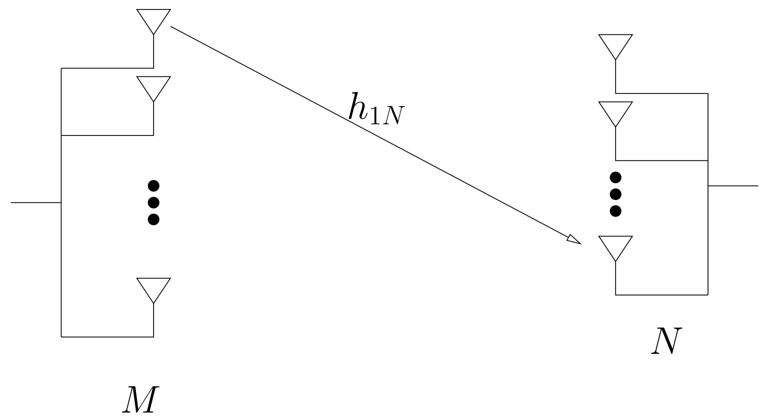

Figure 2.1: MIMO system with M transmit and N receive antennas. One of the entries of the channel matrix H is shown.

transmitted matrix S by

$$X = \sqrt{\frac{\rho}{M}} SH + V, \tag{2.1}$$

where H is $M \times N$ and V is a $T \times N$ noise matrix, both comprised of zero-mean and unit variance circularly symmetric complex Gaussian entries. (See [31], [36], [59], [75] and [69] for capacity analyses with channel correlation.) The matrices H, V and S are assumed to be independent and the values of H and V are unknown to both transmitter and receiver. S satisfies the power constraint $\mathbf{E}\mathrm{tr}SS^* = \mathbf{E}\sum_{i=1}^{T}\sum_{j=1}^{M}|s_{ij}|^2 \stackrel{\Delta}{=} \eta_2 \leq P_{\max}$, where $\mathbf{E}$ and tr denote the expectation and trace operators respectively, and s_{ij} is the (i,j)th entry of S. Throughout this work the * operator denotes the conjugate transpose of a matrix. When $\eta_2 = TM$ the normalization factor $\sqrt{\frac{\rho}{M}}$ in (2.1) makes ρ equal to the signal-to-noise ratio at each receive antenna. Otherwise the SNR at each receive antenna is given by $\mathbf{E}[\mathrm{tr}XX^*]/\mathbf{E}[\mathrm{tr}VV^*] = \rho\eta_2/(TM)$. It is also known that there is no performance gain in having the number of transmit antennas greater than T [38]. Hence we will assume that $T \geq M$.

Computing the capacity of this multiple antenna system for generic ρ is an open problem. In [38], however, it is shown that capacity is achieved when the input signal S has the form

$$S = \Phi D, \tag{2.2}$$

where D is a diagonal $M \times M$ matrix with non-negative entries, and Φ is a $T \times M$ *isotropically distributed unitary* random matrix. This means

- $\Phi^* \Phi = I_M$ (the $M \times M$ identity matrix) although $\Phi \Phi^* \neq I_T$ for $T > M$, and
- the distribution of Φ is unaltered when left-multiplied by a deterministic $T \times T$ unitary matrix or when right-multiplied by a deterministic $M \times M$ unitary matrix [38].

Moreover, Φ and D are independently distributed.

2.3 Mutual Information to Second Order

In this section we prove the following result.

Theorem 1. *Consider the model (2.1) and let $p(S)$ denote the pdf of S.*

1. **First-order result:** *If (i) $\partial p(S)/\partial \rho$ exists at $\rho = 0$ and (ii) $\lim_{\rho \to 0} \rho \mathbf{E} \mathrm{tr}(SS^*)^2 = 0$, the mutual information between the transmitted and received signals S and X for the multiple antenna system (2.1) is zero to first order in ρ, i.e., $I(X; S) = o(\rho)$.*

2. **Second-order result:** *If in addition (i) $\partial^2 p(S)/\partial \rho^2$ exists at $\rho = 0$, (ii) the fourth-order moment of S is finite, i.e., $\mathbf{E} \mathrm{tr}(SS^*)^2 \triangleq \eta_4 < \infty$ and (iii) $\lim_{\rho \to 0} \rho \mathbf{E} \mathrm{tr}(SS^*)^3 = 0$, then the mutual information between S and X up to second order in ρ is given by*

$$I(X; S) = \frac{N \mathrm{tr}[\mathbf{E}(SS^*)^2 - (\mathbf{E} SS^*)^2]}{2M^2} \rho^2 + o(\rho^2). \tag{2.3}$$

The second-order part of the theorem is essentially a result in [46] and [23]. However, we here require a much less stringent condition on the input distribution. Moreover, we shall optimize (2.3) for various signaling schemes.

The reason for the condition on $p(S)$ in Theorem 1, is that the choice of distribution may depend on the SNR ρ. Condition (ii) of the first-order result limits the growth of the fourth-order moment, whereas conditions (ii) and (iii) of the second-order result respectively bound and limit the growth of the fourth- and sixth-order moments. The regularity conditions (i) on $p(S)$ at $\rho = 0$ are required for reasons that will be seen shortly (see sections 3.2 and 5).

For the optimum signaling structure (2.2), equation (2.3) can be replaced by

$$I(X;S) = \frac{N(\eta_4 - \eta_2^2/T)}{2M^2}\rho^2 + o(\rho^2). \tag{2.4}$$

Note that under any reasonable input distribution (and certainly all practical modulation schemes) the mutual information has no linear term in ρ and so the capacity is *much less* than the known channel case where the low SNR expansion of the well-known $\log\det$ formula has a non-zero first-order term. Since η_4 and η_2 are independent of N, (2.4) suggests that the capacity increases linearly in the number of receive antennas. The dependence of the mutual information on M is more complicated since both the denominator (M^2), as well as the numerator (via η_2 and η_4) depend on M. However, careful analysis will show that for most practical signal constraints, the optimal value is $M = 1$ transmit antenna.

Finally, note that the mutual information is affine linear in η_4 suggesting that it increases as the input becomes more peaky, in good agreement with the results of [1] and their multiantenna generalizations.

2.3.1 Conditional Entropy Approximation

We compute $I(X;S) = h(X) - h(X|S)$ via the conditional probability density function $p(X|S)$. Given S, X is zero-mean complex Gaussian with covariance $\mathbf{E}(XX^*|S) = N(I_T + \frac{\rho}{M}SS^*)$ and so as in [38],

$$p(X|S) \quad = \quad \frac{e^{-\mathrm{tr}X^*(I_T+\frac{\rho}{M}SS^*)^{-1}X}}{\pi^{NT}\left[\det(I_T+\frac{\rho}{M}SS^*)\right]^N}. \tag{2.5}$$

Here I_T denotes a $T \times T$ identity matrix. From $p(X|S)$ it is possible to compute the conditional entropy $h(X|S)$ directly:

$$\begin{aligned}
h(X|S) &= -\mathbf{E}\log p(X|S) \\
&= NT\log\pi + \mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*)^N + \mathbf{E}\mathrm{tr}X^*(I_T+\frac{\rho}{M}SS^*)^{-1}X \\
&= NT\log\pi + N\mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*) + \mathbf{E}\mathrm{tr}(I_T+\frac{\rho}{M}SS^*)^{-1}XX^* \\
&\qquad \text{(using } \mathrm{tr}AB = \mathrm{tr}BA\text{)} \\
&= NT\log\pi + N\mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*) + \mathbf{E}_S\mathrm{tr}NI_T \\
&= NT\log\pi e + N\mathbf{E}\log\prod_i(1+\frac{\rho}{M}d_i^2) \\
&\qquad \text{(where } d_i^2 \text{ are the eigenvalues of } SS^*\text{)} \\
&= NT\log\pi e + N\mathbf{E}\sum_i\log(1+\frac{\rho}{M}d_i^2) &(2.6)\\
&\approx NT\log\pi e + N\mathbf{E}\sum_i(\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4) &(2.7)\\
&= NT\log\pi e + N\frac{\rho}{M}\eta_2 - \frac{N\rho^2}{2M^2}\eta_4, &(2.8)
\end{aligned}$$

since $\eta_2 = \mathbf{E}\mathrm{tr}SS^* = \mathbf{E}\sum_i d_i^2$ and $\eta_4 = \mathbf{E}\mathrm{tr}(SS^*)^2 = \mathbf{E}\sum_i d_i^4$.

The approximation step (2.7) is made assuming that the second-order approximation $\mathbf{E}\log(1+\frac{\rho}{M}d_i^2) \approx \mathbf{E}[\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4]$ is valid for each i. Consider the inequality

$$\begin{aligned} \frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4 \leq \quad & \log(1+\frac{\rho}{M}d_i^2) && \leq \frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4 + \frac{\rho^3}{3M^3}d_i^6 \\ \Rightarrow 0 \leq \quad & \frac{\log(1+\frac{\rho}{M}d_i^2) - (\frac{\rho}{M}d_i^2 - \frac{\rho^2}{2M^2}d_i^4)}{\rho^2} && \leq \frac{\rho}{3M^3}d_i^6. \end{aligned} \tag{2.9}$$

For the second-order approximation to be a valid one, the limit of the expression between the two inequalities in (2.9) should go to zero *in expectation* as $\rho \to 0$ for each i. The condition $\rho\mathbf{E}\mathrm{tr}(SS^*)^3 = \rho\mathbf{E}\sum(d_i^2)^3 \to 0$ in the second-order statement of Theorem 1 ensures that this occurs. The first-order condition $\rho\mathbf{E}\mathrm{tr}(SS^*)^2$ similarly ensures that $\rho d_i^2 \to 0$, making $\log(1+\frac{\rho}{M}d_i^2) \approx \frac{\rho}{M}d_i^2$ a valid first-order approximation.

2.3.2 Entropy Approximation

The pdf $p(X)$ depends on the input distribution $p(S)$. Our regularity conditions (i) on $p(S)$ in Theorem 1 guarantee that the distribution can be expanded to second order around $\rho = 0$ as $p(S) = p(S,0) + \rho p'(S,0) + \frac{\rho^2}{2}p''(S,0) + o(\rho^2)$. Also, $p(X|S)$ in (2.5) is a function whose derivatives with respect to ρ at $\rho = 0$ may be calculated. These two facts imply that

$$p(X) = \int p(S)p(X|S)\ dS$$

can also be expanded to second order as

$$p(X) = p(X,0) + \rho p'(X,0) + \frac{\rho^2}{2}p''(X,0) + o(\rho^2),$$

where $p'(X,0)$ and $p''(X,0)$ are used to denote the first and second partial derivatives of $p(X)$ with respect to ρ respectively, evaluated at $\rho = 0$. Also to second order

$$\log p(X) \approx \log p(X,0) + \rho\frac{p'(X,0)}{p(X,0)} + \frac{\rho^2}{2}\left[\frac{p''(X,0)}{p(X,0)} - \left(\frac{p'(X,0)}{p(X,0)}\right)^2\right].$$

This leads us to the following quadratic approximation,

$$\begin{aligned} h(X) &= -\int p(X)\log p(X)\ dX \\ &\approx -\int p(X,0)\log p(X,0)\ dX - \rho\int \big(p'(X,0)\log p(X,0) + p'(X,0)\big)\ dX \\ &\quad - \frac{\rho^2}{2}\int \big(p''(X,0) + (p'(X,0))^2/p(X,0) + p''(X,0)\log p(X,0)\big)\ dX. \quad (2.10) \end{aligned}$$

We now claim that the integrals in (2.10) involving the second derivative $p''(X,0)$ are equal to zero.

Firstly, note that

$$\int \left(p(X,0) + \rho p'(X,0) + \frac{\rho^2}{2}p''(X,0) + o(\rho^2)\right) dX = 1.$$

Comparing coefficients of ρ^n on both sides we conclude

$$\int p^{(n)}(X,0)\ dX = 0, \quad n = 1, 2.$$

Also, since S, H and V are independent, (2.1) implies that

$$\begin{aligned}
\mathbf{E}\mathrm{tr}XX^* &= \mathbf{E}\mathrm{tr}\frac{\rho}{M}SHH^*S^* + \sqrt{\frac{\rho}{M}}\mathbf{E}\mathrm{tr}SHV^* + \sqrt{\frac{\rho}{M}}\mathbf{E}\mathrm{tr}VH^*S^* + \mathbf{E}\mathrm{tr}VV^* \\
&= \frac{\rho}{M}\mathrm{tr}\mathbf{E}HH^*S^*S + 0 + 0 + \mathrm{tr}NI_T \\
&= \frac{\rho}{M}\mathrm{tr}NI_M\mathbf{E}S^*S + NT \\
&= N(\rho\eta_2/M + T). \qquad (2.11)
\end{aligned}$$

Thus,

$$\int \mathrm{tr}XX^*\left(p(X,0) + \rho p'(X,0) + \frac{\rho^2}{2}p''(X,0) + o(\rho^2)\right)dX = N(\rho\eta_2/M + T).$$

Comparing coefficients of ρ^n of on both sides

$$\begin{aligned}
\int \mathrm{tr}X^*Xp(X,0)\ dX &= NT, \\
\int \mathrm{tr}X^*Xp'(X,0)\ dX &= N\eta_2/M, \\
\int \mathrm{tr}X^*Xp''(X,0)\ dX &= 0.
\end{aligned}$$

Now $p(X,0) = \frac{e^{-\mathrm{tr}X^*X}}{\pi^{NT}}$ and so $\log p(X,0) = -\mathrm{tr}X^*X - NT\log\pi$. Hence the zeroth-order term in (2.10) is

$$\begin{aligned}
&\int p(X,0)\log p(X,0)\ dX \\
&= -\int p(X,0)\mathrm{tr}X^*X\ dX - NT\log\pi\int p(X,0)\ dX \\
&= -NT - NT\log\pi \\
&= -NT\log\pi e.
\end{aligned}$$

Similarly,

$$\int p'(X,0)\log p(X,0)\ dX = -N\eta_2/M,$$

and

$$\int p''(X,0)\log p(X,0)\ dX = 0.$$

The above calculations combined with (2.10) lead to

$$h(X) \approx NT\log \pi e + \rho N\eta_2/M - \frac{\rho^2}{2}\int (p'(X,0))^2/p(X,0)\ dX. \tag{2.12}$$

This shows that to express $h(X)$ to second order, it suffices to calculate only the first derivative of $p(X)$ at $\rho = 0$. We use the following result, proved in the appendix.

Lemma 1. *For the model (2.1), the first derivative of the pdf of X evaluated at $\rho = 0$ is given by*

$$p'(X,0) = \frac{e^{-\mathrm{tr}X^*X}}{M\pi^{NT}}(\mathrm{tr}X^*PX - N\eta_2),$$

where $P = \mathbf{E}SS^*$.

This gives us

$$\begin{aligned}\int \frac{(p'(X,0))^2}{p(X,0)}\ dX &= \int \frac{e^{-\mathrm{tr}X^*X}}{M^2\pi^{NT}}(\mathrm{tr}X^*PX - N\eta_2)^2\ dX \\ &= \frac{1}{M^2}\left[\mathbf{E}_G(\mathrm{tr}G^*PG)^2 - 2N\eta_2\mathbf{E}_G\mathrm{tr}G^*PG + N^2\eta_2^2\right],\end{aligned} \tag{2.13}$$

where G is a $T \times M$ random matrix having the pdf $p(G) = \frac{e^{-\mathrm{tr}G^*G}}{\pi^{NT}}$, and $\mathbf{E}_G$ denotes expectation over the random variable G.

To proceed we use the following lemma proved in the appendix.

Lemma 2. *If G is a $T \times N$ matrix with independent zero-mean unit-variance complex*

Gaussian entries, then

- $\mathbf{E}_G \mathrm{tr} G^* P G = N\eta_2$,

- $\mathbf{E}_G (\mathrm{tr} G^* P G)^2 = N^2\eta_2^2 + N \mathrm{tr} P^2$,

for any $T \times T$ *deterministic matrix* P *satisfying* $\mathrm{tr} P = \eta_2$.

For $P = \mathbf{E}SS^*$, from (2.13) and Lemma 2 we have

$$\begin{aligned} \int \frac{(p'(X,0))^2}{p(X,0)}\, dX &= \frac{1}{M^2}\left[N^2\eta_2^2 + N\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right] - 2N\eta_2 N\eta_2 + N^2\eta_2^2\right] \\ &= \frac{N}{M^2}\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right]. \end{aligned}$$

Hence $h(X) \approx NT\log \pi e + N\eta_2/M\rho - \frac{N}{2M^2}\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right]\rho^2$ from (2.12) and this together with (2.8) gives

$$\begin{aligned} I(X;S) &= h(X) - h(X|S) \\ &\approx (NT\log \pi e + N\eta_2/M\rho - \frac{N}{2M^2}\mathrm{tr}\left[(\mathbf{E}SS^*)^2\right]\rho^2) \\ &\quad - (NT\log \pi e + N\eta_2/M\rho - \frac{N\eta_4}{2M^2}\rho^2) \\ &= \frac{N\mathrm{tr}\left[\mathbf{E}(SS^*)^2 - (\mathbf{E}SS^*)^2\right]}{2M^2}\rho^2, \end{aligned} \tag{2.14}$$

as stated in Theorem 1.

We remark that to show the first-order result $I(X;S) = o(\rho)$ we only require first-order expansions of $p(X)$ and $h(X)$, and so the conditions stated in the first-order result of Theorem 1 suffice.

In the special capacity-optimizing case of $S = \Phi D$, we have

$$\begin{aligned}
\mathbf{E}(SS^*)_{ij} &= \mathbf{E}(\Phi D^2 \Phi^*)_{ij} \\
&= \mathbf{E} \sum_k \phi_{ik} d_k^2 \overline{\phi}_{jk} \\
&\quad (d_k^2 \text{ are both the diagonal entries of } D^2 \text{ and the eigenvalues of } SS^*) \\
&= \sum_k \mathbf{E}[d_k^2] \mathbf{E}\left[\phi_{ik}\overline{\phi}_{jk}\right] \quad \text{(since } \Phi \text{ and } D \text{ are independent).}
\end{aligned}$$

The expectation $\mathbf{E}\left[\phi_{ik}\overline{\phi}_{jk}\right]$ is evaluated by noticing that the expectation is unchanged by adding $T - M$ orthonormal columns to Φ to make it a $T \times T$ unitary, denoted, say, by $\Psi = (\psi_{ij})$. Then using the relation $\Psi\Psi^* = I_M$ we have

$$\sum_{k=1}^{T} \psi_{ik}\overline{\psi}_{jk} = \delta_{ij}.$$

Taking expectations of both sides and using the fact that each entry of Ψ has the same distribution, gives us

$$\mathbf{E}\left[\psi_{ik}\overline{\psi}_{jk}\right] = \frac{\delta_{ij}}{T} \quad \text{for } k = 1 \text{ to } T.$$

This implies $\mathbf{E}\left[\phi_{ik}\overline{\phi}_{jk}\right] = \frac{\delta_{ij}}{T}$ for $k = 1$ to M.

Hence

$$\begin{aligned}
\mathbf{E}(SS^*)_{ij} &= \frac{\delta_{ij}}{T} \sum_k \mathbf{E}[d_k^2] \\
&= \frac{\delta_{ij}}{T} \eta_2.
\end{aligned}$$

In other words $\mathbf{E}SS^* = \frac{\eta_2}{T} I_T$, and so $\mathrm{tr}(\mathbf{E}SS^*)^2 = \eta_2^2/T$. This in (2.14) gives

$$I(X;S) = \frac{N(\eta_4 - \eta_2^2/T)}{2M^2}\rho^2 + o(\rho^2).$$

2.4 Examples

We now compute the low SNR mutual information for some cases of interest.

2.4.1 Gaussian Modulation

Suppose the transmitted signal S has independent zero-mean unit-variance complex Gaussian entries. Then $(\mathbf{E}SS^*)_{ij} = \sum_k \mathbf{E}s_{ik}\bar{s}_{jk} = M\delta_{ij}$, so that $\mathbf{E}SS^* = MI_T$. In the appendix we show that for a Gaussian matrix, $\eta_4 = \mathbf{E}(\mathrm{tr}(SS^*)^2) = MT(M+T)$ and so

$$\begin{aligned} \frac{1}{T} I(X;S) &= \frac{N\rho^2}{2M^2T}(MT(M+T) - (TM)^2/T) + o(\rho^2) \\ &= \frac{NT}{2M}\rho^2 + o(\rho^2). \end{aligned}$$

This has two interesting ramifications. First the capacity per channel use increases linearly in T ($I(X;S)$ is actually quadratic in T) and, second, the optimal number of transmit antennas is $M = 1$.

2.4.2 Unitary Space-Time Modulation

In this scheme, we let $S = \Phi\sqrt{T}$ (where Φ has an isotropic unitary distribution), which gives $\eta_2 = TM$ and $\eta_4 = T^2M$. Using this in (2.4) yields

$$\frac{1}{T} I(X;S) = \frac{N(T-M)}{2M}\rho^2 + o(\rho^2),$$

which is strictly less than the Gaussian case. Again, the optimal number of transmit antennas is $M = 1$.

2.4.3 Training-Based Schemes

In these schemes, we have

$$S = \begin{bmatrix} S_\tau \\ S_d \end{bmatrix},$$

where S_τ is a $T_\tau \times M$ fixed training matrix and S_d is a $T_d \times M$ zero-mean random matrix. Furthermore,

$$\mathrm{tr} S_\tau^* S_\tau = \eta_{2,\tau},\ \mathbf{E}\mathrm{tr} S_d S_d^* = \eta_{2,d},\ \eta_{2,\tau} + \eta_{2,d} = \eta_2,\ T_\tau + T_d = T.$$

Under these conditions it can be readily shown that

$$\mathrm{tr}(\mathbf{E} SS^*)^2 = \mathrm{tr}(S_\tau S_\tau^*)^2 + \mathrm{tr}(\mathbf{E} S_d S_d^*)^2,$$

and

$$\mathrm{tr}\mathbf{E}(SS^*)^2 = \mathrm{tr}(S_\tau S_\tau^*)^2 + \mathrm{tr}\mathbf{E}(S_d S_d^*)^2.$$

Therefore, using (2.3), we obtain

$$I(X; S) = \frac{N \mathrm{tr}[\mathbf{E}(S_d S_d^*)^2 - (\mathbf{E} S_d S_d^*)^2]}{2M^2} \rho^2 + o(\rho^2). \tag{2.15}$$

The latter equation shows that the mutual information is independent of S_τ. In fact, the right-hand side of (2.15) is just the mutual information of a system with coherence interval $T_d = T - T_\tau$. Thus, in the low SNR regime, training actually contributes a rate reduction proportional to the fraction of time that one is sending the training symbols. One may contrast this with the result of [25] which shows that training-based schemes achieve capacity at high SNR.

2.5 Optimal Signaling

In this section we shall optimize (2.4) to determine what kind of signaling should be applied to maximize the mutual information between the transmitted and received signals. It is known that, under the standard power (second-order) constraint, capacity approaches up to first order the capacity of a channel where the channel matrix is perfectly known to the receiver. This is achieved by a peaky input distribution [1].

We can show this is also the case for the multiantenna channel as follows. For any $\epsilon > 1$ and assuming $\rho < 1$, define our transmitted signal to satisfy

$$SS^* = \begin{cases} A & \text{w.p. } \rho^\epsilon, \\ 0_{T\times T} & \text{w.p. } 1-\rho^\epsilon, \end{cases}$$

where

$$A = T\rho^{-\epsilon}\begin{bmatrix} I_M & 0_{M\times(T-M)} \\ 0_{(T-M)\times M} & 0_{(T-M)\times(T-M)} \end{bmatrix}.$$

Then S satisfies the power constraint $\mathbf{E}\mathrm{tr}SS^* = \mathrm{tr}(T\rho^{-\epsilon}I_M) \times \rho^\epsilon = TM$. Note that the above distribution does not satisfy the regularity conditions (i) of Theorem 1.

Then from (2.6),

$$\begin{aligned} h(X|S) &= NT\log \pi e + N\mathbf{E}\log\det(I_T + \frac{\rho}{M}SS^*) \\ &= NT\log \pi e + N\rho^\epsilon \log(1+\frac{\rho}{M}T\rho^{-\epsilon})^M \\ &\approx NT\log\pi e + N\rho^\epsilon M\log\left(\frac{T}{M}\rho^{1-\epsilon}\right) \quad \text{(as } \rho^{1-\epsilon} \text{ is large)} \\ &= NT\log\pi e + NM\rho^\epsilon[(1-\epsilon)\log\rho + \log(T/M)] \\ &= NT\log\pi e + o(\rho) \quad \text{since } \epsilon > 1. \end{aligned}$$

Also we have

$$p(X) = \rho^\epsilon \frac{e^{-\mathrm{tr}X^*(I_T+\frac{\rho}{M}A)^{-1}X}}{\pi^{NT}\det(I_T+\frac{\rho}{M}A)^N} + (1-\rho^\epsilon)\frac{e^{-\mathrm{tr}X^*X}}{\pi^{NT}}. \tag{2.16}$$

For ρ small and $\epsilon > 1$, ρ^ϵ is small while $\rho^{1-\epsilon}$ is large. Hence in the first term of (2.16) the determinant in the denominator is $(1+\frac{\rho}{M}T\rho^{-\epsilon})^{MN} \approx \rho^{MN(1-\epsilon)}$ which is large while the numerator is bounded above by 1. Therefore the second term is much larger than the first, and so

$$\begin{aligned} h(X) &= -\mathbf{E}\log p(X) \\ &\approx -\mathbf{E}\log(1-\rho^\epsilon)\frac{e^{-\mathrm{tr}X^*X}}{\pi^{NT}} \\ &= -\log(1-\rho^\epsilon) + \mathbf{E}\mathrm{tr}X^*X + NT\log\pi \\ &\approx \rho^\epsilon + NT\log\pi + \mathbf{E}\mathrm{tr}X^*X \\ &= NT\log\pi + NT(\rho+1) + \rho^\epsilon \quad \text{(using (2.11))} \\ &= NT(\log\pi e + \rho) + o(\rho). \end{aligned}$$

Then $I(X;S)/T = N\rho + o(\rho)$, so the first-order term corresponds to that of the capacity when the channel is known, equal to $\mathbf{E} \log\det(I + \frac{\rho}{M}HH^*) = N\rho + o(\rho)$. However, such signals cannot be used in practice, and so we shall consider signals that are peak constrained or have a fourth-order moment constraint.

2.5.1 Fourth-Order Moment Constraint

Suppose we enforce the fourth-order moment constraint $\eta_4 \leq K\eta_2^2$ for some positive constant K. This may be a practical constraint to impose, but as mentioned in [73], a bounded fourth-order moment will not lead to mutual information optimality at low SNR.

By the root mean square–arithmetic mean inequality we have

$$\sum_{i=1}^{M} d_i^4 \;\geq\; \frac{1}{M}\left(\sum_{i=1}^{M} d_i^2\right)^2,$$

from which we conclude $\eta_4 \geq \eta_2^2/M$. Also, as stated in Section 2.2, $T \geq M$, and so $\eta_4 \geq \eta_2^2/T$. Hence we require that $K > 1/M \geq 1/T$.

Then $\eta_4 - \eta_2^2/T \leq (K-1/T)\eta_2^2$ and as $K - 1/T > 0$, from (2.4) it follows that maximizing the mutual information is equivalent to maximizing η_2. We therefore have the following result.

Theorem 2. *Consider the model (2.1) and suppose that the input signal must satisfy the constraints $\eta_2 \leq TM$ and $\eta_4 \leq K\eta_2^2$. Then, to second order, the mutual information is maximized by any input distribution that simultaneously achieves $\eta_2 = TM$ and $\eta_4 = K\eta_2^2$ and is given by*

$$\frac{1}{T}I(X;S) = \frac{N(K-1/T)T}{2}\rho^2 + o(\rho^2). \tag{2.17}$$

One distribution that achieves this is given by (2.2) where

$$(d_1^2, d_2^2, \ldots, d_M^2) = \begin{cases} (TKM, TKM, \ldots, TKM) & w.p.\ 1/(KM), \\ (0, 0, \ldots, 0) & w.p.\ 1 - 1/(KM). \end{cases}$$

Note here that the optimal mutual information (per channel use) is independent of the number of transmit antennas and is proportional to both N and T.

2.5.2 Peak Constraint

Due to the isotropic unitary matrix in (2.2), it is not possible to directly enforce a peak constraint on the transmitted signals. However, it is possible to force a maximum constraint on the diagonal entries of D (the singular values of S). To this end, assume that $d_i^2 \leq K$ for some positive constant K and for all i. For any fixed M, we wish to maximize $\eta_4 - \eta_2^2/T$ subject to the constraint $\eta_2 \leq P_{\max}$. We also have $\eta_2 = \mathbf{E}\sum_{i=1}^{M} d_i^2 \leq MK$. Now

$$\begin{aligned} \eta_4 - \eta_2^2/T &= \mathbf{E}\left[\sum_{i=1}^{M} d_i^4\right] - \eta_2^2/T \\ &\leq K\mathbf{E}\left[\sum_{i=1}^{M} d_i^2\right] - \eta_2^2/T \\ &= \eta_2(K - \eta_2/T), \end{aligned}$$

with equality if and only if all the d_i's are equal to 0 or K. This quantity is maximized at either $\eta_2 = TK/2$, $\eta_2 = P_{\max}$ or $\eta_2 = MK$, depending on which of the three quantities

$TK/2$, $P_{\max}$ or MK is smallest. This leads to

$$\eta_4 - \eta_2^2/T \quad \leq \quad \begin{cases} TK^2/4, & \text{if } L = TK/2, \\ P_{\max}(K - P_{\max}/T), & \text{if } L = P_{\max}, \\ MK^2(T-M)/T, & \text{if } L = MK, \end{cases} \tag{2.18}$$

where $L = \min\{TK/2, P_{\max}, MK\}$. Equality holds in (2.18) when η_2 is set to $\min\{TK/2, P_{\max}, MK\}$. Corresponding distributions that achieve equality are

$$(d_1^2, d_2^2, \ldots, d_M^2) \quad = \quad \begin{cases} (K, K, \ldots, K) & \text{w.p. } \min\{1, T/2M\}, \\ (0, 0, \ldots, 0) & \text{w.p. } 1 - \min\{1, T/2M\}, \end{cases} \tag{2.19}$$

and

$$(d_1^2, d_2^2, \ldots, d_M^2) \quad = \quad \begin{cases} (K, K, \ldots, K) & \text{w.p. } \min\{1, P_{\max}/MK\}, \\ (0, 0, \ldots, 0) & \text{w.p. } 1 - \min\{1, P_{\max}/MK\}, \end{cases} \tag{2.20}$$

depending on whether $P_{\max} \geq TK/2$ or $P_{\max} < TK/2$ respectively. The mutual information bounds are

$$I(X;S) \quad \leq \quad \begin{cases} NTK^2\rho^2/8M^2, & \text{if } L = TK/2, \\ NP_{\max}(K - P_{\max}/T)\rho^2/2M^2, & \text{if } L = P_{\max}, \\ NK^2(T-M)\rho^2/2TM, & \text{if } L = MK. \end{cases}$$

Note that all the above bounds are decreasing functions of M. Therefore it is clear that the optimal choice of transmit antennas is $M = 1$. Since it is most likely that $K < P_{\max}$

(that is, $L = MK = K$ unless $T = 1$ in which case $L = TK/2 = K/2$), we have the following theorem.

Theorem 3. *In the model (2.1) with optimal signaling as in (2.2) suppose the diagonal entries d_i of D satisfy $d_i^2 < K$ for all i, where K is some constant less than $P_{\max}$. Then for asymptotically low SNR the optimal number of transmit antennas is $M = 1$, and the optimal mutual information is*

$$\frac{1}{T} I(X;S) = \begin{cases} \frac{NK^2(T-1)}{2T}\rho^2 + o(\rho^2) & \text{if } T > 1, \\ \frac{NK^2}{8}\rho^2 + o(\rho^2) & \text{if } T = 1. \end{cases}$$

One distribution that achieves this is given in (2.2) where the diagonal entries of D^2 are given by (2.19) or (2.20) depending on whether $P_{\max} \geq TK/2$ or $P_{\max} < TK/2$ respectively.

2.6 Summary

For the block-fading multiple antenna channel model in which the channel is unknown to the transmitter and receiver, we found that for reasonable input distributions (in particular all practical modulation schemes), the low-SNR asymptotic mutual information has a quadratic leading term. This is much less than the known channel case where it exhibits a linear growth in SNR.

Under various signaling constraints, e.g., Gaussian modulation, unitary space-time modulation, and peak constraints, this mutual information is maximized by using a single transmit antenna. Furthermore, the mutual information per channel use is linear in both the number of receive antennas and the channel coherence interval. Interestingly, when there is a maximum constraint on the singular values of the transmit signal, it is possible to obtain a higher capacity by lowering the signal power from its maximum allowed level. We also show that at low SNR, sending training symbols leads to a rate reduction in proportion to the fraction of training duration time, so that it is best not to perform training.

2.7 Appendix—Proof of Lemmas

Lemma 1. *For the model (2.1), the first derivative of the pdf of X evaluated at $\rho = 0$ is given by*

$$p'(X, 0) = \frac{e^{-\mathrm{tr}X^*X}}{M\pi^{NT}}(\mathrm{tr}X^*PX - N\eta_2),$$

where $P = \mathbf{E}SS^*$.

Proof. We first approximate $p(X|S)$ to first order in ρ. Expanding the numerator to first order:

$$\begin{aligned} e^{-\mathrm{tr}X^*(I_T + \frac{\rho}{M}SS^*)^{-1}X} &\approx e^{-\mathrm{tr}X^*(I_T - \frac{\rho}{M}SS^*)X} \\ &= e^{-\mathrm{tr}X^*X}e^{\mathrm{tr}(\frac{\rho}{M}X^*SS^*X)} \\ &\approx e^{-\mathrm{tr}X^*X}\left[1 + \frac{\rho}{M}\mathrm{tr}X^*SS^*X\right]. \end{aligned}$$

To expand the denominator we use

$$\begin{aligned} \det(I + \rho A) &= e^{\log\det(\mathrm{I}+\rho\mathrm{A})} \\ &= e^{\mathrm{tr}\log(I+\rho A)} \\ &\approx e^{\mathrm{tr}(\rho A)} \\ &\approx 1 + \mathrm{tr}(\rho A), \end{aligned}$$

so that

$$\det\left(I_T + \frac{\rho}{M}SS^*\right)^{-N} \approx \left[1 + \frac{\rho}{M}\mathrm{tr}SS^*\right]^{-N}$$
$$\approx 1 - \frac{\rho N}{M}\mathrm{tr}SS^*.$$

Putting everything together

$$p(X|S) \approx \frac{e^{-\mathrm{tr}X^*X}}{\pi^{NT}}\left[1 + \tfrac{\rho}{M}\mathrm{tr}X^*SS^*X\right]\left[1 - \tfrac{\rho N}{M}\mathrm{tr}SS^*\right],$$

and by taking the coefficient of ρ of both sides it follows that

$$p'(X|S,\rho=0) = \frac{e^{-\mathrm{tr}X^*X}}{M\pi^{NT}}(\mathrm{tr}X^*SS^*X - N\mathrm{tr}SS^*). \tag{2.21}$$

To find $p'(X,0)$ we take the expectation of (2.21) over S, leading to the required result. $\square$

Lemma 2. *If G is a $T \times N$ matrix with independent zero-mean unit-variance complex Gaussian entries, then*

- $\mathbf{E}_G\mathrm{tr}G^*PG = N\eta_2$,

- $\mathbf{E}_G(\mathrm{tr}G^*PG)^2 = N^2\eta_2^2 + N\mathrm{tr}P^2$,

for any $T \times T$ matrix P satisfying $\mathrm{tr}P = \eta_2$.

Proof. 1. Denoting the (i,j)th entries of G and P by g_{ij} and p_{ij} respectively we have

$$\begin{aligned}
\mathbf{E}_G \mathrm{tr} G^* P G &= \mathbf{E}_G \sum_{i,j,k} \overline{g}_{ji} p_{jk} g_{ki} \\
&= \sum_{i,j,k} p_{jk} \mathbf{E}_G [\overline{g}_{ji} g_{ki}] \\
&= \sum_{i,j,k} p_{jk} \delta_{jk} \\
&\quad \text{(since } g_{ki} \text{ and } g_{ji} \text{ are independent unless } j=k, \\
&\quad\quad \text{in which case the expectation is unity)} \\
&= N \sum_{j,k} p_{jk} \delta_{jk} \\
&= N \sum_{j} p_{jj} \\
&= N \mathrm{tr} P
\end{aligned}$$

as required.

2. We have

$$\begin{aligned}
\mathbf{E}_G (\mathrm{tr} G^* P G)^2 &= \mathbf{E}_G \left(\sum_{i,j,k} \overline{g}_{ji} p_{jk} g_{ki} \right)^2 \\
&= \mathbf{E}_G \sum_{i,j,k,l,m,n} \overline{g}_{ji} p_{jk} g_{ki} \overline{g}_{ml} p_{mn} g_{nl} \\
&= \sum_{j,k,m,n} p_{jk} p_{mn} \sum_{i,l} \mathbf{E}_G \left[\overline{g}_{ji} g_{ki} \overline{g}_{ml} g_{nl} \right].
\end{aligned}$$

$\mathbf{E}[\overline{g}_{ji} g_{ki} \overline{g}_{ml} g_{nl}]$ will be non-zero only when terms pair up as $|g|^4$ or $|g_i|^2 |g_j|^2$. This will occur in the following instances:

- $j = k = m = n$, $i = l$,
- $j = k = m = n$, $i \neq l$,
- $j = k \neq m = n$,

- $j = n \neq k = m$, $i = l$.

Using the Kronecker delta to indicate non-zero terms when all its subscripts are equal, we have

$$
\begin{aligned}
\mathbf{E}_X(\mathrm{tr}G^*PG)^2 &= \sum_{j,k,m,n} p_{jk}p_{mn} \sum_{i,l} \mathbf{E}_X\left[\overline{g}_{ji}g_{ki}\overline{g}_{ml}g_{nl}\right] \\
&= \sum_{j,k,m,n} p_{jk}p_{mn} \sum_{i,l} (2\delta_{jkmn}\delta_{il} + \delta_{jkmn}(1-\delta_{il}) + \delta_{jk}\delta_{mn}(1-\delta_{km}) \\
&\qquad\qquad + \delta_{jn}\delta_{km}(1-\delta_{kn})\delta_{il}) \\
&= \sum_{j,k,m,n} p_{jk}p_{mn}\left[2\delta_{jkmn}N + \delta_{jkmn}(N^2-N) + \delta_{jk}\delta_{mn}(1-\delta_{km})N^2\right. \\
&\qquad\qquad \left. + \delta_{jn}\delta_{km}(1-\delta_{kn})N\right] \\
&= 2N\sum_j p_{jj}^2 + (N^2-N)\sum_j p_{jj}^2 + N^2\sum_{j,m} p_{jj}p_{mm} - N^2\sum_j p_{jj}^2 \\
&\quad + N\sum_{j,k} p_{jk}p_{kj} - N\sum_j p_{jj}^2 \\
&= N^2\sum_{j,m} p_{jj}p_{mm} + N\sum_{j,k} p_{jk}p_{kj} \\
&= N^2(\mathrm{tr}P)^2 + N\mathrm{tr}P^2,
\end{aligned}
$$

completing the proof.

□

The following lemma was used when considering a Gaussian modulation scheme in Section 2.4.1.

Lemma 3. *Let S be a $T \times M$ matrix with independent zero-mean unit-variance complex Gaussian entries. Then $\mathbf{E}\mathrm{tr}(SS^*)^2 = MT(M+T)$.*

Proof. Denote the (i,j)th entry of S by s_{ij}. Then as the *pdf* of the Gaussian matrix S is

$p(S) = \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}}$, we have

$$\begin{aligned}
\mathbf{E}\mathrm{tr}(S^*S)^2 &= \int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} \mathrm{tr}(SS^*)^2 \, dS \\
&= \int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} \sum_{i,j,k,l} s_{ij}\overline{s}_{kj}s_{kl}\overline{s}_{il} \, dS \\
&= \sum_{i,j,k,l} \int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} s_{ij}\overline{s}_{kj}s_{kl}\overline{s}_{il} \, dS.
\end{aligned}$$

In this summation the indices i and k each range from 1 to T while the indices j and l each range from 1 to N. If both $i \neq k$ and $j \neq l$ the integral $\int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} s_{ij}\overline{s}_{kj}s_{kl}\overline{s}_{il} \, dX = 0$ as the integrand is an odd function of the variable s_{ij}.

Using the elementary integrals $\int \frac{e^{-|s|^2}}{\pi} \, ds = \int \frac{e^{-|s|^2}}{\pi} |s|^2 \, ds = 1$, $\int \frac{e^{-|s|^2}}{\pi} |s|^4 \, ds = 2$ where the integrals are over $s \in \mathbb{C}$, we have

$$\begin{aligned}
\int \frac{e^{-\mathrm{tr} SS^*}}{\pi^{MT}} \mathrm{tr}(S^*S)^2 \, dS &= \sum_{\substack{i,j,l \\ j \neq l}} \int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} |s_{ij}|^2 |s_{il}|^2 \, dS + \sum_{\substack{i,j,k \\ i \neq k}} \int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} |s_{ij}|^2 |s_{kj}|^2 \, dS \\
&\quad + \sum_{i,j} \int \frac{e^{-\mathrm{tr} S^* S}}{\pi^{MT}} |s_{ij}|^4 \, dS \\
&= \sum_{\substack{i,j,l \\ j \neq l}} 1 + \sum_{\substack{i,j,k \\ i \neq k}} 1 + \sum_{i,j} 2 \\
&= TM(M-1) + TM(T-1) + 2TM \\
&= MT(M+T),
\end{aligned}$$

as required. $\square$

Chapter 3

High-SNR Analysis through the Diversity-Multiplexing Gain Trade-off

3.1 Introduction

In the last chapter we saw that when the channel is unknown to both transmitter and receiver, a MIMO channel performs poorly with a mutual information only quadratic in the SNR. The real benefit of MIMO communications is at moderate to high SNR, where capacity scales more favorably [16, 66].

We again assume the channel is unknown to the transmitter, and consider how the transmitter should best deal with the fading channel. If the transmitter knew the fading characteristics, it could alter the rate at which it is sending symbols accordingly, and thus transmit more reliably. Unlike last time, we will assume the receiver does have channel knowledge—at higher SNR levels channel estimates become more reliable, and so this becomes a reasonable assumption. While previously we were interested in mutual information and capacity, here we measure performance by *error probability* as a function of SNR.

Since the transmitter does not know the channel state, it will adopt a coding strategy that should work well on average over all channel instantiations. However there will always be some non-zero probability that the fade is sufficiently bad that the channel cannot support a given data rate—such a channel state is called *outage*. The outage probability is one of the ways in which a coding scheme can be evaluated.

To provide some intuition, consider a point-to-point scalar fading channel:

$$x = \sqrt{P}sh + v, \quad x, s, h, v \in \mathbb{C}, P > 0, \mathbf{E}|s|^2 = 1, v \sim \mathbb{C}\mathcal{N}(0, 1). \tag{3.1}$$

Here $\sqrt{P}s$ is the complex number transmitted over the channel where it is subject to a fade h as well as additive noise v before arriving as x at the receiver. The receiver knows h and so can determine s from x by maximum-likelihood decoding: find the symbol $\hat{s}$ which minimizes $|x - \sqrt{P}\hat{s}h|$.

Figure 3.1 shows how an error can occur in such a channel. In the left side of the figure the center of each circle represents a codeword (element of $\mathbb{C}$) having average power P. Suppose the symbol represented by the filled circle is sent. We see that if h has small magnitude (of order $1/\sqrt{P}$), the presence of noise means that the received symbol could lie anywhere in the dashed circle shown in the right side of the figure. Since the received signal space has shrunk due to h, this dashed circle includes other symbols, so in this case one symbol is likely to be mistaken for another, leading to an error. The outage probability is related to that of $|h|$ being sufficiently small.

Unfortunately, in the MIMO case, the probability of outage is difficult to obtain analytically for a given SNR, since the channel can be poorly conditioned in many ways. One approximate analysis that has gained a lot of attention in recent years, has been to look at how error probability scales in the *high SNR regime*. In the seminal work of Zheng and Tse [82] the *diversity-multiplexing gain trade-off* has been used to evaluate the performance of

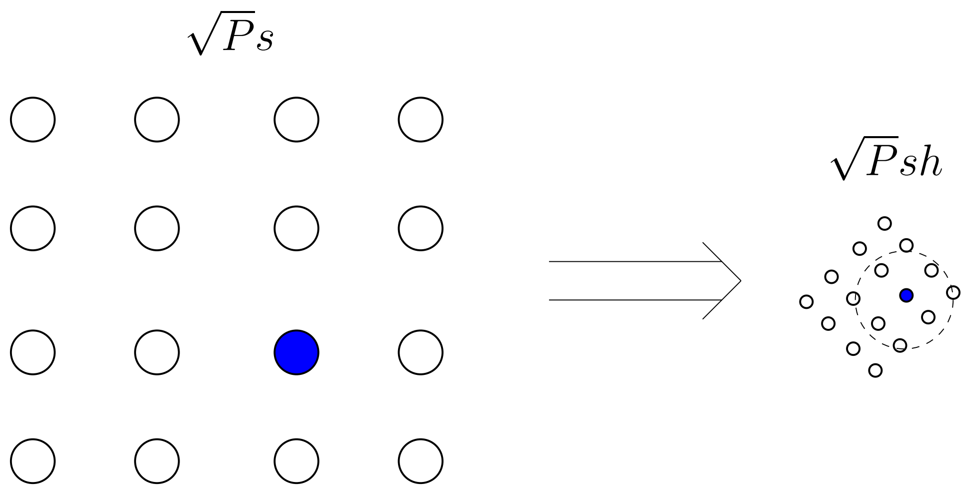

Figure 3.1: Fading plus noise may result in any of the symbols shown in the dashed circle being mistaken for the filled dot.

MIMO systems and codes. This studies the relationship between the two main advantages of MIMO communication: increased data rate and reliability. This analysis can be extended from point-to-point systems, to systems with relay nodes. The nodes improve reliability by providing additional paths between transmitter and receiver. That way, if some paths undergo poor fading, alternative paths can still provide reliable communication.

We will begin by introducing the notions of diversity and multiplexing gain at high SNR and then summarize some existing results for MIMO channels. In the next chapter we will state results for systems with relays.

3.2 Diversity and Multiplexing Gain

Two typical error probability curves for two communication systems are shown in figure 3.2. The left curve gives superior error performance, since for a given SNR it has a lower probability of error. The definition of diversity is the negative slope of the error probability curve for asymptotically high SNR. The greater the rate of decay of error probability with

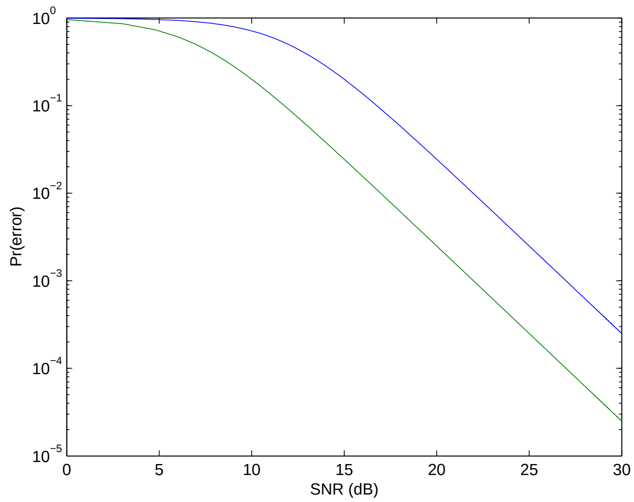

Figure 3.2: Error probability curves for two different communication systems. Both systems have the same diversity.

SNR, the higher the diversity and the more reliable that system is at high SNR. The two curves shown have the same diversity, so we see that at moderate SNR, diversity does not reveal the full story in performance. The SNR offset between the curves (known as *coding gain*) should also be considered. For MIMO and network problems however, this is far more difficult to calculate.

If there were no fading, and error events were only due to Gaussian noise, error probability decays exponentially in the SNR, this is due to the exponentially decaying tails of the Gaussian distribution (the probability of noise being large is very small). Fading severely degrades high SNR performance since the probability of error due to the channel being bad is far greater than that of noise being large. Figure 1.1 illustrates this.

Consider a MIMO system with M transmit and N receive antennas. As in the last chapter we have the model

$$X = \sqrt{P}SH + V. \tag{3.2}$$

At high SNR the capacity of a fast fading MIMO system, when the transmitter has no channel knowledge, is given by [66]

$$C(P) = \mathbf{E} \log\det\left(I + PHH^*\right) \to \min\{M, N\} \log P \quad \text{as } P \to \infty.$$

That is, capacity increases logarithmically in SNR, as in the well-known scalar case studied by Shannon. For a fixed rate (number of codewords) the probability of outage tends to zero as SNR tends to infinity, since the finitely many codewords simply grow arbitrarily far apart. If instead we allow the number of codewords to grow with SNR, thus keeping codewords spaced closer together, the outage probability becomes quantifiable and different codes and systems can be compared. One code outperforms another if it makes better use of the degrees of freedom (number of signal dimensions) available to it. The *multiplexing gain* measures how the rate increases with $\log P$. It can be thought of as the effective number of dimensions utilized by the code. We consider a family of codes indexed by the SNR P and say the scheme has multiplexing gain r if the data rate (number of symbols per channel use) scales as $r \log P$:

$$r = \lim_{P \to \infty} \frac{\text{rate}}{\log P}. \tag{3.3}$$

This can be thought of as a normalized data rate. The maximum r can be interpreted as the number of dimensions made available for the code.

The Zheng-Tse definition of *diversity* is for this scheme of codes, rather than a fixed code: at rate r the scheme has diversity $d(r)$ if

$$d(r) = \lim_{P \to \infty} \frac{\log \Pr(\text{error})}{\log P}. \tag{3.4}$$

Throughout this work we shall use exponential equality notation and say $\Pr(\text{error}) \doteq P^{-d(r)}$ to mean (3.4).

The traditional definition of diversity and error exponent analysis looks at fixed codes, that is, the point $d(0)$, when the multiplexing gain does not grow with P. Here we will view d as a function of r and since reliability decreases as rate increases, there is a natural trade-off between the two. Among all coding schemes, there exists an optimal trade-off, one that gives the best diversity performance for a given multiplexing gain. Zheng and Tse provide this for the MIMO case: what makes this possible is that in the high SNR regime error events are dominated by outage (attributed to the statistics of the channel). Hence one is interested in the probability that a channel is poorly conditioned.

Next we will look at this trade-off for a few known examples. A useful description of this work is given in the textbook by Tse and Viswanath [68].

3.2.1 Scalar Channel

Consider the scalar fading channel (3.1) where $h \sim \mathbb{C}\mathcal{N}(0,1)$. Then for a target data rate $r \log P$ the outage probability is given by

$$\begin{aligned}
\Pr(\text{outage}) &= \Pr(\log(1+|h|^2 P) < r \log P) \\
&= \Pr\left(|h|^2 < \frac{P^r - 1}{P}\right) \\
&\approx \frac{1}{P^{1-r}},
\end{aligned} \tag{3.5}$$

since for an exponential distribution

$$\Pr(|h|^2 < \epsilon) = 1 - \exp(-\epsilon) \approx \epsilon. \tag{3.6}$$

It is equation (3.5) which establishes why for Rayleigh channels diversity can be thought of as the number of independent paths between transmitter and receiver. A single Rayleigh fading path has diversity at most one. If there are k independent Rayleigh fading paths between transmitter and receiver (as can occur in a MIMO system), the probability of all paths being in outage is a product of the individual outage probabilities, so the diversity adds to k.

It can be shown that $\Pr(\text{error}) \doteq \Pr(\text{outage})$, and so from (3.5) $d(r) = 1 - r$ for $0 < r < 1$ is the optimal trade-off. It can be shown [68] that uncoded QAM achieves the optimal diversity and multiplexing gain (i.e., has this optimal trade-off curve).

3.2.2 MIMO Channel

This idea generalizes when we consider vector channels, except this time we have an $M \times N$ random channel *matrix* H corresponding to M transmit and N receive antennas. Consider the MIMO system model (3.2) where H has independent $\mathbb{C}\mathcal{N}(0, 1)$ entries (Rayleigh fading). The generalization of $|h|$ being small is that some of the eigenvalues of H are small—this means the channel squeezes the constellation of codewords in certain directions, making them difficult to distinguish once perturbed by noise.

We have

$$
\begin{aligned}
\Pr(\text{outage}) &= \Pr\left(\log\det(I + PHH^*) < r\log P\right) \\
&= \Pr\left(\prod_{i=1}^{\min\{M,N\}} (1 + P\lambda_i) < P^r\right) \\
&\quad \text{where } \lambda_1 > \ldots > \lambda_{\min\{M,N\}} \text{ are the ordered eigenvalues of } HH^* \\
&= \Pr\left(\prod_{i=1}^{\min\{M,N\}} (1 + PP^{-\alpha_i}) < P^r\right) \quad \text{where } \lambda_i = P^{-\alpha_i} \\
&= \Pr\left(\prod_{i=1}^{\min\{M,N\}} P^{(1-\alpha_i)^+} < P^r\right) \quad \text{where } (x)^+ = \max\{x, 0\} \\
&= \Pr\left(\sum_{i=1}^{\min\{M,N\}} (1-\alpha_i)^+ < r\right) \\
&= \int_{\alpha:\sum_i (1-\alpha_i)^+ < r} p(\alpha)\, d\alpha \qquad (3.7) \\
&\doteq \int_{\substack{\alpha:\sum_i (1-\alpha_i)^+ < r \\ \alpha_1 \geq \ldots \geq \alpha_{\min\{M,N\}}}} P^{-\sum_i (2i-1+|M-N|)\alpha_i}\, d\alpha, \qquad (3.8)
\end{aligned}
$$

where the last step uses knowledge of the joint eigenvalue distribution of HH^*, when H has independent $\mathbb{C}\mathcal{N}(0, 1)$ entries.

By Laplace's method [84], a result we will be making frequent use of, this integral may be approximated for large P by $P^{-d(r)}$, where

$$
d(r) = \inf_{\substack{\alpha:\sum_i (1-\alpha_i)^+ < r \\ \alpha_1 \geq \ldots \geq \alpha_{\min\{M,N\}}}} \sum_{i=1}^{\min\{M,N\}} (2i - 1 + |M - N|)\alpha_i.
$$

After solving this optimization problem, the optimal diversity-multiplexing gain trade-off in this case is given by the piecewise linear curve joining the points $(k, (M - k)(N - k))$ $k = 0, 1, \ldots, \min\{M, N\}$ [82]. This is plotted in figure 3.3 for $M = N = 2$. For fixed rate codes ($r = 0$) the classical MN diversity value is obtained. This can be achieved when each symbol is allowed to pass through all MN paths between transmitter and receiver. The maximum multiplexing gain corresponds to full use of the dimensions of the signal space: it is equal to the rank of H.

For more general distributions of H, when the distribution of HH^* is not invariant under similarity transformations, the probability density function of the channel matrix itself is easier to find than the distribution of its eigenvalues. In this case, instead of using (3.7), which depends on the eigenvalue distribution, we can proceed as follows.

3.2.2.1 Probability of Outage Formula in Terms of Distribution of Channel Matrix

Let $W = HH^*$. Writing the eigenvalue decomposition $W = U^*\Sigma U$, where U is unitary and Σ is diagonal with entries $\sigma_1, \sigma_2, \ldots, \sigma_{\min\{M,N\}}$, we have the change of variable [13]

$$dW = (\det V_\Sigma)^2 \; d\Sigma \; dU, \tag{3.9}$$

where $\det V_\Sigma$ is the Vandermonde determinant $\prod_{i<j}(\sigma_i - \sigma_j)$.

Let $Z = -\log W/\log P := -U^*(\log \Sigma)U/\log P$ where

$$\log \Sigma = \mathrm{diag}(\log \sigma_1, \log \sigma_2, \ldots, \log \sigma_{\min\{M,N\}}).$$

Let $\Lambda = -\log \Sigma/\log P$ so that $Z = U^*\Lambda U$ and $\Sigma = P^{-\Lambda}$. Similar to (3.9) we have the change of variable

$$dZ = (\det V_\Lambda)^2 \; d\Lambda \; dU, \tag{3.10}$$

where $\det V_\Lambda$ is the Vandermonde determinant corresponding to Λ.

Also, from $\Sigma = P^{-\Lambda}$, we can show

$$d\Sigma = (\log P)^{\min\{M,N\}} \det \Sigma \; d\Lambda. \tag{3.11}$$

Finally, we have

$$W = U^* \Sigma U = U^* P^{-\Lambda} U = P^{-Z}, \tag{3.12}$$

since $Z = U^* \Lambda U$. Combining (3.9)–(3.11) gives us the change of variable from W to Z given $W = P^{-Z}$:

$$dW = \frac{(\det V_\Sigma)^2 (\log P)^{\min\{M,N\}} \det \Sigma}{(\det V_{\log \Sigma / \log P})^2} \, dZ. \tag{3.13}$$

This allows us to redo the calculation of (3.8) in terms of an integral over the entries of W rather than its eigenvalues:

$$\begin{aligned}
\Pr(\text{outage}) &= \Pr(\log \det(I + PW) < r \log P) \\
&= \Pr\left(\det(I + PW) < P^r\right) \\
&= \Pr\left(\det(I + PP^{-Z}) < P^r\right) \\
&= \Pr\left(\det(I + PP^{-Z}) < P^r\right) \\
&= \Pr\left(\det(I + P^{I-Z}) < P^r\right) \\
&= \Pr\left(\det P^{(I-Z)^+} < P^r\right) \\
&= \Pr\left(\operatorname{tr}(I - Z)^+ < r\right),
\end{aligned}$$

where, if $W = U^* \Sigma U$, and $\lambda_1, \ldots, \lambda_{\min\{M,N\}}$ are the eigenvalues of Z, then

$$(I - Z)^+ := U^* \operatorname{diag}((1 - \lambda_1)^+, (1 - \lambda_2)^+, \ldots, (1 - \lambda_{\min\{M,N\}})^+) U.$$

Using (3.13), this gives us

$$
\begin{aligned}
\Pr(\text{outage}) &= \int_{W:\log\det(I+PW)<r\log P} p(W)\, dW \\
&= \int_{Z:\mathrm{tr}(I-Z)^+<r} p(W=P^{-Z})\frac{(\log P)^{\min\{M,N\}}}{(\det V_{\log\Sigma/\log P})^2}(\det V_\Sigma)^2 \det\Sigma\; dZ \\
&\doteq \int_{Z:\mathrm{tr}(I-Z)^+<r} p(W=P^{-Z})(\det V_\Sigma)^2 \det\Sigma\; dZ \\
&\quad \text{(since the logarithmic terms do not contribute to the exponent of } P) \\
&= \int_{Z:\mathrm{tr}(I-Z)^+<r} p(W=P^{-Z})(\det V_\Sigma)^2 P^{-\mathrm{tr}Z}\; dZ \qquad (3.14) \\
&\quad \text{(since } \det\Sigma = \det W = \det P^{-Z} = P^{-\mathrm{tr}Z} \text{ using (3.12)).}
\end{aligned}
$$

Equation (3.14) allows us to find the outage behavior of $W = HH^*$ via its distribution, without having to find the distribution of its eigenvalues Σ. For example, in the case when H has independent $\mathbb{C}\mathcal{N}(0,1)$ entries, $p(W) = Ke^{-\mathrm{tr}W}(\det W)^{M-N}$ [13] and after some manipulation, one can arrive at (3.8) from (3.14).

3.2.2.2 Achieving the Optimal Trade-off

The optimal trade-off curve is plotted in the case $M=2, N=2$ in figure 3.3, together with those of some particular coding schemes. Note that none of the schemes shown achieve the optimal diversity for all multiplexing gains. The original work [82] left open the question of codes that were trade-off optimal. Since then there have been many works addressing this problem. For example there are rotation-based constructions by Yao-Wornell and Dayal-Varanasi [79, 12], the Golden code by Belfiore et al. [4] (later generalized to perfect space-time codes [44]), lattice space-time (LAST) codes by El Gamal et al. [17], division-algebra based codes by Elia et al. [14], space-time trellis codes by Vaze and Sundar Rajan [71], and permutation codes by Tavildar and Viswanath [65].

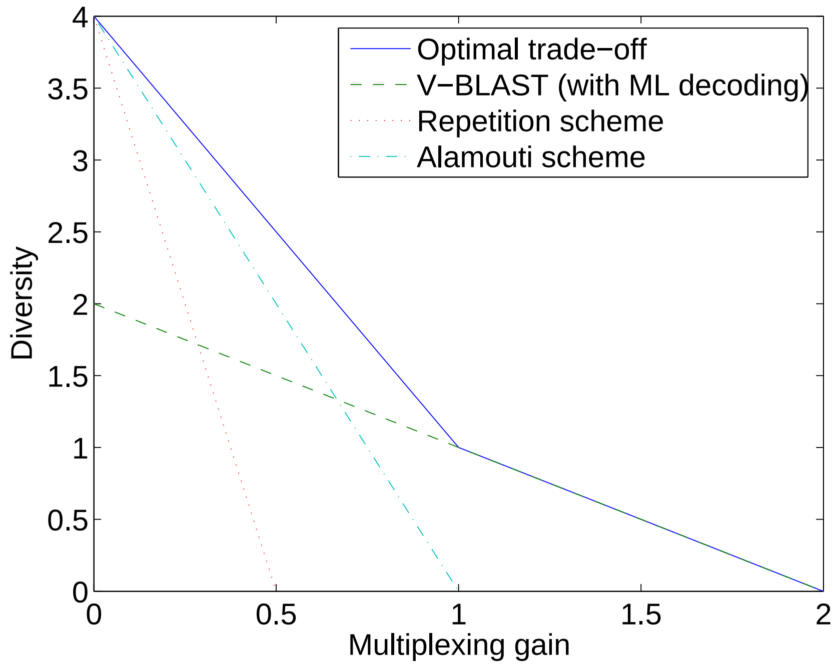

Figure 3.3: Diversity-multiplexing gain trade-off curves

The optimal trade-off has also been found for automatic retransmission request (ARQ) MIMO channels [18, 45]. Here one is allowed one bit of feedback from receiver to transmitter to indicate a success or failed transmission. Diversity can be increased for larger multiplexing gains, subject to the number of allowable rounds of retransmission.

In the next chapter we will see that the diversity-multiplexing trade-off also can be applied to systems with relays.

Chapter 4

A MIMO System with Relays

4.1 Introduction

The work of this chapter and the next belong to the recently studied area of *cooperative diversity*. The essential idea is that relay nodes can provide diversity by acting as a distributed MIMO system (see figure 4.1), setting up multiple independent paths between each transmitter and receiver. In this way errors will only occur in the rare event that all of the links are in a deep fade.

In this chapter we will find the diversity-multiplexing gain trade-off for a system with a single transmitter with M antennas, a single receiver with N antennas, and R single-antenna relay nodes to assist communication (see figure 4.4). Prior to introducing this system, we shall review some results on cooperative diversity and the diversity-multiplexing trade-off in the case of single-antenna systems with relays. We assume *half-duplex* communication, meaning that nodes cannot simultaneously transmit and receive information.

4.1.1 Relay Channels

Sendonaris et al. in [56, 57] introduced the idea of cooperative diversity in a wireless setting, showing how the capacity of users in the uplink of a cellular network (mobiles to base station)

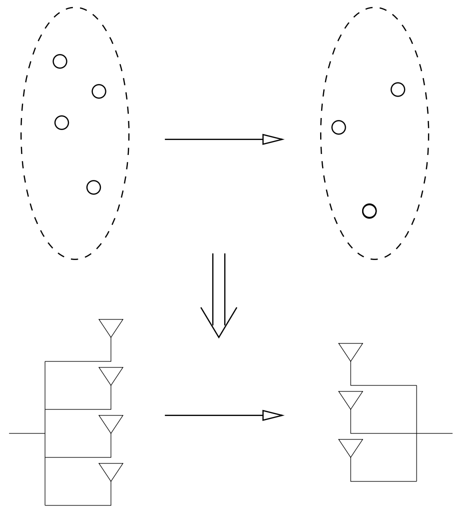

Figure 4.1: Nodes can cooperate to provide diversity, acting as a distributed MIMO system.

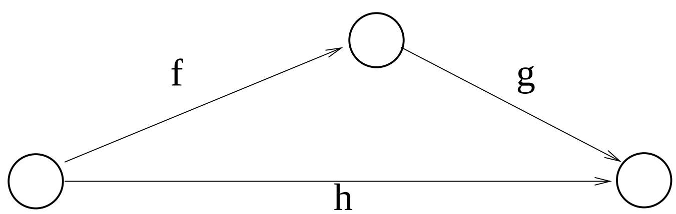

Figure 4.2: Wireless network with a single relay.

can be increased by cooperation.

Laneman et al. [33, 34, 35] compare several protocols by computing diversity through outage probability calculations. Suppose we have a single transmitter, receiver and relay, each having single antennas. Let f be the source-relay gain, g the relay-receiver gain and h the source-destination gain as shown in figure 4.2. The relay can either *amplify and forward* what it receives (thus amplifying additional noise received at the relay), or *decode and forward*.

In the amplify and forward case, the system is described by

$$y_1 = hA_1x_1 + v_1,$$

$$y_2 = fg\beta Bx_1 + hA_2x_2 + gBw + v_2.$$

Here $y_1, x_1, v_1 \in \mathbb{C}^{T'}$, $y_2, x_2, v_2 \in \mathbb{C}^{(T-T')}$ while A_1, B, A_2 are matrices of dimension $T' \times T'$, $(T - T') \times T'$ and $(T - T') \times (T - T')$ respectively. Both A_1 and A_2 are diagonal while B is the linear transformation performed by the relay on its received signals.

That is, for the first T' channel uses, the relay and receiver each obtain a faded version of the sent signal plus noise. For the remaining $(T - T')$ channel uses, the relay transmits an amplified version of what it has received to the receiver.

Laneman et al. look at the special case of $T' = T/2$, $A_1 = B = I_{T/2}$ and $A_2 = 0$. The diversity-multiplexing gain trade-off is given by $d(r) = 2(1 - 2r)$. A maximum diversity of two is obtained, but since only $T/2$ symbols are transmitted per T channel uses, the multiplexing gain is at most $1/2$.

Nabar et al. [41] consider $A_2 \neq 0$, that is, the source will continue to transmit directly to the receiver while the relay does the same. They show that this protocol outperforms that considered in [33]. The trade-off-optimal strategy is $d(r) = (1-r)+\max\{1-2r, 0\}$ obtained through the choice $T' = T/2$, $A_1 = A_2 = B = I_{T/2}$. Azarian et al. [3] generalize this to more than one relay (referring it to as a *non-orthogonal amplify and forward* protocol).

They also introduce a *dynamic decode and forward* procedure, whereby a relay waits a variable time before transmitting, depending on the current channel state. If the source-relay channel is in a good state, it will decode and retransmit sooner than if the channel were in a poor state. This protocol is trade-off optimal for $0 < r < 0.5$, achieving the

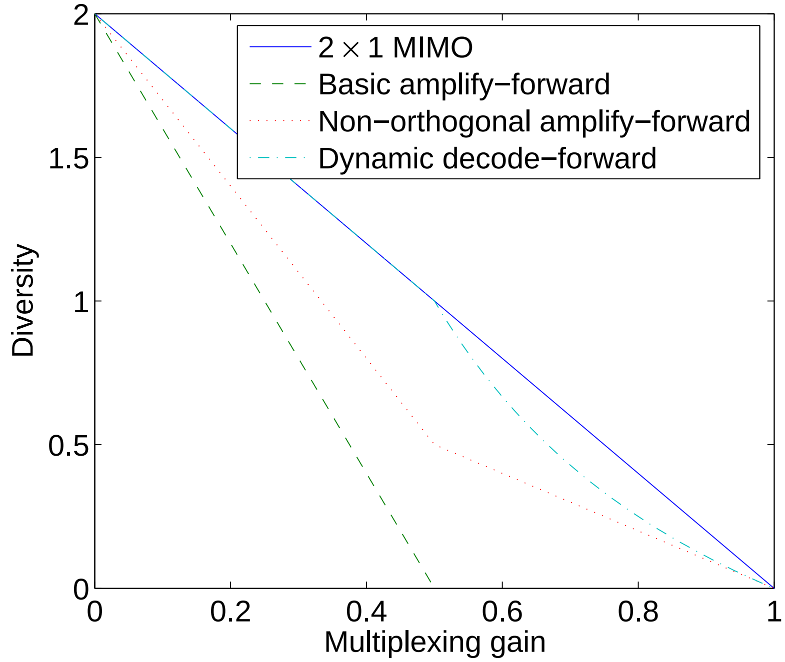

Figure 4.3: Trade-off curves for a single relay.

diversity of that of a 2×1 MIMO system (as though the relay and transmitter are in full cooperation). This result is generalized to where the transmitter and receiver have multiple antennas in [77].

The diversity-multiplexing gain trade-off curves for a single relay are plotted in figure 4.3. These schemes shown may be generalized to multiple relays. For low multiplexing gains at least, the dynamic decode and forward procedure has optimal trade-off. It remains an open problem to find the optimal trade-off behavior for larger multiplexing gains.

The issue of whether a relay node should amplify its received signal or decode has been studied extensively. Generally it is better to decode and forward when the source-to-relay channel is good and the signal-to-noise ratio at the relay is large [34]. This is done to suppress relay noise. However this means the rate one can achieve is restricted to the performance of the source-relay channel, without regarding the relay-destination channel.

In this work we assume the relays amplify and forward their signal, as we show that at high SNR, the amplified relay noise does not result in a significant degradation in diversity.

4.1.2 Rayleigh Product Channel

Yang and Belfiore [77] consider the diversity-multiplexing trade-off of a MIMO system where the channel matrix H is the product of two independent complex Gaussian matrices. That is, consider (3.2) with $H = H_1 H_2$ where H_1 and H_2 and independent complex Gaussian entries. Such a model takes into account instances of lower rank channel matrices. Suppose H_1 is $(m \times l)$, H_2 is $(l \times n)$ and let $M \leq N \leq L$ be the ordered permutation of l, m, n. The optimal trade-off curve is shown to be

$$d(r) = (M-r)(N-r) - \frac{1}{2}\left\lfloor \frac{[(M-L+N-r)^+]^2}{2} \right\rfloor, \quad \text{where } M \leq N \leq L.$$

This is upper bounded by $\min\{d_{H_1}(r), d_{H_2}(r)\}$ and coincides with it for large r.

The model we will work in this chapter reduces to the Rayleigh product channel in the special case that the relays do no processing of their received signal.

4.2 Model

We will compute the diversity-multiplexing trade-off curve for a relay network first described by Jing and Hassibi in [28] and shown in figure 4.4. This can be considered a generalization of [29] in which there were single-antenna nodes. In these works the pairwise error probability of the system was evaluated and conditions placed on the coding to achieve optimal performance. An explicit code construction is given in [42]. The single-antenna case has

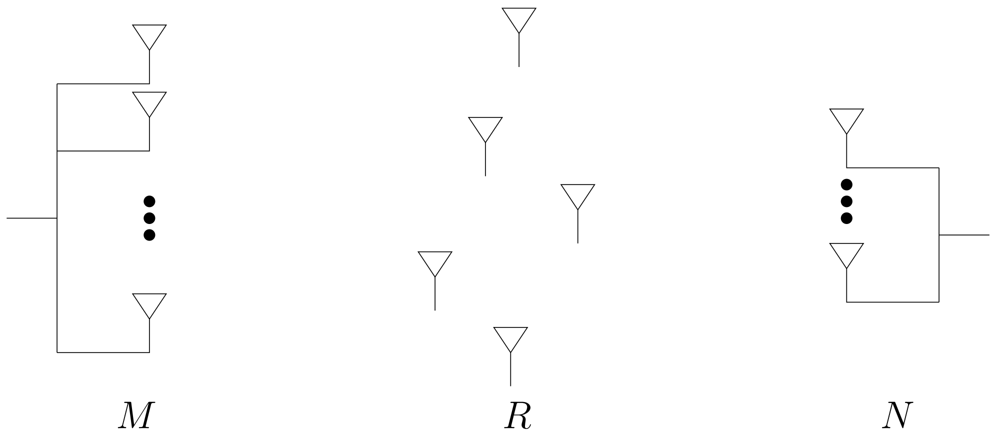

Figure 4.4: Relay network with multiple antennas at transmitter and receiver.

also been studied by Elia et al. [14] and Susinder Rajan and Sundar Rajan [48]. The coding used is known as *distributed space-time coding* as it is spread over the relays in the network. Codes that perform well even without channel knowledge at the transmitter have also been constructed independently by Jing and Jafarkhani [30], Oggier and Hassibi [43] and Kiran and Sundar Rajan [63].

We have a single transmitter with M antennas, a receiver with N antennas and R single-antenna relay nodes which assist communication. Let f_{ij} be the fading coefficient from the ith transmit antenna to the jth relay, and g_{jk} be the fading coefficient from the jth relay to the kth transmit antenna. As usual all fading coefficients are drawn from a $\mathbb{C}\mathcal{N}(0,1)$ distribution. Next define

$$f_i := \begin{bmatrix} f_{1i} \\ f_{2i} \\ \vdots \\ f_{Mi} \end{bmatrix}, \quad g_i := \begin{bmatrix} g_{i1} & g_{i2} & \dots & g_{iN} \end{bmatrix},$$

as column and row vectors corresponding to connections to the ith relay.

Let s be the $T \times M$ matrix sent by the transmitter over T transmission steps and its M antennas. That is, $s = [\ s_1 \quad s_2 \quad \ldots \quad s_M\ ]$, where s_i is the T-dimensional vector sent by the ith transmit antenna.

In the first T time steps, the transmitter sends s to the relay nodes. The receiver is assumed to be inactive at this stage. The ith relay node receives a faded version of s plus noise:

$$r_i = \sqrt{P} s f_i + v_i,$$

where v_i is a T-dimensional noise vector. Here we assume $\mathbf{E}||s||^2 = TM$, so P is a normalization constant proportional to the SNR at the receiver.

Next each relay left-multiplies its T-dimensional signal by a unitary matrix A_i (thus forming linear combinations of the rows, coding over time) and transmits the resulting vector to the receiver. The receiver obtains a $T \times N$ matrix: the sum of the R faded copies of s processed by the relay nodes received by N antennas over T further time steps:

$$\begin{aligned} X &= \sum_{i=1}^{R} A_i r_i g_i + w \\ &= \sqrt{P} \sum_{i=1}^{R} A_i s f_i g_i + \sum_{i=1}^{R} A_i v_i g_i + w \\ &= \sqrt{P} S H + W, \end{aligned} \tag{4.1}$$

where

$$S = \left[\ A_1 s \quad A_2 s \quad \ldots \quad A_R s\ \right], \quad H = \begin{bmatrix} f_1 g_1 \\ f_2 g_2 \\ \vdots \\ f_R g_R \end{bmatrix}, \quad W = \sum_{i=1}^{R} A_i v_i g_i + w. \tag{4.2}$$

Here S is a $T \times MR$ transmission matrix, H is an $MR \times N$ matrix describing the channel connections as R outer products $f_i g_i$ stacked together, and W is a $T \times N$ noise matrix. The noise sources v_i from each relay and w at the receiver are also assumed to be independent complex Gaussian random variables with zero mean and unit variance for each time step. Equation (4.1) has the same form as a space-time code for a MIMO system, first described in the previous chapter. As before we assume only the receiver knows H, while the transmitter and relays do not have any channel knowledge. The difference between this system and a MIMO system is that there is no cooperation amongst the relays, and none of them have direct access to s. The code to describe s and the unitary matrices A_i form what is known as a *distributed space-time code* [28].

We will assume T is sufficiently large for codewords to be allowed to be spaced appropriately distant from each other (later on a scheme with $T = MR$ achieving the trade-off will be proposed). This will allow an error event to be outage dominated.

The model differs from that of [29] in that we do not concern ourselves with power allocation at the relays. Any constants of proportionality become absorbed in the forthcoming high-SNR analysis.

4.3 Diversity-Multiplexing Gain Trade-off of a MIMO System with Relays

To carry out the trade-off analysis we proceed as follows.

1. Find the probability of outage taking into account there is additional relay noise. We claim that at high SNR the outage behavior is the same as though there were no relay noise.

2. Apply the results of [77] to find the exponent of the probability density function of the joint eigenvalue distribution of the channel matrix. For this particular channel, the distribution of the channel itself is of similarly difficulty to compute, so we apply (3.8) instead of (3.14).

3. Optimize this exponent subject to constraints that prevent exponential decay of the pdf—this gives the dominant behavior of the integral by Laplace's method. We assume $M \geq N$ initially.

4. We argue that the case $M < N$, the optimal trade-off can be obtained by disregarding $N - M$ receive antennas.

The maximum diversity of this scheme ($r = 0$) is already known [28] to be $R \min\{M, N\}$. The maximum multiplexing gain will be determined by the maximum dimension of the signal space that can be exploited: $\min\{M, N, R\}$. In the analysis we now will see what happens for intermediate values of r.

We remark that since this is a two stage protocol and that the direct link between transmitter and receiver is not used, all multiplexing gains obtained should strictly be divided by 2, since two channel uses are used to transmit information. Adopting the direct path for relay networks with slightly different models are considered in [3], [78] and [48] amongst others. Adopting lines similar to Yang and Belfiore [78], we conjecture the following, adopting a non-orthogonal amplify and forward protocol (see Sec. 4.1.1).

Conjecture 1. *If $d(r)$ is the trade-off curve we obtain for our system in the following analysis, and $d_1(r)$ that of the direct path between transmitter and receiver (a MIMO link),* *then the overall diversity is lower-bounded by*

$$d_{NAF}(r) \geq d_1(r) + d(2r).$$

The first term here corresponds to the source-destination link used all the time; the second to the source-relays-destination system used at half the rate.

4.3.1 Outage Calculation

Since the equivalent noise W in (4.2) consists of Gaussian noise amplified by g_i we need to verify that it does not adversely impact the diversity-multiplexing gain trade-off. First we determine the noise covariance. Considering the kth row of W for $k = 1, \ldots, T$ we have from (4.2) $W_k = \sum_{i=1}^{N}(A_i v_i)_k g_i + w_k$, where $(A_i v_i)_k$ and w_k denote the kth row of $A_i v_i$ and w respectively. Then given g_i,

$$\begin{aligned}
\mathbf{E}_{v,w} W_k^* W_l &= \mathbf{E}_{v,w} \left(\sum_{i=1}^{N} (A_i v_i)_k g_i + w_k \right)^* \left(\sum_{j=1}^{N} (A_j v_j)_l g_j + w_l \right) \\
&= \mathbf{E}_w w_k^* w_l + \sum_{i=1}^{R} \sum_{j=1}^{R} g_i^* \mathbf{E}_v [((A_i v_i)_k)^* (A_j v_j)_l] g_j \\
&\quad \text{since } w_k \text{ and } v_j \text{ are independent vectors} \\
&= \delta_{kl} I_N + \sum_{i=1}^{R} \sum_{j=1}^{R} g_i^* \mathbf{E}_v [(A_i^{(k)} v_i)^* (A_j^{(l)} v_j)] g_j \\
&\quad \text{where } A_i^{(k)} \text{ denotes the } k\text{th row of } A_i \\
&= \delta_{kl} I_N + \sum_{i=1}^{R} \sum_{j=1}^{R} g_i^* \mathbf{E}_v [v_i^* A_i^{(k)*} A_j^{(l)} v_j] g_j \\
&= \delta_{kl} I_N + \sum_{i=1}^{R} g_i^* \mathbf{E}_v [v_i^* A_i^{(k)*} A_i^{(l)} v_i] g_i \\
&\quad \text{since } v_i \text{ and } v_j \text{ are independent for } i \neq j.
\end{aligned}$$

Let $B_{kl} = A_i^{(k)*} A_i^{(l)}$ with $(k,m)th$ entry $B_{kl,jm}$. Then

$$\mathbf{E}_v[v_i^* B_{kl} v_i] = \mathbf{E}_v \sum_{j,m} \overline{v}_{ij} B_{kl,jm} v_{im} = \sum_{j,m} B_{kl,jm} \mathbf{E}_v[\overline{v}_{ij} v_{im}] = \sum_{j,m} B_{kl,jm} \delta_{jm} = \mathrm{tr} B_{kl}.$$

Hence

$$\begin{aligned}
\mathbf{E}_{v,w} W_k^* W_l &= \delta_{kl} I_N + \sum_{i=1}^R g_i^* \mathrm{tr}(A_i^{(k)*} A_i^{(l)}) g_i \\
&= \delta_{kl} I_N + \sum_{i=1}^R g_i^* \mathrm{tr}(A_i^{(l)} A_i^{(k)*}) g_i \\
&= \delta_{kl} I_N + \sum_{i=1}^R g_i^* A_i^{(l)} A_i^{(k)*} g_i \\
&= \delta_{kl} I_N + \sum_{i=1}^R g_i^* \delta_{kl} g_i \quad \text{since } A_i\text{'s are unitary} \\
&= \delta_{kl}(I_N + G^* G) \quad \text{where } G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_R \end{bmatrix}.
\end{aligned}$$

Note that this is the same for all time steps $k = 1, \ldots, T$. Hence the instantaneous mutual information between transmitter and receiver is given by [66]

$$I(X; S) = \log\det(I_N + P(I_N + G^* G)^{-1} H^* R_S H), \tag{4.3}$$

where R_S is the $MR \times MR$ input covariance matrix of X for each time step.

We remark that S given by (4.2) has a particular structure, namely its columns span the M-dimensional space generated by the columns of s. It is known that with the channel unknown to the transmitter the mutual information (4.3) is maximized when $R_S = I_{MR}$.

Hence we make that substitution here. Although S has rank M, R_S takes an average over all realizations of S and so is allowed to have full rank MR.

Since G is a Gaussian matrix, the probability that an eigenvalue of $I + G^*G$ is large is exponentially small. That is, $\lambda_{\max}(G^*G) \leq \mathrm{tr}G^*G$ which is a χ^2 random variable with $2N^2$ degrees of freedom (equivalently a sum of N^2 exponential random variables). Hence it has an exponentially decaying tail and its probability of being large is exponentially small:

$$(1 + \lambda_{\max}(GG^*))^{-1} \doteq P^0.$$

We therefore have justified the following sequence of steps:

$$\begin{aligned}
& P_{\text{outage}}(r \log P) \\
= \quad & \min_{R_S : \mathrm{tr} R_S \leq MR} \Pr\left(I(X;S) < r \log P\right) \\
= \quad & \min_{R_S : \mathrm{tr} R_S \leq MR} \Pr\left(\log\det(I_N + R_W^{-1} H^* R_S H) < r \log P\right) \\
\leq \quad & \Pr\left(\log\det(I_N + R_W^{-1} H^* H) < r \log P\right) \\
\doteq \quad & \Pr\left(\log\det(I_N + H^* H) < r \log P\right).
\end{aligned}$$

To proceed with the calculation of this probability we first observe that

$$
\begin{aligned}
H^*H &= [(f_1g_1)^* \ldots (f_Rg_R)^*] \begin{bmatrix} f_1g_1 \\ \vdots \\ f_Rg_R \end{bmatrix} \\
&= \sum_{i=1}^{R} g_i^* f_i^* f_i g_i \\
&= \sum_{i=1}^{R} g_i^* ||f_i||^2 g_i \\
&= G^*DG \quad \text{where } D = \mathrm{diag}(||f_1||^2, \ldots, ||f_R||^2).
\end{aligned}
$$

This shows that the rank of H is the smaller of the ranks of G and D: $\min\{R, N\}$ which is independent of M. For the remainder of the analysis we will assume $M \geq N$. At the end of the section we will comment on the $M < N$ case.

We proceed along lines similar to recent work by Yang and Belfiore [77] which was applied to the Rayleigh product channel. The non-zero eigenvalues of $H^*H = G^*DG$ are the same as those of QQ^* where $Q_{R\times N} = D^{1/2}G$. We know that given D, QQ^* has a *Wishart* distribution $QQ^* \sim \mathcal{W}_R(N, D)$. That is, the columns of Q are zero-mean independent complex Gaussian vectors having covariance matrix D.

The distribution of the ordered eigenvalues $\lambda_1 > \ldots > \lambda_{\min\{N,R\}}$ of QQ^* given D is given by the following lemma, applying a result in [70].

Lemma 4. *Let $\mu_1^2 > \ldots > \mu_R^2$ be the ordered values of $||f_1||^2, \ldots, ||f_R||^2$, the diagonal elements of D, assumed to be distinct. The distribution of the ordered eigenvalues $\lambda_1 > \ldots > \lambda_{\min\{N,R\}}$ of $QQ^* = D^{1/2}GG^*D^{1/2}$ for fixed D and for G a random $R \times N$ matrix* *with independent $\mathbb{C}\mathcal{N}(0,1)$ entries, is given by*

$$p(\lambda|D) \quad = \quad \begin{cases} K_{R,N}\det[e^{-\lambda_j/\mu_i^2}]\prod_{i=1}^R \mu_i^{2(R-N-1)}\lambda_i^{N-R}\prod_{i<j}^R \frac{\lambda_i-\lambda_j}{\mu_i^2-\mu_j^2}, & N \geq R, \\ \\ G_{R,N}\det[\Xi]\prod_{i<j}^R \frac{1}{\mu_i^2-\mu_j^2}\prod_{i<j}^N(\lambda_i-\lambda_j), & N < R, \end{cases} \tag{4.4}$$

where $K_{R,N}$ and $G_{R,N}$ are normalization constants, $[e^{-\lambda_j/\mu_i^2}]$ is a $(\min\{R,N\}\times\min\{R,N\})$ matrix with (i,j) entry $e^{-\lambda_j/\mu_i^2}$, and

$$\Xi = \begin{bmatrix} 1 & \mu_1^2 & \ldots \mu_1^{2(R-N-1)} & \mu_1^{2(R-N-1)}e^{(-\lambda_1/\mu_1^2)} & \ldots & \mu_1^{2(R-N-1)}e^{(-\lambda_N/\mu_1^2)} \\ \vdots & & \vdots & & \vdots & \\ 1 & \mu_R^2 & \ldots \mu_R^{2(R-N-1)} & \mu_R^{2(R-N-1)}e^{(-\lambda_1/\mu_R^2)} & \ldots & \mu_R^{2(R-N-1)}e^{(-\lambda_N/\mu_R^2)} \end{bmatrix}. \tag{4.5}$$

With each entry being the sum of M independent exponential random variables, the joint distribution of the diagonal entries of D is

$$p(\mu) := p(\mu_1^2,\ldots,\mu_R^2) = \prod_{i=1}^R \frac{e^{-\mu_i^2}\mu_i^{2(M-1)}}{(M-1)!}. \tag{4.6}$$

Define

$$\alpha_i = \frac{-\log\lambda_i}{\log P}, \quad \beta_i = \frac{-\log\mu_i^2}{\log P} \quad \Rightarrow \quad \lambda_i = P^{-\alpha_i}, \mu_i^2 = P^{-\beta_i}, \tag{4.7}$$

and

$$p(\alpha,\beta) = p(\lambda,\mu).\prod_{i=1}^N\left|\frac{d\lambda_i}{d\alpha_i}\right|\prod_{i=1}^R\left|\frac{d\mu_i^2}{d\beta_i}\right| = p(\lambda,\mu)\prod_{i=1}^N(\log P)P^{-\alpha_i}\prod_{i=1}^R(\log P)P^{-\beta_i}. \tag{4.8}$$

Then after making this simple change of variables, and defining the vectors $\alpha = (\alpha_1, \ldots, \alpha_{\min\{N,R\}})$, $\beta = (\beta_1, \ldots, \beta_R)$, we may write the joint eigenvalue distribution of α and β from (4.4) to (4.8) as follows:

$$
\begin{aligned}
& p(\alpha, \beta) \\
= \; & p(\lambda, \mu)(\log P)^{N+R} \prod_{i=1}^{N} P^{-\alpha_i} \prod_{i=1}^{R} P^{-\beta_i} \quad \text{from (4.8)} \\
= \; & p(\mu) p(\lambda \mid \mu)(\log P)^{N+R} \prod_{i=1}^{N} P^{-\alpha_i} \prod_{i=1}^{R} P^{-\beta_i} \\
= \; & p(\lambda \mid \mu)(\log P)^{N+R} \prod_{i=1}^{N} P^{-\alpha_i} \prod_{i=1}^{R} P^{-\beta_i} \frac{\exp\left(-P^{-\beta_i}\right) P^{-(M-1)\beta_i}}{(M-1)!} \quad \text{using (4.6)–(4.7)} \\
= \; & \begin{cases}
C_{R,M,N} (\log P)^{N+R} \prod_{i=1}^{R} P^{-(N-R+1)\alpha_i} P^{-(R+M-N-1)\beta_i} \\
\quad \cdot \prod_{i<j} \frac{P^{-\alpha_i} - P^{-\alpha_j}}{P^{-\beta_i} - P^{-\beta_j}} \cdot \exp\left(-\sum_{i=1}^{R} P^{-\beta_i}\right) \det\left[\exp\left(-P^{-(\alpha_j - \beta_i)}\right)\right]_{i,j=1}^{R}, & N \geq R, \\
D_{R,M,N} (\log P)^{N+R} \prod_{i=1}^{N} P^{-\alpha_i} \prod_{i=1}^{R} P^{-M\beta_i} \prod_{i<j}^{R} \frac{1}{P^{-\beta_i} - P^{-\beta_j}} \\
\quad \cdot \prod_{i<j}^{N} (P^{-\alpha_i} - P^{-\alpha_j}) \cdot \exp\left(-\sum_{i=1}^{R} P^{-\beta_i}\right) . \det A, & N < R,
\end{cases}
\end{aligned}
\tag{4.9}
$$

where A is the transformation of the matrix Ξ in (4.5) under this change of variables (4.7), while $C_{R,M,N}$ and $D_{R,M,N}$ are normalization constants.

Note that from the exponential factor, $p(\alpha, \beta)$ will decay exponentially unless $\beta_i \geq 0$ for $i = 1, \ldots, R$.

The next task is to find the exponent of P in (4.9). To deal with the determinant factors we make use of the results proved by induction on the matrices' dimensions, in Lemmas 2 and 3 of [77]:

Lemma 5. *We have the following expressions for the exponential orders of the determinants* *in (4.9).*

1.

$$\det\left[\exp\left(-P^{-(\alpha_j-\beta_i)}\right)\right]_{i,j=1}^{R} \doteq \exp\left(-\sum_{i=1}^{R} P^{-(\alpha_i-\beta_i)}\right) P^{-\sum_{i<j}(\alpha_i-\beta_j)^+} \tag{4.10}$$

This determinant decays exponentially in P unless $\alpha_i \geq \beta_i$.

2.

$$\begin{aligned}\det A \quad &\doteq \quad \prod_{i=1}^{N+1} P^{-(R-N-1)\beta_i} \prod_{i=N+2}^{R} P^{-(R-i)\beta_i} \prod_{i=1}^{N}\prod_{j=N+1}^{R} P^{-(\alpha_i-\beta_j)^+} \\ &\quad \cdot \prod_{i<j}^{N} P^{-(\alpha_i-\beta_j)^+} \exp\left(-\sum_{i=1}^{N} P^{-(\alpha_i-\beta_i)}\right) \end{aligned} \tag{4.11}$$

For $i = 1, \ldots, N$ this decays exponentially in P unless $\alpha_i \geq \beta_i$ for $\beta_i \geq 0$.

Since $\lambda_i > \lambda_j$ and $\mu_i > \mu_j$ for $i < j$, we have $\alpha_i \leq \alpha_j$ and $\beta_i \leq \beta_j$ for $i < j$. Hence

$$\frac{P^{-\alpha_i} - P^{-\alpha_j}}{P^{-\beta_i} - P^{-\beta_j}} \doteq \frac{P^{-\alpha_i}}{P^{-\beta_i}} = P^{-(\alpha_i-\beta_i)}. \tag{4.12}$$

Combining (4.9), (4.10), (4.11) and (4.12), and using the fact that constants and powers of $(\log P)$ do not contribute to the exponential order allows us to find the exponential order $\epsilon(\alpha, \beta)$ of $p(\alpha, \beta)$. We will repeatedly use the identity

$$\sum_{i<j}^{R} a_i := \sum_{j=1}^{R}\sum_{i<j} a_i = (R-1)a_1 + (R-2)a_2 + \ldots + a_R = \sum_{i=1}^{R}(R-i)a_i.$$

For $N \geq R$,

$$\begin{aligned}
\epsilon(\alpha,\beta) &= \sum_{i=1}^{R}(N-R+1)\alpha_i + \sum_{i=1}^{R}(R+M-N-1)\beta_i + \sum_{i<j}^{R}(\alpha_i-\beta_i) + \sum_{i<j}(\alpha_i-\beta_j)^+ \\
&= \sum_{i=1}^{R}(N-R+1)\alpha_i + \sum_{i=1}^{R}(R+M-N-1)\beta_i + \sum_{i=1}^{R}(R-i)(\alpha_i-\beta_i) + \sum_{i<j}(\alpha_i-\beta_j)^+ \\
&= \sum_{i=1}^{R}(N+1-i)\alpha_i + \sum_{i=1}^{R}(M-N+i-1)\beta_i + \sum_{i<j}(\alpha_i-\beta_j)^+, \qquad (4.13)
\end{aligned}$$

while for $N < R$,

$$\begin{aligned}
\epsilon(\alpha,\beta) &= \sum_{i=1}^{N}\alpha_i + \sum_{i=1}^{R}M\beta_i + \sum_{i<j}^{N}(\alpha_i-\beta_i) + \sum_{j=N+1,i<j}^{R} -\beta_i + \sum_{i=1}^{N+1}(R-N-1)\beta_i \\
&\quad + \sum_{i=N+2}^{R}(R-i)\beta_i + \sum_{i=1}^{N}\sum_{j=N+1}^{R}(\alpha_i-\beta_j)^+ + \sum_{i<j}^{N}(\alpha_i-\beta_j)^+ \\
&= \sum_{i=1}^{N}\alpha_i + M\sum_{i=1}^{R}\beta_i + \sum_{i=1}^{N}(N-i)\alpha_i + \sum_{i=1}^{R}(R-i)(-\beta_i) + \sum_{i=1}^{N+1}(R-N-1)\beta_i \\
&\quad + \sum_{i=N+2}^{R}(R-i)\beta_i + \sum_{i=1}^{N}\sum_{j=N+1}^{R}(\alpha_i-\beta_j)^+ + \sum_{i<j}^{N}(\alpha_i-\beta_j)^+ \\
&= \sum_{i=1}^{N}(N-i+1)\alpha_i + \sum_{i=1}^{N+1}(R-N-1+M-R+i)\beta_i \\
&\quad + \sum_{i=N+2}^{R}(R-i+M-R+i)\beta_i + \sum_{i=1}^{N}\sum_{j=N+1}^{R}(\alpha_i-\beta_j)^+ + \sum_{i<j}^{N}(\alpha_i-\beta_j)^+ \\
&= \sum_{i=1}^{N}(N-i+1)\alpha_i + \sum_{i=1}^{N}(M-N+i-1)\beta_i + M\sum_{i=N+1}^{R}\beta_i \\
&\quad + \sum_{i=1}^{N}\sum_{j=N+1}^{R}(\alpha_i-\beta_j)^+ + \sum_{i<j}^{N}(\alpha_i-\beta_j)^+.
\end{aligned}$$

(4.14)

The diversity of our wireless scheme is then $\inf \epsilon(\alpha, \beta)$ where

$$\sum_{i=1}^{N}(1-\alpha_i)^+ < r, \tag{4.15}$$

$$\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{\min\{N,R\}}, \tag{4.16}$$

$$\beta_1 \leq \beta_2 \leq \ldots \leq \beta_R, \tag{4.17}$$

and

$$\alpha_i \geq \beta_i \geq 0 \quad \text{for } i = 1, \ldots, N. \tag{4.18}$$

Next we find this infimum by optimizing (4.13) and (4.14) over α and β.

4.3.2 Optimization over α and β

To minimize (4.13) and (4.14) subject to the constraints (4.15)-(4.18), we first fix $\alpha_1 \leq \ldots \leq \alpha_{\min\{N,R\}}$ satisfying (4.15) and find the optimal β. We will assume $M \geq N$.

4.3.2.1 Case 1: $N \geq R$

As done in [77] we begin with an initial configuration

$$0 \leq \beta_1 = \alpha_1 \leq \beta_2 = \alpha_2 \leq \ldots \leq \beta_R = \alpha_R$$

so that $\sum_{i<j}(\alpha_i - \beta_j)^+ = 0$.

The sum (4.13) is of the form $\sum_{i=1}^{R}(a_i\alpha_i + b_i\beta_i)$. With the initial configuration just described we would have $a_i = N+1-i$ and $b_i = M-N+i-1 \geq 0$. For some fixed j, if β_j is decreased below α_i for some $i < j$, the term $(\alpha_i - \beta_j)^+$ is no longer zero. Here a_i increases by 1, b_j decreases by 1 and the overall sum (4.13) decreases provided b_j remains positive.

How many α_i's should β_j cross? Denote this number by c_j. Since b_j decreases by 1 each time from an initial value of $M-N+j-1$, we have $c_j = \min\{j-1, M-N+j-1\} = j-1$ since $M \geq N$. In other words, for $M \geq N$ it is optimal for β_j to be decreased all the way to zero. Substituting this into (4.13) leads to

$$\begin{aligned}
\epsilon(\alpha, 0) &= \sum_{i=1}^{R}(N+1-i)\alpha_i + \sum_{i<j}^{R}\alpha_i \\
&= \sum_{i=1}^{R}(N+1-i)\alpha_i + \sum_{i=1}^{R}(R-i)\alpha_i \\
&= \sum_{i=1}^{R}(N+R+1-2i)\alpha_i
\end{aligned} \tag{4.19}$$

4.3.2.2 Case 2: $N < R$

Again we assume an initial configuration

$$0 \leq \beta_1 = \alpha_1 \leq \beta_2 = \alpha_2 \leq \ldots \leq \beta_N = \alpha_N \leq \beta_{N+1} \leq \ldots \leq \beta_R,$$

so that the last two terms of (4.14) are 0.

This time (4.14) is of the form $\epsilon(\alpha, \beta) = \sum_{i=1}^{N} a_i\alpha_i + \sum_{j=1}^{N} b_j\beta_j + \sum_{N+1}^{R} b_j\beta_j$. With the initial configuration, $a_i = N-i+1$, $b_j = M-N+j-1$. Proceeding similarly to Case 1, but treating the cases $j \leq N$ and $j > N$ separately, we see that if β_j is decreased below α_i for some $i < j$, a_i increases by 1 while b_j decreases by 1. The overall sum (4.14) decreases provided b_j remains positive. If c_j is the number of α_i's that b_j should cross, we have as before $c_j = \min\{j-1, M-N+j-1\} = j-1$. Hence β_j may cross all α_i's all the way to zero to minimize the sum. Substituting this into (4.14) leads to

$$\begin{aligned}
\epsilon(\alpha,0) &= \sum_{i=1}^{N}(N-i+1)\alpha_i + \sum_{i=1}^{N}\sum_{j=N+1}^{R}\alpha_i + \sum_{i<j}^{N}\alpha_i \\
&= \sum_{i=1}^{N}(N+1-i)\alpha_i + \sum_{i=1}^{N}(R-N)\alpha_i + \sum_{i=1}^{N}(N-i)\alpha_i \\
&= \sum_{i=1}^{N}(N+R+1-2i)\alpha_i. \qquad (4.20)
\end{aligned}$$

Combining (4.19) and (4.20),

$$\epsilon(\alpha) = \sum_{i=1}^{\min\{N,R\}}(N+R+1-2i)\alpha_i. \qquad (4.21)$$

Now we optimize this over α subject to (4.15) and (4.16), so we assume $0 < \alpha_i < 1$ for $i = 1, \ldots, \min\{N,R\}$. Initially suppose $r = 0$ which forces $\alpha_i = 1$ for all i. This corresponds to a maximum diversity of

$$\begin{aligned}
\sum_{i=1}^{\min\{N,R\}}(N+R+1-2i) &= (N+R+1)\min\{N,R\} - (1+\min\{N,R\})\min\{N,R\} \\
&= \min\{N,R\}(N+R-\min\{N,R\}) \\
&= \min\{N,R\}\max\{N,R\} \\
&= NR.
\end{aligned}$$

As r increases, we are free to lower the values of α_i. Since the coefficients $(N+R-1-2i)$ of α_i are positive and strictly decreasing in i, to minimize (4.21) it is optimal to push α_i to zero one at a time beginning with α_1.

That is,

- For $r = 0$ set all $\alpha_i := 1$ achieving diversity NR.

- For $0 < r < 1$ push α_1 to zero, $\alpha_i = 1$ for $i > 1$. The diversity decreases by $N+R-1$ to $NR - (N+R-1) = (N-1)(R-1)$ for $r = 1$.

- For $1 < r < 2$ push α_2 to zero, while $\alpha_1 = 0$, $\alpha_i = 1$ for $i > 2$. The diversity decreases by $N+R-3$ to $(N-1)(R-1) - (N+R-3) = (N-2)(R-2)$ for $r = 2$.

- ...

- For $\min\{N,R\} - 1 < r < \min\{N,R\}$ push $\alpha_{\min\{N,R\}}$ to zero, while $\alpha_i = 0$ for $i < \min\{N,R\}$. The diversity decreases by $|N-R|+1$ to zero for $r = \min\{N,R\}$.

We hence obtain a piecewise linear curve joining points $(k, (N-k)(R-k))$ for $k = 1, 2, \ldots, \min\{N,R\}$. This is precisely the same as the optimal trade-off curve for a MIMO system with R transmit and N receive antennas. Next we will verify that this trade-off curve is achievable with a particular code construction.

4.3.3 Achieving the Trade-off

By adopting an analysis similar to that presented by Elia et al. in [14] one can construct an explicit coding scheme which satisfies a *non-vanishing determinant criterion*, and in doing so achieves the optimal trade-off. This property states that if S_1 and S_2 are two distinct codewords then

$$\det(S_1 - S_2)(S_1 - S_2)^* \doteq P^0.$$

That is, the determinant does not decay to zero as $P \to \infty$. Codes based on cyclic division algebras [42, 14, 48] can be made to have this property.

4.3.4 The Case $M < N$

The optimization of Section 4.3.2 would not have worked for the case $M < N$. In this instance we reason as follows.

Allowing for full cooperation amongst the relay nodes, the diversity of the system can be upper bounded by the diversity obtained by the each of the first stage ($d_{\mathrm{MIMO}(M,R)}(r)$) and second stage ($d_{\mathrm{MIMO}(R,N)}(r)$) of transmission. That is,

$$\begin{aligned} d(r) &\leq \min\{d_{\mathrm{MIMO}(M,R)}(r), d_{\mathrm{MIMO}(R,N)}(r)\} \\ &= \min\{d_{\mathrm{MIMO}(R,M)}(r), d_{\mathrm{MIMO}(R,N)}(r)\} \\ &= d_{\mathrm{MIMO}(R,\min\{M,N\})}(r). \end{aligned} \tag{4.22}$$

We have already seen this bound to be achievable in the $M \geq N$ case. When $M < N$, simply ignore the signal received at the final $N - M$ receive antennas. This allows us to apply the result we have proved and gives a lower bound on diversity:

$$\begin{aligned} d(r) &\geq d_{\mathrm{MIMO}(M,R)}(r) \\ &= d_{\mathrm{MIMO}(R,\min\{M,N\})}(r). \end{aligned} \tag{4.23}$$

Comparing (4.22) and (4.23), we conclude

$$d(r) = d_{\mathrm{MIMO}(R,\min\{M,N\})}(r).$$

4.4 Discussion

We make the following remarks.

- Large diversity benefits can be reaped by increasing the number of relay antennas, provided they maintain independent channels. Implementing cheap relay nodes may be easier to do than adding antennas at the transmitter or receiver.

- The trade-off behavior is a function of $\min\{M, N\}$ so there is no point in having more transmit than receive antennas or vice versa.

- Compared with the results of Yang and Belfiore [77] for the product Rayleigh channel, we see the benefit of having different unitary matrices at the relays. The Rayleigh product channel's result apply to our model in the case A_i all equal to the identity matrix, and there the diversity is at most MN if $R > \max\{M, N\}$.

- The optimization problem we solved to find the dominant exponent was different from the Zheng-Tse work, but led to a similar answer. This arose from the eigenvalues of D being of order P^0 ($\beta = 0$), so that outage was again related to the eigenvalues of a Wishart matrix. It would be interesting to investigate precisely how this arises, and why this is not seen for the Rayleigh product channel.

4.5 Summary

We have established the diversity-multiplexing gain trade-off for a MIMO(M,N) system with R relay nodes. This was found to coincide with that of a MIMO system with R transmit and $\min\{M, N\}$ receive antennas. The calculation involves working with an equivalent MIMO channel model, then performing an outage probability calculation on the equivalent channel. Whether this calculation can be streamlined and applied to other networks remains a challenging problem.

It would also be interesting to consider a more general network setup where there is more than one transmitter/receiver pair or if the relay nodes themselves had more than one antenna each. We will look at the former case in more detail in the next chapter, where every node has a single antenna. Interference is now introduced into the system and we will see how high diversity can still be achieved.

Chapter 5

Diversity in Wireless Networks

5.1 Introduction

Up until now we have looked at wireless systems with one transmitter and one receiver. For the remainder of the thesis we study networks with more than one transmitter-receiver pair. This introduces *interference* into the system as users compete for access to the shared wireless medium. The receivers then have to decode signals perturbed not only by fading and noise but also by additive interference from other users. The essential question we ask here is, how can we ensure reliable communications without jeopardizing on the total rate of the system? For example, we do not wish to have the transmitter-receiver pairs communicate one at a time. While this avoids the interference problem and is reliable, it gives a low rate per user. Are there cleverer schemes? Here we will investigate ways that other nodes cooperate with communication between source and destination, ensuring a higher rate while maintaining reliable communication. Again we study this in the high SNR regime, where we can measure reliability by diversity, the exponent of the error probability due primarily to fading. From now we also suppose all communication nodes have a single antenna.

We will study two scenarios for the wireless network, both subject to Rayleigh fading and additive Gaussian noise. We are interested in some of the possible relationships between diversity and rate (measured again by the multiplexing gain).

In the first case we have m source-destination pairs (a total of $2m$ nodes). Initially we focus on $m = 2$ and the non-cooperative case (an interference channel). For such a channel we investigate the diversity-multiplexing gain trade-off for four well-known transmission/reception schemes. It will be seen that both diversity and rate are maximized when each receiver decodes both transmitted messages jointly. These results can be generalized to any value of m. With a view to increasing diversity, two more schemes are then considered for the interference channel, now allowing cooperation between nodes. The nodes provide increased diversity but at the expense of rate. One scheme is shown to increase diversity up to a factor of three while reducing rate by four, while the other has diversity two but cuts rate by a factor of three.

In the second scenario we will consider n relay nodes in addition to the m source-destination pairs (a total of $2m + n$ nodes). This time the relay nodes will be used to eliminate interference at the receivers and provide a diversity linear in the number of relay nodes. The diversity analysis of this network is more involved. Compared to an alternative protocol where receivers decode all transmitted messages, this scheme is seen to achieve higher diversity at higher rates.

The capacity of interference channels, with or without relays, is an open problem in information theory. Here we are considering particular transmission strategies and aim to achieve high diversity over a range of transmission rates.

5.1.1 Cooperative Diversity with More than One Transmitter or Receiver

As the demand for higher reliable data rates in wireless communications continues, so too does the importance of understanding the interplay between rate and diversity in the general network setting. In the previous two chapters we reviewed some of the recent work concerning the optimal diversity-multiplexing gain trade-off for point-to-point systems with and without relays. It was seen that at least for small multiplexing gains high diversity gains can be realized, as though the relays collectively behaved as a multiple antenna system. The optimal trade-off for large multiplexing gains remains an open problem.

In the case of more than one transmitter or receiver, there exist recent results for multiple access and broadcast channels. Zheng and Tse [83] generalize their results for the point-to-point MIMO system [82] to the case where K users each have m transmit antennas and communicate to a single receiver with n receive antennas. The set of multiplexing gains $\{r_1, \ldots, r_K\}$ that allow each user to achieve a diversity gain d is quantified.

Azarian et al. [3] study the half-duplex cooperative broadcast and multiple access channels as generalizations of their work on relay channels. Here they assume each user has a single antenna. They show that the trade-off achieved by a *dynamic decode and forward* protocol for a broadcast channel with N destinations, is the same as that for a corresponding relay channel with N relays. Hence the added requirement of receiving nodes having to decode their intended message does not lead to any cost in the trade-off curve. For multiple access channels they show that a non-orthogonal amplify and forward protocol is trade-off optimal: for N transmitters a diversity gain of $N(1-r)$ can be achieved for a multiplexing gain of r.

Using nodes for cooperation in interference channels is investigated by Høst-Madsen in [27], inheriting methods known for the interference and relay channels (see for example [32] and [76]). Here achievable rates for several schemes are analyzed, which may be categorized under:

- compress and forward: a relay node transmits a compressed version of what it has received using Wyner-Ziv coding. Amplify-forward is the special case of no compression.

- decode and forward: the relay decodes the message sent by the relevant transmitter and reencodes it for transmission.

The receivers can employ individual or joint decoding. In the former, one decodes only the intended signal treating any other users as part of the noise. In the latter, one attempts to decode all signals whether or not they are intended for that receiver.

Ideally one would like to analyze the outage behavior of these schemes at high SNR, but the rate equations have difficult descriptions for such analysis. Here we consider simpler protocols for the interference channel that allow for outage analysis.

It should be emphasized that a wireless network with cooperation amongst nodes can never perform as well as a corresponding MIMO system. For one thing, the achievable multiplexing gain of each user is always limited by the number of antennas at the source or destination, which in our case is 1. Yuksel and Erkip note this in [80] and show that by considering cut set bounds, a source destination pair with two additional nodes as relays can perform at best as a 1×3 MIMO system with maximum diversity 3. One of the schemes we shall consider achieves precisely this diversity.

Throughout this work we assume all nodes behave synchronously, transmitting and receiving in discrete time slots. The trade-off problem has also been studied in the asynchronous case [74]. For a list of many spatial diversity techniques in wireless networks, refer to [7] and the references therein.

As in the previous chapters, we follow the work of Zheng and Tse [82] in the definition of rate and diversity: we consider a family of codes of fixed block length and increasing signal-to-noise ratio (SNR) and say that user i supports a multiplexing gain of r_i if its data rate $R_i(\mathrm{SNR})$ satisfies $\lim_{\mathrm{SNR}\to\infty} \frac{R_i(\mathrm{SNR})}{\log \mathrm{SNR}} = r_i$. The family has a *diversity* d if the average error probability P_e behaves as $\lim_{\mathrm{SNR}\to\infty} \frac{\ln P_e(\mathrm{SNR})}{\ln \mathrm{SNR}} = -d$, that is $\mathrm{P}_e \sim \mathrm{SNR}^{-d}$. Hence the higher value of d for a given multiplexing gain, the more reliable the system at that corresponding rate. This quantity is meaningful only at high SNR.

5.1.2 The Gaussian Interference Channel

Figure 5.1 illustrates the two-user interference channel. There are two transmitters (single antenna) each with one intended receiver $(1 \to 3, 2 \to 4)$. However the received signals (complex valued) are each a linear combination of both transmitted signals (plus Gaussian noise) and the receivers have the task of determining their intended signal. Again we assume that each of the propagation parameters h_{ij} is a complex Gaussian random variable with zero mean and unit variance (Rayleigh fading).

Using subscripts which correspond to the node(s) labeled in Figure 5.1, the system equations are described by

$$\begin{aligned} y_3 &= h_{13}x_1 + h_{23}x_2 + n_3, \qquad (5.1) \\ y_4 &= h_{14}x_1 + h_{24}x_2 + n_4. \end{aligned}$$

We assume the noise terms have zero mean and unit variance and that x_1, x_2 have the power constraint $\mathbf{E}|x_i|^2 \leq P$, $i = 1, 2$. Throughout this work we assume all fading parameters h_{ij} are known to all nodes.

The interference channel has been studied for over four decades and except for the case of sufficiently strong interference, finding the capacity region for general SNR has remained

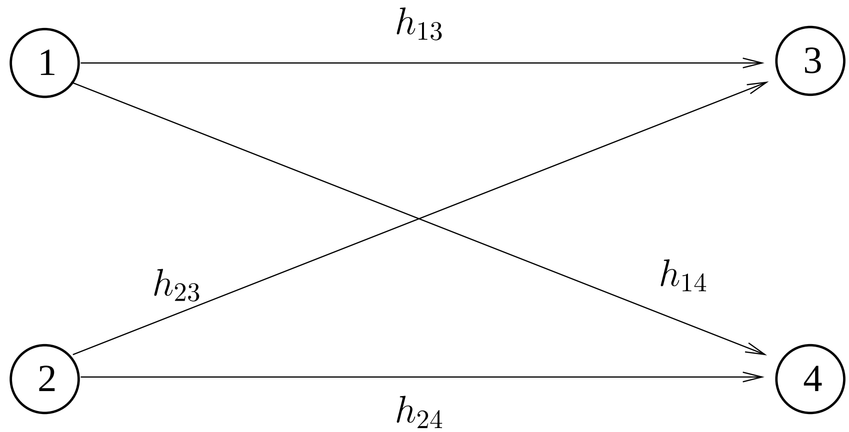

Figure 5.1: Interference channel.

an elusive open problem. One can talk about four regimes for which the following best-known transmission and decoding strategies are used, as mentioned in the work by Costa [9]. Recent work by Etkin et al. [15] has established the capacity region to within one bit per channel use.

1. Zero or low interference: both users transmit simultaneously, receivers decode their intended signal treating the other user as noise.

2. Moderate interference: Etkin et al. [15] show that a *Han-Kobayashi* scheme can achieve within one bit of capacity (thus the asymptotic capacity at high SNR). The idea is to divide the information of the two users into a private part (only decoded by its intended receiver) and common part. The resulting rate equations are not so easy to analyze, so instead we employ the simpler scheme of time or frequency division. In this case the receivers know when or over which frequency band each transmitter is transmitting, and so each knows when to decode its intended signal. The channels are othogonally separated.

3. Strong interference: both users transmit simultaneously, each receiver treats the system as a two-user multiple access channel and decodes each user jointly [24].

4. Very strong interference: here the receivers treat their intended signal as noise and first decode the other user's signal before subtracting off this interference; the capacity region is the same as if there were no interference present at all [8].

We are interested in seeing what rate and diversity each of these strategies gives us while assuming fading in the channel. In addition we will consider two more schemes in which the existing nodes are used additionally as cooperative relays.

We will seek to exploit multiple paths in the system, thus increasing diversity. No claims are made about the optimal rate-diversity trade-off for the interference channel, but as will be seen this work enables us to understand better the preference of joint decoding over successive decoding in increasing diversity.

Motivated by [33], we calculate diversity from an outage calculation for a given multiplexing gain. For fixed propagation parameters each scheme defines a region of achievable rates. When the randomness of the channels is incorporated this region changes in size—an outage corresponds to the event that, given a fixed rate pair (R_1, R_2), the achievability region does not include this pair.

5.2 Interference Channel with and without Cooperation

In this section we derive the results given in table 5.1. This shows the sum-rate and corresponding diversity one can achieve by the four transmission/reception strategies already mentioned when considering the nodes as an interference channel. The low and very strong interference cases both involve treating one of the users as noise. In Section 5.2.2 we consider two schemes involving cooperation amongst nodes, corresponding to the last two lines of the table. Here r_i ($i = 1, 2$ where $0 < r_i < 1$) are the multiplexing gains of user i and the sum-rate defined as the sum of multiplexing gains employed by each user per channel use.

Table 5.1: Summary of rate-diversity results ($K = 1 - \max\{r_1, r_2\}$)

Scheme	Sum-Rate	Diversity
Alternate transmission	$(r_1 + r_2)/2$	$1 - \max\{r_1, r_2\}$
MAC	$r_1 + r_2$	$\min\{1 - r_1, 1 - r_2, 2(1 - r_1 - r_2)\}$
Treat one of users as noise	$r_1 + r_2$	0
Alternate, using nodes as relays	$(r_1 + r_2)/4$	$3K - 4\frac{\ln(\ln P)}{\ln P}$
Interference channel, then receiver cooperation	$(r_1 + r_2)/3$	$2K - 2\frac{\ln(\ln P)}{\ln P}$

5.2.1 Using Nodes as an Interference Channel

First let us recap the diversity-multiplexing gain relationship in the point-to-point channel, as explained in (3.5). For a point-to-point complex Rayleigh fading channel $y = hx + v$ with power constraint $\mathbf{E}|x|^2 \leq P$ and zero-mean unit-variance noise v we find the diversity as a function of $R = r \log P$ $(0 < r < 1)$ via the approximation $\Pr(\log(1 + P|h|^2) < r \log P) \approx \frac{1}{P^{1-r}}$, valid for high P. This tells us that for multiplexing gain r the diversity is $1 - r$.

5.2.1.1 Alternating Transmission

If the two users alternate their transmission, where user i transmits with multiplexing gain r_i for $i = 1, 2$, then we have two point-to-point channels. The sum-rate is $(r_1 + r_2)/2$ (as there are two separate channel uses) and each has diversity $1 - r_i$ as seen above. This leads to an overall diversity of the system of $1 - \max\{r_1, r_2\}$.

5.2.1.2 MAC: Simultaneous Transmission, Decoding Both Users

Suppose both users transmit simultaneously and each of the two receivers decodes the messages of both senders. The sum-rate is then given by $(r_1 + r_2)$. The achievable rate region is given by the intersection of the capacity regions of two multiple access channels, one corresponding to each receiver. We make use of the following well known result, which can be considered as a special case of [83].

Theorem 4. *Consider the fading multiple access channel (MAC) $y = h_1x_1 + h_2x_2 + v$ where h_1 are h_2 are independent zero-mean unit-variance complex Gaussian (Rayleigh fading) parameters and v is zero-mean unit-variance complex Gaussian noise. Additionally suppose the two transmitters have the power constraint $\mathbf{E}|x_1|^2 \leq P$ and $\mathbf{E}|x_2|^2 \leq P$. Then for transmission rates r_1 and r_2 the diversity is* $\min\{1 - r_1, 1 - r_2, 2(1 - r_1 - r_2)\}$.

Proof. The capacity region of the MAC defined this way can readily be found:

$$\begin{aligned} R_1 &\leq \log(1 + P|h_1|^2), \\ R_2 &\leq \log(1 + P|h_2|^2), \\ R_1 + R_2 &\leq \log(1 + P(|h_1|^2 + |h_2|^2)). \end{aligned}$$

Then we wish to find the probability that a point (R_1, R_2) lies outside this region due to the randomness of h_1 and h_2. This probability is at most the sum of the probabilities of one of the three inequalities above being violated (union bound). The sum of two independent exponential random variables of unit mean can be shown to have distribution $\Pr(|h_1|^2 + |h_2|^2 < x) = 1 - e^{-x}(x+1) \leq x^2$, so the third inequality is violated with probability at most $((2^{R_1+R_2} - 1)/P)^2$. As in the point-to-point case each of the first two inequalities is violated with probability approximately $(2^{R_i} - 1)/P = P^{-(1-r_i)}$. Hence the outage probability is

$$\begin{aligned} \mathrm{P}_{\text{out}} &\leq P^{-(1-r_1)} + P^{-(1-r_2)} + P^{-2(1-r_1-r_2)} \\ &\leq 3P^{-\min\{1-r_1, 1-r_2, 2(1-r_1-r_2)\}}. \end{aligned}$$

From this the diversity is $\min\{1 - r_1, 1 - r_2, 2(1 - r_1 - r_2)\}$. □

The interference channel achievability region is considered as the intersection of two MACs, each with diversity $\min\{1-r_1, 1-r_2, 2(1-r_1-r_2)\}$. Hence the overall interference channel using this scheme has diversity $\min\{1-r_1, 1-r_2, 2(1-r_1-r_2)\}$, with the constraint $r_1 + r_2 < 1$.

Comparing this scheme to the previous alternating scheme, one can show that for equal sum-rate, the diversity for this second scheme is at least as high as that of the first scheme. That is, for $r_1 + r_2 < 1$,

$$\min\{1 - r_1, 1 - r_2, 2(1 - r_1 - r_2)\} \geq 1 - \max\{1 - 2r_1, 1 - 2r_2\}.$$

5.2.1.3 Treating One of the Users as Noise

Finally we consider schemes in which one of the users is decoded before the other. This is optimal when there is zero or very strong interference and there is no alternating in transmission this time. The sum-rates in both cases are $r_1 + r_2$.

Referring to (5.1), as x_1 and x_2 are independent, and assuming we use Gaussian codebooks, we have

$$I(x_1; y_3) \quad = \quad \log\left(1 + \frac{P|h_{13}|^2}{1 + P|h_{23}|^2}\right).$$

The presence of P in both the numerator and denominator of this expression means that for high P this mutual information does not become large. The probability of outage becomes

$$\mathrm{P_{out}} = \Pr\left(\frac{P|h_{13}|^2}{1 + P|h_{23}|^2} < 2^R - 1\right),$$

which does not decay at all with P. Hence the diversity is zero.

This completes our derivation of diversity expressions for four commonly chosen transmission/reception strategies for the interference channel. We see that decoding both users is superior in both rate and diversity over all other schemes. From a diversity view point we have also seen that successive cancellation is ineffective.

5.2.2 Using Nodes as Relays

Here we look at cooperative diversity schemes in which nodes are now able to forward information intended for others. Some of the ideas here are motivated by [27].

5.2.2.1 Alternating Transmission Using Nodes as Relays

In figure 5.2 we see that there are potentially three independent paths via which node 1 can send its information to node 3, (labeled I, II and III). Similarly node 2 can send its information to node 4 via three paths. Firstly consider the transmission of the information of node 1. In order to maximize diversity we assume nodes 2 and 4 are amplifying and forwarding their received signal. Node 3 as a result also receives amplified noise generated at nodes 2 and 4.

Diversity in amplify-forward networks is studied in [35] and [29] amongst others but they do not look at the specific case of two relays plus a direct link between transmitter and receiver.

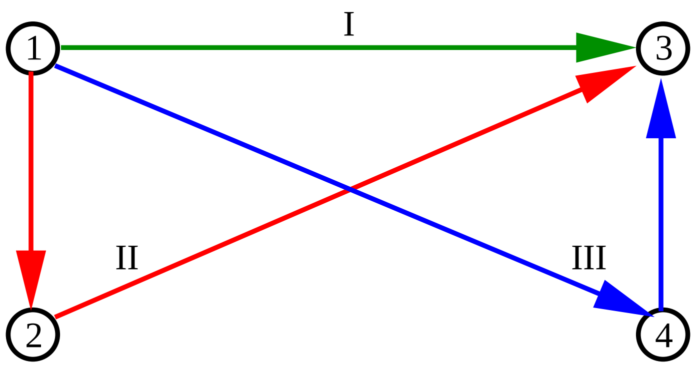

Figure 5.2: Possible independent paths between nodes 1 and 3.

Suppose that in the first time slot node 1 transmits to nodes 2, 3 and 4 while in the second time slot nodes 2 and 4 amplify and forward their received signal to node 3 while node 1 retransmits its message to node 4. Note that due to the half-duplex condition we have imposed node 2 is not yet transmitting its information to node 4. In the third and fourth time slots node 2 then transmits to node 4 via nodes 1 and 2 similarly as depicted in figure 5.3. The sum-rate of this scheme is $(r_1 + r_2)/4 \leq 1/2$ as a total of four channel uses is employed.

For node 3 the system equation is given by

$$\begin{bmatrix} y_3^{(1)} \\ y_3^{(2)} \end{bmatrix} = \begin{bmatrix} h_{13} \\ (h_{13} + \alpha_1 h_{23} h_{12} + \alpha_2 h_{43} h_{14}) \end{bmatrix} x_1 + \begin{bmatrix} n_3^{(1)} \\ (h_{23} n_2 + h_{43} n_4 + n_3^{(2)}) \end{bmatrix}, \quad (5.2)$$

where we have separated the signal and noise component, and superscripts are used to

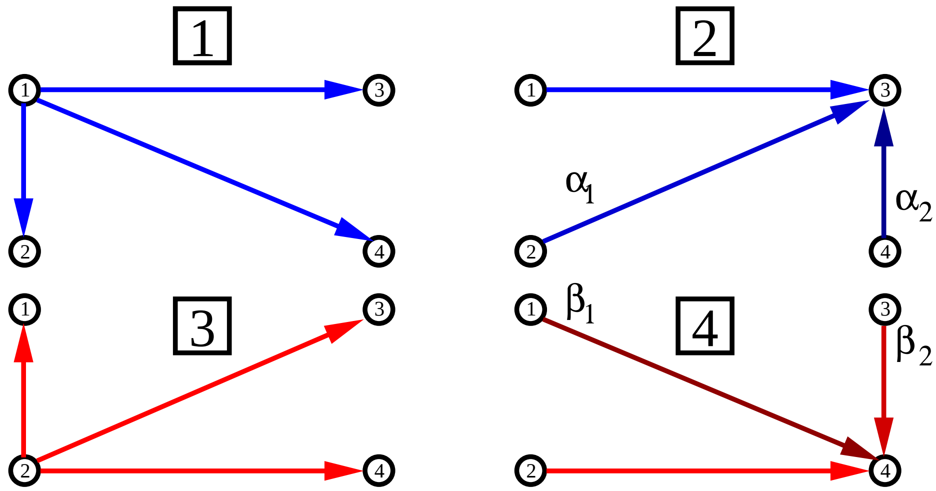

Figure 5.3: First cooperative diversity scheme.

denote time steps. The relays do not have any peak power constraint but note that since the fading parameters have unit variance, the average power of the relays is also of the order of P.

The parameters α_1 and α_2 are phases (complex numbers of unit magnitude) chosen so that the three signals received by node 2 add coherently (with the same phase). This coherent addition is highly necessary. For example if h_1 and h_2 are independent $\mathbb{C}\mathcal{N}(0,1)$ random variables then without coherent addition,

$$\begin{aligned}
\Pr(|h_1+h_2|^2 < \epsilon) &= \Pr(|h|^2 < \epsilon) \quad \text{where } h \sim \mathbb{C}\mathcal{N}(0,2) \\
&= 1 - e^{-\epsilon/2} \\
&\approx \epsilon/2,
\end{aligned}$$

while with coherent addition,

$$\begin{aligned}
\Pr(|h_1|^2 + |h_2|^2 < \epsilon) &= 1 - e^{-\epsilon}(\epsilon + 1) \\
&\approx 1 - (1 - \epsilon)(\epsilon + 1) \\
&= \epsilon^2,
\end{aligned} \tag{5.3}$$

leading to diversity 1 and 2 respectively.

One may write a similar equation to (5.2) for node 4. We now show that the diversity of the network employing this scheme is close to 3. The main idea behind the analysis that follows is that we wish to choose bounds that do not impact diversity by being aware of the most significant fading terms that lead to outage.

From the system equation, the achievable rate for node 3 is found to be

$$\begin{aligned}
R_1 &\leq \log\left(1 + P\left(|h_{13}|^2 + \frac{|h_{13} + \alpha_1 h_{23} h_{12} + \alpha_2 h_{43} h_{14}|^2}{1 + |h_{23}|^2 + |h_{43}|^2}\right)\right) \\
&= \log(1 + PL_1),
\end{aligned}$$

where

$$L_1 := |h_{13}|^2 + \frac{(|h_{13}| + |h_{23} h_{12}| + |h_{43} h_{14}|)^2}{1 + |h_{23}|^2 + |h_{43}|^2},$$

which is possible for suitable choices of α_1 and α_2. Similarly $R_2 \leq \log(1 + PL_2)$ where

$$L_2 := |h_{24}|^2 + \frac{(|h_{24}| + |h_{14} h_{21}| + |h_{34} h_{23}|)^2}{1 + |h_{14}|^2 + |h_{34}|^2}.$$

Then

$$
\begin{aligned}
\mathrm{P_{out}} &= \Pr\left(L_1 < P^{-(1-r_1)} \text{ or } \Pr(L_2 < P^{-(1-r_2)}\right) \\
&\leq \Pr(L_1 < P^{-(1-r_1)}) + \Pr(L_2 < P^{-(1-r_2)}), \qquad (5.4)
\end{aligned}
$$

by the union bound.

We upper bound the first of these terms (the second term is treated similarly) as follows:

$$
\begin{aligned}
& \Pr(L_1 < P^{-(1-r_1)}) \\
= \; & \Pr\left(|h_{13}|^2 + \frac{(|h_{13}| + |h_{23}h_{12}| + |h_{43}h_{14}|)^2}{1 + |h_{23}|^2 + |h_{43}|^2} < \epsilon_1\right) \\
& \text{(setting } \epsilon_1 := P^{-(1-r_1)} \text{ which is small)} \\
\leq \; & \Pr\left(|h_{13}|^2 + \frac{|h_{23}h_{12}|^2 + |h_{43}h_{14}|^2}{1 + |h_{23}|^2 + |h_{43}|^2} < \epsilon_1\right) \\
& \text{(decreasing the left side, thus loosening the bound)} \\
\leq \; & \Pr\left(|h_{13}|^2 < \epsilon_1, \frac{|h_{23}h_{12}|^2 + |h_{43}h_{14}|^2}{1 + |h_{23}|^2 + |h_{43}|^2} < \epsilon_1\right) \\
& \text{(if the sum of positive terms is upper bounded,} \\
& \text{each term individually has the same bound)} \\
= \; & \Pr(|h_{13}|^2 < \epsilon_1) \Pr\left(\frac{|h_{23}h_{12}|^2 + |h_{43}h_{14}|^2}{1 + |h_{23}|^2 + |h_{43}|^2} < \epsilon_1\right). \qquad (5.5)
\end{aligned}
$$

As $|h_{13}|^2$ has exponential distribution with unit mean, the first term in this product is $1 - e^{-\epsilon_1} \leq \epsilon_1$. Concentrating on the second term, for convenience we set $M_1 := |h_{23}h_{12}|^2$, $M_2 := |h_{43}h_{14}|^2$, $S := |h_{23}|^2 + |h_{43}|^2$. As seen in (5.3), S as the sum of two exponential random variables can be shown to have distribution

$$\Pr(S > x) = e^{-x}(x + 1). \qquad (5.6)$$

Then we have

$$
\begin{aligned}
& \Pr\left(\frac{|h_{23}h_{12}|^2+|h_{43}h_{14}|^2}{1+|h_{23}|^2+|h_{43}|^2}<\epsilon_1\right) \\
= \;& \Pr\left(M_1+M_2<\epsilon_1(1+S)\right) \\
= \;& \Pr\big(M_1+M_2<\epsilon_1(1+S), S<\ln(1/\epsilon')\big) \\
& \quad +\Pr\big(M_1+M_2<\epsilon_1(1+S)|S>\ln(1/\epsilon')\big).\Pr\big(S>\ln(1/\epsilon')\big) \\
& (\epsilon' \text{ is some positive function of } \epsilon_1 \text{ to be chosen later}) \\
\leq \;& \Pr\big(M_1+M_2<\epsilon_1(1+\ln(1/\epsilon')), S<\ln(1/\epsilon')\big)+1\times\Pr\big(S>\ln(1/\epsilon')\big) \\
& (\text{maximizing the upper bound in the first condition} \\
& \text{and using a bound of 1 in the second term}) \\
\leq \;& \Pr\big(M_1+M_2<\epsilon_1(1+\ln(1/\epsilon'))\big)+\epsilon'(1+\ln(1/\epsilon')) \\
& (\text{dropping the second condition in the first term and using (5.6) in the second term}) \\
\leq \;& \Pr\big(M_1<\epsilon_1(1+\ln(1/\epsilon')), M_2<\epsilon_1(1+\ln(1/\epsilon'))\big)+\epsilon'(1+\ln(1/\epsilon')) \\
= \;& \big[\Pr\big(M_1<\epsilon_1(1+\ln(1/\epsilon'))\big)\big]^2+\epsilon'(1+\ln(1/\epsilon')) \qquad (5.7) \\
& (M_1 \text{ and } M_2 \text{ are independent}).
\end{aligned}
$$

To proceed we use the following lemma concerning the distribution of the product of exponential random variables near the origin.

Lemma 6. *Let λ_1 and λ_2 be independent exponential random variables each with unit mean. Then*

$$\Pr\left(\lambda_1\lambda_2<\epsilon\right)=\epsilon\ln(1/\epsilon)+o(\epsilon\ln(1/\epsilon)). \qquad (5.8)$$

In fact numerically one may verify that $\Pr(\lambda_1\lambda_2<\epsilon)\leq\epsilon\ln(1/\epsilon)$ for $\epsilon<0.07$.

Proof.

$$\begin{aligned}
\Pr(\lambda_1\lambda_2 < \epsilon) &= \int_0^\infty \Pr(\lambda_1 < \epsilon/x)\, e^{-x}\; dx \\
&= \int_0^\infty e^{-x}(1 - e^{-\epsilon/x})\; dx \\
&= 1 - \int_0^\infty e^{-(x+\epsilon/x)}\; dx \\
&= 1 - 2\sqrt{\epsilon}K_1(2\sqrt{\epsilon}), \qquad (5.9)
\end{aligned}$$

where $K_1(x) := \int_0^\infty e^{-x\cosh t}\cosh t\; dt$ is a modified Bessel function of the second kind. Referring to 9.6.11 and 9.6.7 of Abramowitz and Stegun [2],

$$K_1(z) = z^{-1} + (z/2)\ln(z/2) + o(z\ln z). \qquad (5.10)$$

Substituting (5.10) into (5.9) gives us (5.8). One can also compute the next significant term in the expansion and verify that its coefficient ($1 - 2\gamma$, where γ is the Euler constant) is negative. Hence for sufficiently small ϵ, $\Pr(\lambda_1\lambda_2 < \epsilon) \le \epsilon\ln(1/\epsilon)$. □

Applying this lemma to (5.7) gives

$$\begin{aligned}
&\Pr(M_1 + M_2 < \epsilon_1(1+S)) \\
\le\; & \left[\Pr\left(M_1 < \epsilon_1(1+\ln(1/\epsilon'))\right)\right]^2 + \epsilon'(1+\ln(1/\epsilon')) \\
\le\; & \left[\epsilon_1\left(1+\ln(1/\epsilon')\right)\ln\left[\frac{1}{\epsilon_1(1+\ln(1/\epsilon'))}\right]\right]^2 \\
&+ \epsilon'(1+\ln(1/\epsilon')).
\end{aligned}$$

This holds for all choices of $\epsilon' > 0$. To ensure a reasonably tight upper bound choose

$\epsilon' := {\epsilon_1}^2$ so that

$$\begin{aligned}
& \Pr\left(M_1 + M_2 < \epsilon_1(1+S)\right) \\
\leq\ & \left[\epsilon_1\left(1 + \ln(1/{\epsilon_1}^2)\right)\ln\left[\frac{1}{\epsilon_1(1+\ln(1/{\epsilon_1}^2))}\right]\right]^2 + {\epsilon_1}^2(1+\ln(1/{\epsilon_1}^2)) \\
\leq\ & \left[\epsilon_1(1+2\ln(1/\epsilon_1))\ln\left[\frac{1}{\epsilon_1}\right]\right]^2 + {\epsilon_1}^2(1+2\ln(1/\epsilon_1)) \\
\leq\ & C{\epsilon_1}^2\left(\ln(1/\epsilon_1)\right)^4,
\end{aligned}$$

for some positive constant C. From (5.5) we conclude that $\Pr(r_1 \log P > \log(1+PL_1)) \leq C{\epsilon_1}^3\left(\ln(1/\epsilon_1)\right)^4$. By the same argument we can show $\Pr(r_2 \log P > \log(1+PL_2)) \leq C{\epsilon_2}^3\left(\ln(1/\epsilon_2)\right)^4$ where $\epsilon_2 := P^{-(1-r_2)}$. Hence from (5.4) we have

$$\begin{aligned}
\mathrm{P}_{\mathrm{out}} &\leq C{\epsilon_1}^3\left(\ln(1/\epsilon_1)\right)^4 + C{\epsilon_2}^3\left(\ln(1/\epsilon_2)\right)^4 \\
&\leq 2Cf(\max\{\epsilon_1,\epsilon_2\}),
\end{aligned}$$

where $f(x) = x^3(\ln(1/x))^4$.

Since $\max\{\epsilon_1,\epsilon_2\} = P^{-(1-\max\{r_1,r_2\})}$ the diversity is lower bounded by

$$-\lim_{P\to\infty}\log \mathrm{P}_{\mathrm{out}}/\log P \geq 3K - \lim_{P\to\infty} 4\frac{\log(K\log P)}{\log P} = 3K,$$

where $K = 1 - \max\{r_1, r_2\}$.

We remark that the exponent $4\frac{\log(K\log P)}{\log P}$ represents the slight penalty in diversity over a multiantenna system with three independent Rayleigh fading paths. Two main factors—the composition of Rayleigh fading paths, and presence of fading-dependent noise—have contributed to this degradation which may become significant at moderate or lower power levels.

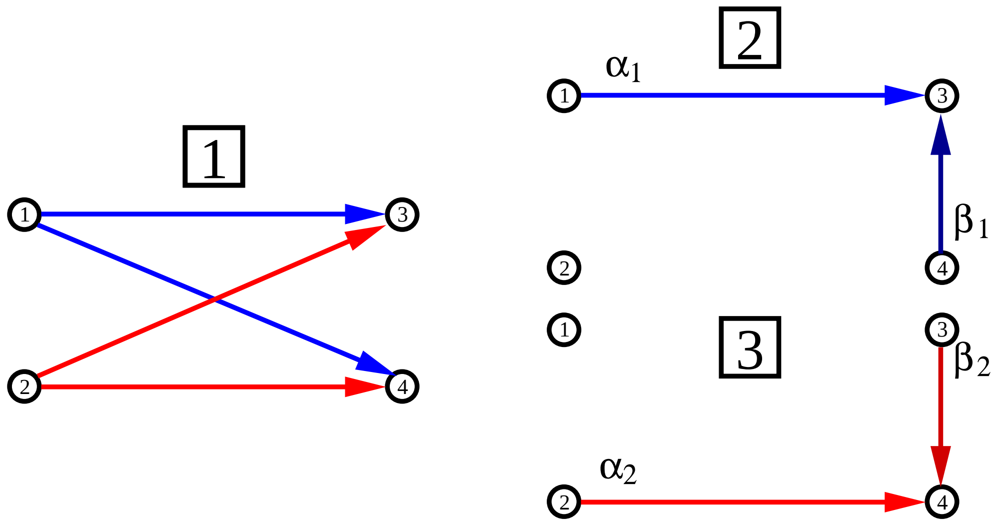

Figure 5.4: Second cooperative diversity scheme.

As mentioned in the previous chapter, the noise from intermediate nodes may be amplified by a fade, leading to large denominators in the rate equations. However since the tail of a Gaussian distribution decays exponentially, the probability that outage occurs due to the fade being sufficiently large is correspondingly small. This motivates the $\ln(1/\epsilon')$ bounds in (5.7). In other words, at high SNR, relay noise does not alter diversity.

5.2.2.2 Interference Channel with Receiver Cooperation

Now we consider a scheme in which three instead of four channel uses are employed (so the sum-rate is $(r_1+r_2)/3$) but due to the half-duplex constraint, only two of the three possible independent transmission paths are utilized. Hence we cannot hope to achieve a diversity greater than two. Figure 5.4 shows the steps involved in transmission. Furthermore the nodes apply phases α_i, β_i as shown in the figure.

The system equations are then given by

$$\begin{bmatrix} y_3^{(1)} \\ y_3^{(2)} \end{bmatrix} = \begin{bmatrix} h_{13} & h_{23} \\ \alpha_1 h_{13} + \beta_1 h_{43} h_{14} & \beta_1 h_{24} h_{43} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_3^{(1)} \\ h_{43} n_4^{(1)} + n_3^{(2)} \end{bmatrix},$$
$$\begin{bmatrix} y_4^{(1)} \\ y_4^{(3)} \end{bmatrix} = \begin{bmatrix} h_{14} & h_{24} \\ \alpha_1 h_{13} h_{34} & \beta_2 h_{24} + \alpha_2 h_{23} h_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_3^{(1)} \\ h_{34} n_3^{(1)} + n_4^{(3)} \end{bmatrix}.$$

Referring to Telatar's results on mutual information in a vector Gaussian channel [66] the achievable rate region for node 3 is found to be

$$\begin{aligned} R_1 &\leq \log\left(1 + P\left(|h_{13}|^2 + \frac{|\alpha_1 h_{13} + \beta_1 h_{43} h_{14}|^2}{1 + |h_{43}|^2}\right)\right), \\ R_2 &\leq \log\left(1 + P\left(|h_{23}|^2 + \frac{|h_{24} h_{43}|^2}{1 + |h_{43}|^2}\right)\right), \end{aligned}$$

$$R_1 + R_2 \leq \log\det(I + P\Sigma^{-1} H^* H). \tag{5.11}$$

where

$$\begin{aligned} \Sigma &= \begin{bmatrix} 1 & 0 \\ 0 & 1 + |h_{43}|^2 \end{bmatrix}, \\ H &= \begin{bmatrix} h_{13} & h_{23} \\ \alpha_1 h_{13} + \beta_1 h_{43} h_{14} & \beta_1 h_{24} h_{43} \end{bmatrix}. \end{aligned}$$

Another three equations can be written for node 4, so we have a total of six rate constraints. The phases α_i, β_i ($i = 1, 2$) are chosen to prevent cancellation amongst the complex fading parameters. For example α_1, β_1 are chosen so that the terms in $\det H$ must add coherently.

An outage occurs when at least one of these constraints is violated for some fixed choice of rate pair (R_1, R_2). The outage probability can therefore be upper bounded by the union bound sum of the individual probabilities of each constraint violation. Four of the six outage calculations are treated similarly to those in the previous section except here we have a single relay instead of two. It can be shown that for any of these $\mathrm{P_{out}} \le C\epsilon^2(\ln(1/\epsilon))^2$ (where $\epsilon = P^{-(1-\max\{r_1,r_2\})}$).

For constraint (5.11) for example, assuming phases have been appropriately chosen, we may upper bound the outage probability as

$$\begin{aligned}
\mathrm{P_{out}} &= \Pr\left(\log\det(I + P\Sigma^{-1}H^*H)) < R_1 + R_2\right) \\
&\le \Pr\left(\det(P\Sigma^{-1}H^*H)) < P^{r_1+r_2}\right) \\
&\quad \text{(decreasing the left side increases its probability)} \\
&= \Pr\left(\frac{|\det H|^2}{\det\Sigma} < P^{-(2-(r_1+r_2))}\right).
\end{aligned}$$

Assuming phases have been chosen appropriately, $|\det H|^2$ may be lower bounded as $|\det H|^2 = c(|h_{13}h_{43}h_{24}| + |h_{23}h_{13}| + |h_{23}h_{43}h_{14}|)^2 \ge c(|h_{13}h_{43}h_{24}|^2 + |h_{23}h_{13}|^2 + |h_{23}h_{43}h_{14}|^2)$. Hence we wish to approximate the probability

$$\Pr\left(\frac{|h_{13}h_{43}h_{24}|^2 + |h_{23}h_{13}|^2 + |h_{23}h_{43}h_{14}|^2}{1+|h_{43}|^2} < \epsilon\right).$$

This is a sum of three dependent variables so we cannot apply the same method as the previous section. Instead we argue that this probability behaves like $o(\epsilon^2)$ by noting that if any two of the exponential variables $|h_{13}|^2$, $|h_{43}|^2$, $|h_{24}|^2$, $|h_{23}|^2$, $|h_{14}|^2$ are small, this does not guarantee that the inequality is satisfied. Hence the probability is expected to be strictly less than $O(\epsilon^2)$.

We conclude that the other four inequalities are more stringent at high P, so the outage probability of this scheme behaves like $\mathrm{P}_{\mathrm{out}} \sim P^{-2K}(K \ln P)^2$ where $K = 1 - \max\{r_1, r_2\}$. Hence the diversity is $2K - 2\ln(K \ln P)/\ln P$ and this enables us to complete table 5.1.

5.2.3 Generalization

The results of the previous sections can be readily generalized to more than two source-destination pairs, as shown in figures 5.5 and 5.6. The diversity can be shown to be the number of independent paths utilized by each pair. For m such pairs each scheme works as follows:

- Scheme 1: In Step 1 a source transmits to every other node. Then in Step 2 those nodes transmit with appropriate phases to the intended receivers. Steps 1 and 2 are repeated for every other source-destination pair, leading to a total of $2m$ time slots. The maximum diversity obtainable is $2m-1$ as there are $2m-2$ paths via other nodes from source to destination, not including the direct path.

- Scheme 2: In Step 1 all sources transmit to all nodes (an m-user interference channel). In the following m steps each receiver takes its turn in receiving signals that were previously received by the other $m-1$ receivers. Hence a total of $m+1$ time slots are used. The receivers decode their intended signals by treating the system as a vector Gaussian channel. Here the maximum diversity obtainable is m, with one contributed by each receiver.

Again this assumes that signals can be made to add coherently at the receivers. To sum up:

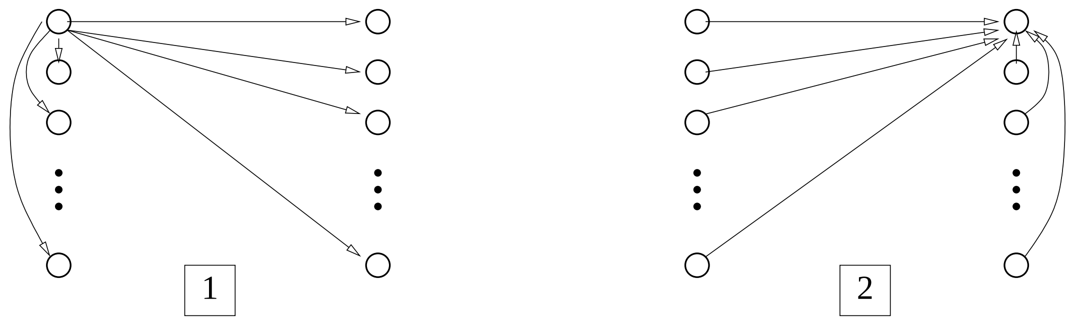

Figure 5.5: Generalization of Scheme 1.

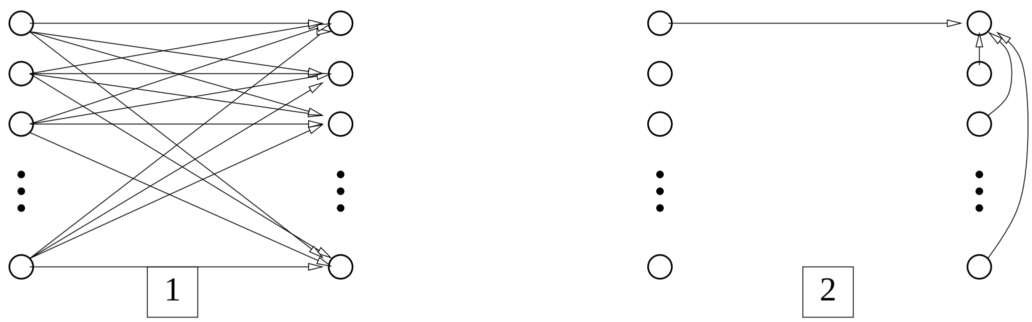

Figure 5.6: Generalization of Scheme 2.

Theorem 5. *For a Rayleigh fading network with m source-destination pairs, consider the generalizations of schemes 1 and 2 as described above and illustrated in figures 5.5 and 5.6. Then generalized Scheme 1 can provide a diversity of at most $2m-1$ while reducing rate by $2m$. Scheme 2 has diversity m but cuts rate by a factor of $m+1$.*

5.3 High Diversity Scheme Based on Interference Cancellation

Now we move on to a different setup. Consider a wireless network with m transmitter-receiver pairs and an additional n relay nodes to assist communication. All nodes have single antennas. We are interested in the diversity-multiplexing gain trade-off of such a system.

Since the presence of interference is known to reduce diversity significantly, we propose a transmission scheme based on interference cancellation by the relay nodes. This scheme achieves a diversity linear in the number of relay nodes (over all rates up to the maximum possible). Compared to a MAC-based protocol where receivers decode all transmitted messages, the new scheme is seen to achieve higher diversity at higher rates.

In this work we look at a network with m transmitters each having a receiver they wish to communicate with. Assisting us in this communication are n relay nodes which provide redundant paths from sender to receiver. For reliable communication each receiver needs to decode its intended signal with low probability of error.

We will be defining a wireless scheme in which the diversity increases linearly in the number of relay nodes. The model has also been investigated in [11] with power efficiency in mind. We then find its diversity through an outage probability calculation. This will be expressed as a function of the sum of the multiplexing gains of the users. We postpone proofs to the final section.

We also compare the diversity-multiplexing trade-off curve to the scheme we considered in the previous section, in which there was no interference cancellation and receivers decoded all transmitted signals.

5.3.1 Model and Transmission Scheme

We have m transmitter-receiver pairs and n relay nodes, shown in figure 5.7.

Let f_{ik}, g_{ij}, h_{jk} be fading coefficients between nodes, where $i = 1$ to m indexes the transmitters, $j = 1$ to n indexes the relays and $k = 1$ to m indexes the receivers respectively. All coefficients are once again assumed to be independent and identically drawn from a $\mathbb{CN}(0, 1)$ complex Gaussian distribution. We assume each relay node has knowledge of all

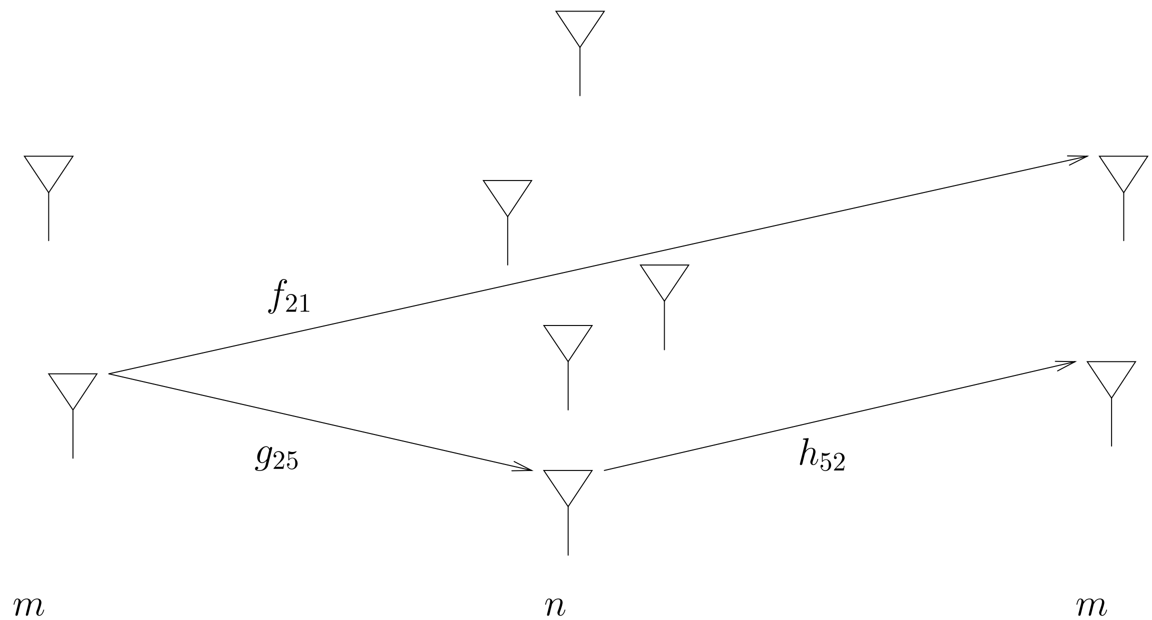

Figure 5.7: Wireless network with m transmitter-receiver pairs and n relay nodes. In Stage 1 the transmitters send to the relays and receivers, in Stage 2 the relays forward a scaled version of what they have received to the receivers. Three of the fading coefficients are shown.

fading coefficients. This is a rather strong assumption. To justify it, suppose that none of the channels are changing rapidly and some form of feedback is used. We also note that our results can be viewed as outer bounds for schemes that do not assume this channel knowledge. The receivers, however, need not have knowledge of all the channels.

Initially the m transmitter nodes send their signals simultaneously. Each of the relays and receivers obtains a faded linear combination of transmitted signals. The jth relay node then multiplies its received signal by a scalar d_j (a function of the fading coefficients) and forwards its resulting signal to the receivers while the m transmitting nodes are silent. Thus we have a two-stage process.

In Stage 1, if y_k denotes the signal received at the kth receiver and s_j the signal received at the jth relay node we may write

$$
\begin{aligned}
y_k^{(1)} &= \sqrt{P}\sum_{i=1}^{m} f_{ik}x_i + w_k^{(1)} \quad k = 1, 2, \ldots, m, && (5.12)\\
s_j &= \sqrt{P}\sum_{i=1}^{m} g_{ij}x_i + v_j \quad j = 1, 2, \ldots, n, && (5.13)
\end{aligned}
$$

where w_k and v_j are additive complex Gaussian noise at the kth receiver and jth relay node respectively. The noise sources are independent. The constant P represents the power of each transmitted signal so as to normalize x_i: we assume $\mathbf{E}|x_i|^2 = 1$ and that each transmitter has the same average power P. Also assume w_k is zero-mean complex Gaussian noise with $\mathbf{E}|w_k|^2 = 1$.

In Stage 2, the kth receiver obtains a faded linear combination of signals from the relays:

$$
\begin{aligned}
y_k^{(2)} &= \sum_{j=1}^{n} s_j d_j h_{jk} + w_k^{(2)}\\
&= \sqrt{P}\sum_{j=1}^{n}\sum_{i=1}^{m} x_i g_{ij} d_j h_{jk} + \sum_{j=1}^{n} v_j d_j h_{jk} + w_k^{(2)},
\end{aligned}
$$

for $k = 1, 2, \ldots, m$. In matrix form,

$$
[(\mathbf{y}^{(1)})^T (\mathbf{y}^{(2)})^T] = \sqrt{P}\mathbf{x}^T[F \quad GDH] + \mathbf{v}^T[0 \quad DH] + [(\mathbf{w}^{(1)})^T (\mathbf{w}^{(2)})^T], \qquad (5.14)
$$

where f_{ik} is the (i, k) entry of F, g_{ij} is the (i, j) entry of G, D is diagonal with d_k as its kth entry, and H has h_{jk} as its (j, k) entry.

Stage 1 of transmission can be shown to add one to the diversity established by Stage 2—effectively it adds one independent path f_{kk} from sender to receiver. From now we focus on the diversity arising from Stage 2 of transmission.

Note that the m receivers receive both their desired signals as well as (potentially) $m-1$ interference terms. In [52] it has been shown that decoding signals assuming the interference terms are noise results in a diversity of zero, for any transmission rate. Therefore [52] focused on MAC-type decoding (see, e.g., [83]) to obtain diversity results. Here the relays assist in eliminating the interference. Thus we choose D to satisfy:

$$G_{m\times n} D_{n\times n} H_{n\times m} = cI_{m\times m} \tag{5.15}$$

for a scalar c, and where $I_{m\times m}$ represents the $m\times m$ identity matrix. As mentioned earlier, the relay nodes are assumed to have knowledge of the channels and so are able to compute D. Of course, since we have m^2 equations in (5.15) and n unknowns (not including c), this is only possible if $n \geq m^2$. We will presently assume this and in Section 5.3.3 comment on the case $n < m^2$.

Then we have

$$\mathbf{y}^T = \sqrt{P} c \mathbf{x}^T + \mathbf{v}^T DH + \mathbf{w}. \tag{5.16}$$

That is, each receiver obtains only its intended signal (plus additive noise). To ensure unit amplification by each relay on average we additionally impose the condition $\sum_{j=1}^{n} {d_j}^2 = n$. From this constraint, we find c as follows: solve the equation $G\hat{D}H = I$ for $\hat{D}$, then let $D = \hat{D}/||\hat{D}||$ which has the required norm n. Then $cI = GDH = G\hat{D}H/||\hat{D}||$ from which $c = 1/||\hat{D}||$. When $n > m^2$ there are infinitely many solutions to $G\hat{D}H = I$, and we will later argue that the optimal choice is the minimum norm solution. We assume each receiver has knowledge of the scalar c.

We remark that the power efficiency of this interference-cancellation scheme was studied in Section V-G of [11]. A similar interference cancellation scheme is considered in [6] where capacity scaling laws are derived in the asymptotic case of a large number of relays. This requires less channel knowledge at the relays, but at least m^3 relay nodes are required. Unlike [6], in this work we are primarily interested in diversity.

5.3.2 Main Result—Outage Behavior

We now analyze the error probability behavior of the interference cancellation scheme defined in the previous section. From (5.16) and recalling that each receiver knows c we can find the mutual information for the ith transmitter-receiver pair and say that that pair is in outage if for a given rate R_i the instantaneous mutual information is below R_i:

$$R_i > \log\left(1 + \frac{Pc^2}{1 + \sigma_v^2 ||(DH)_i||^2}\right), \tag{5.17}$$

for $i = 1, 2, \ldots, m$. Here σ_v^2 represents the noise variance at each relay (from now we may assume it to equal 1 as this does not affect trade-off analysis) and $(DH)_i$ is the ith column of DH.

Assume we use a coding scheme which is *approximately universal*, meaning that at high SNR an arbitrarily low probability of error can be achieved when the channel is not in outage [68]. If no such coding scheme exists, then the analysis which follows will provide a *lower bound* on the error probability (upper bound on diversity).

Hence from now we assume the dominant error event for the interference cancellation network is outage—that is one of the users' data rates cannot be supported due to fading.

We claim that the probability of outage described by (5.17) has the same behavior at high SNR as $\Pr(\log(1 + Pc^2) < R_i)$, or in exponential equality notation we have the following, proved in the appendix:

Lemma 7. *For the model described by (5.16) and the text thereafter, for $P \to \infty$ we have*

$$\Pr\left(\log\left(1+\frac{Pc^2}{1+||(DH)_i||^2}<R_i\right)\right) \dot{\leq} \Pr\left(\log\left(1+Pc^2\right)<R_i\right).$$

Hence it suffices to consider

$$\Pr\left(\log\left(1+Pc^2\right)<R_i\right).$$

If $n > m^2$ we can minimize this probability over the infinitely many choices of D satisfying (5.15) by choosing D to have minimal norm. We can rewrite the system $GDH = cI$ in the form $Ad = b$, where

$$A = \begin{bmatrix} g_{11}h_{11} & g_{12}h_{21} & \cdots & g_{1n}h_{n1} \\ g_{11}h_{12} & g_{12}h_{22} & \cdots & g_{1n}h_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ g_{11}h_{1m} & g_{12}h_{2m} & \cdots & g_{1n}h_{nm} \\ g_{21}h_{11} & g_{22}h_{21} & \cdots & g_{2n}h_{n1} \\ g_{21}h_{12} & g_{22}h_{22} & \cdots & g_{2n}h_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ g_{21}h_{1m} & g_{22}h_{2m} & \cdots & g_{2n}h_{nm} \\ \vdots & \vdots & \vdots & \vdots \\ g_{m1}h_{1m} & g_{m2}h_{2m} & \cdots & g_{mn}h_{nm} \end{bmatrix}, \tag{5.18}$$

$$d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}, \quad b = \mathrm{vec}(I_{m \times m}). \tag{5.19}$$

(The vec operator stacks the columns of a matrix, forming a column vector.) The d which minimizes $||d||^2 = ||D||^2$ in this underdetermined system of equations is found by applying the pseudoinverse:

$$d = A^*(AA^*)^{-1}b.$$

Hence

$$\min ||D||^2 = d^*d = b^*(AA^*)^{-1}b.$$

Determining the outage probability behavior at high SNR then amounts to finding $\Pr\left(\frac{1}{b^*(AA^*)^{-1}b} < \epsilon\right)$. It turns out that this is of order ϵ^{n-m^2+1} which leads to the following main result whose proof is given in Section 5.3.4.

Theorem 6. *Consider the two-stage interference cancellation scheme described by (5.14), with m transmitter-receiver pairs and n relay nodes (where $n > m^2$). Choose D to satisfy (5.15) and have norm $||D||^2 = n$. Then if the ith transmitter has multiplexing gain r_i, the maximum diversity of the system is at least $d = (n - m^2 + 2)(1 - \max_i r_i)$.*

5.3.3 Discussion

Theorem 6 implies that provided relay nodes allow for independent fades to and from the transmitter/receiver pairs, diversity can be made to grow linearly in the number of relay nodes for rates up to the maximum possible sum rate $S = m/2$.

Of course, we need not insist that all m nodes are transmitting at once. In fact, we cannot do so when $n < m^2$. In this case, one can choose $K < m$ source destination pairs and perform the interference cancellation scheme using the remaining $2m + n - 2K$ nodes as relays (which requires $2m + n - 2K \geq K^2$ or $K \leq \sqrt{2m+n+1} - 1$). If we perform this for all $\binom{M}{k}$ possible combinations of K transmit/receive pairs, then each pair will be communicating a fraction K/m of the time. Thus, assuming equal multiplexing gains $r_i = r$, the sum-rate now is $S = rK/2$. We therefore have the following result.

Theorem 7. *Consider the two staged interference cancellation scheme just described. Then, if the multiplexing gains of all users are equal, $r_i = r$, then for a given sum-rate S, the diversity is at least*

$$d(S) = \max_{K \leq \sqrt{2m+n+1}-1} (2m + n - 2K - K^2 + 2)(1 - 2S/K). \tag{5.20}$$

5.3.3.1 A MAC-Based Scheme

For the sake of comparison, consider a variant of the second cooperative diversity scheme protocol analyzed in Section 5.2.2.2, where D no longer is chosen to cancel interference but rather a MAC-based decoder [83] is used to deal with interference. In this scheme, initially K of the m transmitters transmit their signals simultaneously as before while the corresponding K receivers listen. In the next K stages the remaining $2(m - K) + n$ nodes transmit what they received to one of the receivers by choosing their phases in such away that the signals are received coherently for that receiver. This is repeated K times for each receiver. This $(K + 1)$-stage process then is repeated for all $\binom{m}{K}$ combinations of k transmit/receive pairs. If we assume all the multiplexing gains are equal the sum rate of this scheme is $S = \frac{K}{K+1} m r_i$. In the previous section we saw that such a scheme achieves diversity equal to the number of independent paths established between transmitter and receiver: $2(m-K)+n+1$.

Maximizing this over choices of K we obtain the following trade-off curve:

$$d(S) = \max_{K\in\{1,\ldots,m\}} (2m-2K+n+1)\left(1-\frac{1+K}{K}S\right). \tag{5.21}$$

Functions (5.20) and (5.21) are plotted together with the interference cancellation scheme in figure 5.8 for the cases $m=2$, $n=6$ and $m=3$, $n=10$. We see that both curves match at low rates, at intermediate rates the MAC-based scheme achieves higher diversity, but the interference cancellation scheme is able to achieve diversity at higher rates. In fact, the main motivation for considering the scheme of this work is that the MAC scheme allows only sum-rates up to $\frac{m}{m+1}$, whereas interference cancellation allows for transmission up to a rate of $S = \frac{\sqrt{2m+n+1}-1}{2}$.

5.3.4 Proof of Main Result

In this section we prove Theorem 6. We wish to show that if A and b have the form given in (5.18) then $\Pr\left(\frac{1}{b^*(AA^*)^{-1}b} < \epsilon\right) \sim \epsilon^{n-m^2+1}$. The final result follows from noting that the direct transmission of Stage 1 adds one to the diversity.

Order the singular values of A as $0 < \sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_{m^2}$. Then the eigenvalues of $(AA^*)^{-1}$ are $\frac{1}{\sigma_1^2} > \frac{1}{\sigma_2^2} > \ldots > \frac{1}{\sigma_{m^2}^2}$. This implies

$$\begin{aligned}\frac{1}{b^*(AA^*)^{-1}b} \geq \frac{\sigma_1^2}{||b||^2} &= \frac{\sigma_1^2}{m}\\ \Rightarrow \Pr\left(\frac{1}{b^*(AA^*)^{-1}b} < \epsilon\right) &\leq \frac{1}{m}\Pr\left(\sigma_1^2 < m\epsilon\right).\end{aligned}$$

By performing this bound we lose the structure of b but encouraged by figure 5.9 the

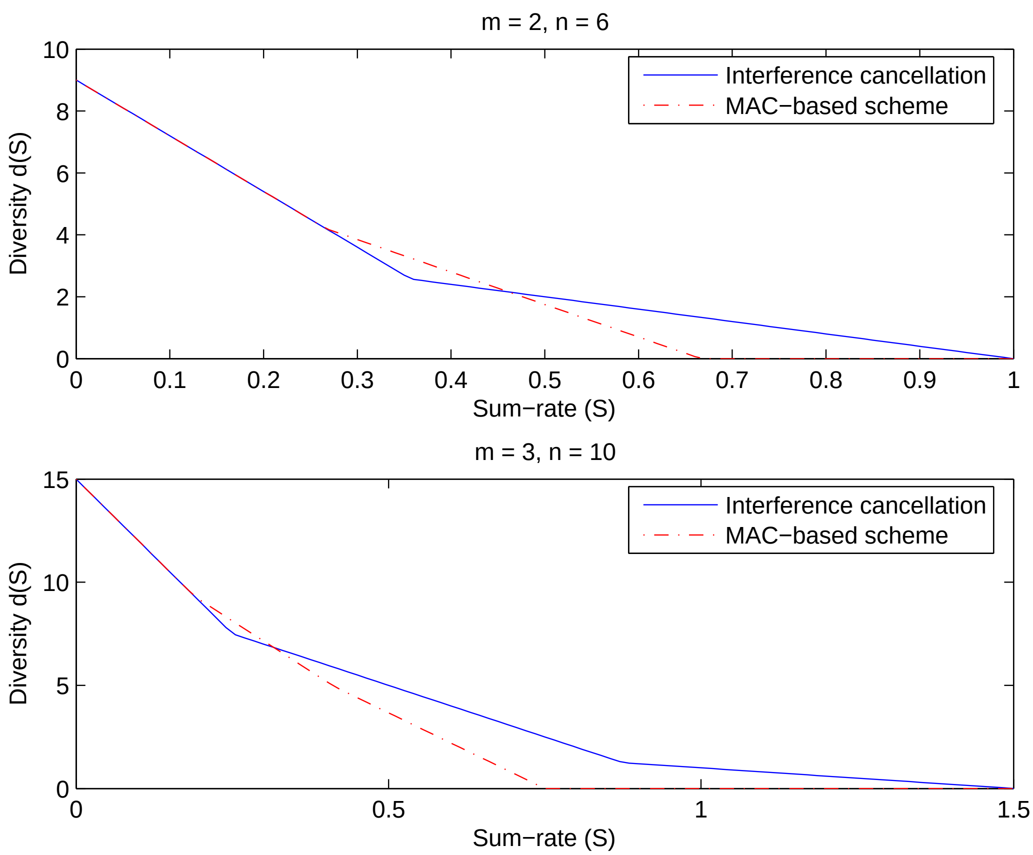

Figure 5.8: Interference cancellation and MAC-based diversity-rate trade-off curves.

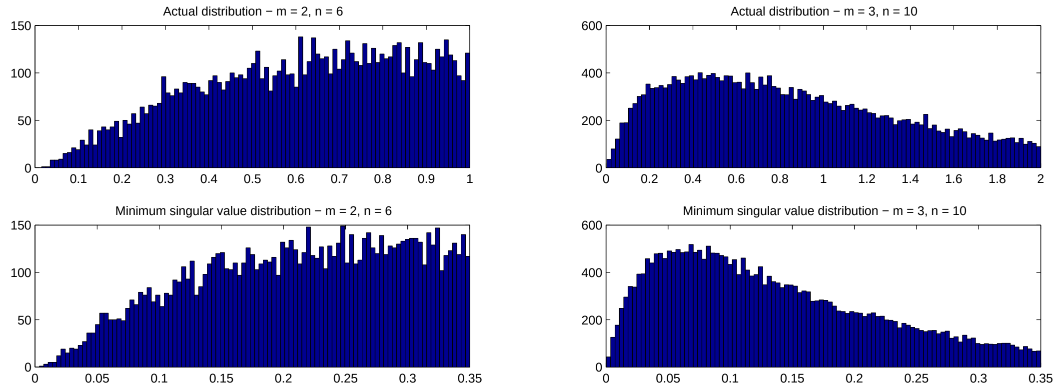

Figure 5.9: Simulated probability density of $\frac{1}{b^*(AA^*)^{-1}b}$ (top histograms) and the minimum singular value of A defined in (5.18) (lower histograms). We claim that both have diversity $n - m^2 + 1$, which can be seen by the derivative at the origin behaving like ϵ^{n-m^2} ($m = 2$, $n = 6$ for the left plots, $m = 3$, $n = 10$ for the right plots).

diversity is unchanged when making this approximation. In other words the bound is sufficiently tight for our purposes.

Now we wish to prove that $\Pr\left(\sigma_1^2(A) < \epsilon\right) \sim \epsilon^{n-m^2+1}$. The problem has been reduced to identifying the distribution of the smallest singular value of A near the origin. If the matrix A were to have independent $\mathbb{C}\mathcal{N}(0,1)$ entries, this result can be shown to be true by adopting the approach of [82] using the known joint distribution of the eigenvalues. In our case the entries of A in (5.18) are dependent, and so computing the eigenvalue distribution is far more involved, and we adopt an alternative approach.

The set of $m^2 \times n$ matrices may be viewed as points in the vector space $\mathbb{C}^{m^2 n}$. Matrices of the form A in (5.18) form a lower-dimensional submanifold; denote this space by T. That is,

$$T := \{A \in \mathbb{C}^{m^2 \times n} : A_{ik}^j = g_{ij} h_{jk}, \text{where } g_{ij}, h_{jk} \in \mathbb{C},\ i = 1, \ldots, m,\ j = 1, \ldots, n,\ k = 1, \ldots, m\}. \tag{5.22}$$

Assigned to T is a probability distribution induced by the complex Gaussian variables g_{ij} and h_{jk}. The matrices with smallest squared singular value less than ϵ will lie within a neighborhood of radius $\sqrt{\epsilon}$ of the submanifold of matrices having rank lower than m^2 (call this submanifold U and the neighborhood U_ϵ). In other words,

$$U := \{A \in \mathbb{C}^{m^2 \times n} : \sigma_{\min}(A) = 0\}, \tag{5.23}$$

$$U_\epsilon := \{B \in \mathbb{C}^{m^2 \times n} : \exists A \in U \text{ such that } ||B - A||^2 < \epsilon\}. \tag{5.24}$$

To see this, writing the singular value decomposition of A as $Q_1 \Sigma Q_2$ where Q_1 and Q_2 are unitary and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{m^2})$. Then letting $\Sigma' = \mathrm{diag}(0, \sigma_2, \ldots, \sigma_{m^2})$ the matrix $A' := Q_1 \Sigma' Q_2$ is not of full rank and

$$\begin{aligned}
||A - A'||^2 &= ||Q_1(\Sigma - \Sigma')Q_2||^2 \\
&= ||Q_1 \mathrm{diag}(\sigma_1, 0, \ldots, 0) Q_2||^2 \\
&= ||q_1 \sigma_1 q_2||^2 \\
&\quad \text{where } q_1 \text{ is the first column of } Q_1 \text{ and } q_2 \text{ is the first row of } Q_2 \\
&= \sigma_1^2 \mathrm{tr}(q_1 q_2)(q_1 q_2)^* \\
&= \sigma_1^2 \mathrm{tr} q_1 q_2 q_2^* q_1^* \\
&= \sigma_1^2 \mathrm{tr} q_1 q_1^* \quad \text{since } q_2 \text{ is a unit row vector} \\
&= \sigma_1^2 \mathrm{tr} q_1^* q_1 \\
&= \sigma_1^2 \mathrm{tr} 1 \quad \text{since } q_1 \text{ is a unit vector} \\
&= \sigma_1^2.
\end{aligned}$$

Hence if $\sigma_1^2 < \epsilon$ then A will be within distance $\sqrt{\epsilon}$ of a matrix of lower rank (A').

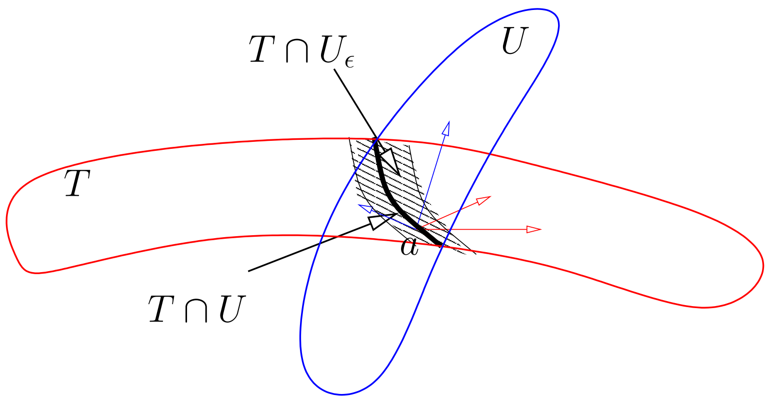

Figure 5.10: Intersection of submanifolds T and U. To identify the region $T \cap U_e$ we find for each $a \in T \cap U$ the vectors tangent to T and U at a. This is found from the parametrizations of T and U.

We are interested in the probability that a matrix from T is in U_ϵ. This will be given by integration of the probability density function of A over the region $T \cap U_\epsilon$. As we shall see the density function is sufficiently well behaved, so the answer will be directly related to the volume of the region $T \cap U_\epsilon$. This requires knowledge of how the submanifolds T and U intersect, which in turn involves computation of their tangent spaces at a (see figure 5.10). This is based on the respective parametrizations of T and U at a, which we consider next.

5.3.4.1 Parametrization of Submanifolds

All $m^2 \times n$ matrices of rank at most $m^2 - 1$ (recall $n > m^2$) may be specified by $m^2 - 1$ of the vectors, with the remaining vectors being linear combinations of these. Hence we may write

$$A = [X|XG], \tag{5.25}$$

where X is an $m^2 \times (m^2 - 1)$ matrix and G is an $(m^2 - 1) \times (n - m^2 + 1)$ matrix of complex numbers specifying the linear combinations.

Observe that the columns of A are independent. For any $j \in \{1, 2, \ldots, n\}$ the $2m - 1$

values,

$$g_{1j}h_{j1}, g_{1j}h_{j2}, \ldots, g_{1j}h_{jm}, g_{2j}h_{j1}, g_{3j}h_{j1}, \ldots, g_{mj}h_{j1}, \tag{5.26}$$

specify any other entry of A. This is true by the identity

$$g_{ij}h_{jk} = \frac{g_{1j}h_{jk} \times g_{ij}h_{j1}}{g_{1j}h_{j1}}.$$

Hence by permuting rows as necessary, a parametrization for A may be given by

$$A = \left[\begin{array}{c} Y \\ \hline f(Y) \end{array} \right], \tag{5.27}$$

where Y is a $(2m-1) \times n$ matrix, and $f(Y)$ is an $(m-1)^2 \times n$ matrix whose entries are of the form $y_{ij}y_{kj}/y_{1j}$, where i ranges from 2 to m, k ranges from $m+1$ to $2m-1$ and j ranges from 1 to n.

5.3.4.2 Integration over Submanifolds

For a given point $a \in T \cap U$ we identify the number of linearly independent directions from a normal to U and tangential to T. This allows us to carry out the integration over two stages. Firstly we integrate over $T \cap U_\epsilon$ these dimensions. In the second stage we integrate over all $a \in T \cap U$. That is, we perform the change of variable $A \to (a, w)$ where $a \in T \cap U$ and $w \in T \cap U_\epsilon$ and the desired probability may be written in the form

$$\begin{aligned} \Pr\left(\sigma_1^2 < \epsilon\right) &= \int_{A \in T \cap U_\epsilon} p(A)\; dA \\ &= \int_{a \in T \cap U} \int_{\substack{w \in T \\ \|w-a\|^2 < \epsilon}} p(a, w) \sqrt{\det[J(a,w)J(a,w)^*]}\; dw da, \end{aligned} \tag{5.28}$$

where $p(a, w)$ is the probability density function of Y evaluated at $(a, w) \in T$ (given in Lemma 8). The inner integral is done in directions normal to U at a.

The determinant factor here represents the volume of an infinitesimal element in the m^2n-dimensional space: $J(a, w)$ is the matrix whose columns span the tangent space of $T \cap U_\epsilon$ at a. The inner integral is done in directions normal to U at a and may be approximated by an integral over a sphere of radius $\sqrt{\epsilon}$ centered at a.

The following lemma (proof in the appendix of this chapter) gives the pdf of matrices in T explicitly.

Lemma 8. *The density of the $(2m-1) \times n$ matrix Y, having jth column with $(2m-1)$ values described in (5.26) (where the g's and h's are independent $\mathbb{C}\mathcal{N}(0,1)$ variables), is given by*

$$p(Y) = \prod_{j=1}^{n} \frac{2}{\pi^{2m-1}|y_{1j}|^{2(m-1)}} \times K_0\left(\frac{2}{|y_{1j}|}\sqrt{\left(\sum_{i=1}^{m}|y_{ij}|^2\right)\left(|y_{1j}|^2 + \sum_{i=m+1}^{2m-1}|y_{ij}|^2\right)}\right),$$

where $K_0(x) := \int_0^\infty \exp(-t^2 - x^2/4t^2)\, dt$ is a modified Bessel function of the second kind.

The Bessel function $K_0(x)$ is monotonically decreasing in x and known [2] to behave like $\ln(1/x)$ as $x \to 0$ and like $\exp(-x)$ as $x \to \infty$. Furthermore the determinant in (5.28) is a ratio of two polynomials in its entries, so the asymptotic behavior of the integrand is dominated by that of the density function p. Knowledge of the asymptotic behavior of p enables us to upper bound the inner integral of (5.28) as

$$(\epsilon \ln(1/\epsilon))^{k(a)} g(a), \tag{5.29}$$

where $g : T \cap U \to \mathbb{R}$ is some function independent of ϵ and $g(a) \sim \ln(1/||a||)$ as $||a|| \to 0$ and

$g(a) \sim \exp(-||a||)$ as $||a|| \to \infty$. The inner integral is over a small neighborhood of a, and since the integrand is continuous, its value by the mean value theorem is close to the volume of that neighborhood times the value of the density function at a. This neighborhood is approximated by a sphere of dimension $k(a)$ centered at a and of radius $\sqrt{\epsilon}$. The additional $\ln(1/\epsilon)$ factor takes into account the behavior of p near the origin.

The next result ensures that the small neighborhood in this integral is of the appropriate dimension for our main result to hold.

Lemma 9. *In (5.29), the upper bound for the inner integral of (5.28), $k(a) = n - m^2 + 1$ almost surely.*

Proof. Given the parametrizations we take derivatives of each entry of $a \in T$ with respect to each variable to form a matrix whose columns are tangent vectors. Let B and E respectively be the $m^2n \times (2m-1)n$ and $m^2n \times (n+1)(m^2-1)$ matrices corresponding to T and U.

From (5.25) the corresponding matrix for the submanifold U of matrices of smallest singular value zero is

$$E = \begin{bmatrix} I_{m^2(m^2-1)} & 0_{m^2(m^2-1)\times(m^2-1)(n-m^2+1)} \\ G^T \otimes I_{m^2} & I_{n-m^2+1} \otimes X_{m^2\times(m^2-1)} \end{bmatrix}, \tag{5.30}$$

where $\otimes$ represents the Kronecker product of matrices [26]. E has dimension $m^2n \times (n+1)(m^2-1)$. From (5.27) the corresponding matrix for submanifold T is

$$B = \operatorname{diag}(B_1, B_2, \ldots, B_n), \tag{5.31}$$

where

$$B_j = \begin{bmatrix} I_{2m-1} \\ \\ C_j \end{bmatrix},$$

and C_j is an $(m-1)^2 \times (2m-1)$ block containing derivatives of $f(Y)$, thus having terms of the form y_{ij}/y_{1j} or $-y_{ij}y_{kj}/y_{1j}^2$.

E can be shown to be full rank, which tells us there are $n - m^2 + 1$ linearly independent directions from a normal to U (in the nullspace of U). For all of these directions to be in the tangent space of $a \in T$ require the augmented matrix $[B|E]$ to be full rank (i.e., its columns span the m^2n-dimensional space). Equivalently this condition requires

$$\begin{aligned} \det[B|E][B|E]^* &\neq 0, \\ \text{i.e.,} \quad \det(BB^* + EE^*) &\neq 0. \end{aligned} \tag{5.32}$$

If this were not the case, the submanifolds S and T as illustrated in figure 5.10 would not intersect transversally and we can no longer claim that there are $n - m^2 + 1$ independent directions in which $T \cap U_\epsilon$ is small.

This determinant (5.32) can be shown to be not identically equal to zero by finding an instance of a for which the determinant is non-zero. As the determinant is a smooth function of its entries, this tells us the determinant is non-zero almost surely for $a \in U \cap T$.

From (5.30) and (5.31) we observe that

$$
\begin{aligned}
& BB^* + EE^* \\
= \; & \operatorname{diag}\{B_1B_1^*, B_2B_2^*, \ldots, B_nB_n^*\} + \begin{bmatrix} I & \bar{G} \otimes I \\ G^T \otimes I & (G^T\bar{G} \otimes I) + (I \otimes XX^*) \end{bmatrix} \\
= \; & \begin{bmatrix} I + \operatorname{diag}\{B_1B_1^*, \ldots, B_{m^2-1}B_{m^2-1}^*\} & \bar{G} \otimes I \\ G^T \otimes I & (G^T\bar{G} \otimes I) + (I \otimes XX^*) + \operatorname{diag}\{B_{m^2}B_{m^2}^*, \ldots, B_nB_n^*\} \end{bmatrix}.
\end{aligned}
$$

Using the result

$$
\det \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} = \det \mathcal{A} \det(\mathcal{D} - \mathcal{C}\mathcal{A}^{-1}\mathcal{B}),
$$

we see that $\det(BB^*+EE^*)$ is non-zero if and only if both $\det(I+\operatorname{diag}(B_1B_1^*, \ldots, B_{m^2-1}B_{m^2-1}^*))$, and the determinant of its Schur complement [26] in $BB^* + EE^*$ (i.e., $\det(\mathcal{D} - \mathcal{C}\mathcal{A}^{-1}\mathcal{B})$) are each non-zero. The former is true since it is the determinant of the sum of a positive definite and positive semidefinite matrix. To see that the Schur complement does not have identically zero determinant, we evaluate it for one easily computable case of $a \in U \cap T$.

One such instance is letting

$$
G = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}.
$$

This corresponds to the last $n - m^2 + 1$ columns of A each being copies of the first column.

In this case G^TG is a matrix having all 1s. Since

$$\begin{bmatrix} I & 0 & 0 & \dots & 0 \\ I & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I & 0 & 0 & \dots & 0 \end{bmatrix} \mathcal{A} \begin{bmatrix} I & I & \dots & I \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{11} & \dots & \mathcal{A}_{11} \\ \mathcal{A}_{11} & \mathcal{A}_{11} & \dots & \mathcal{A}_{11} \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{A}_{11} & \mathcal{A}_{11} & \dots & \mathcal{A}_{11} \end{bmatrix},$$

where $\mathcal{A}_{11}$ is the top left subblock of $\mathcal{A}$, the Schur complement is

$$\mathrm{diag}\{XX^*+B_{m^2}B_{m^2}^*, \dots, XX^*+B_{n^2}B_{n^2}^*\}+\begin{bmatrix} I-(I+B_1B_1^*)^{-1} & \dots & I-(I+B_1B_1^*)^{-1} \\ I-(I+B_1B_1^*)^{-1} & \dots & I-(I+B_1B_1^*)^{-1} \\ \vdots & \vdots & \vdots \\ I-(I+B_1B_1^*)^{-1} & \dots & I-(I+B_1B_1^*)^{-1} \end{bmatrix}.$$

This, being the sum of a positive definite and positive semidefinite matrix, has non-zero determinant. Hence $\det(BB^* + EE^*) \neq 0$ in this instance and we conclude that almost surely there exist $n - m^2 + 1$ linearly independent directions from a normal to U contained within the tangent space of $a \in T$. $\square$

From Lemma 9 the outer integral becomes simple—the integrand is now bounded above by a constant multiple of $p(a)(\epsilon \ln(1/\epsilon))^{k(a)}$. Since $g(a)$ has the appropriate asymptotic properties and k is almost surely $n - m^2 + 1$, the outer integral will be some constant multiplied by $(\epsilon \ln(1/\epsilon))^{(n-m^2+1)} \sim \epsilon^{n-m^2+1}$, and we are done. That is, the maximum diversity of the system is $n - m^2 + 2$.

5.3.4.3 Special case: $m = 1$

The previous analysis is unnecessary in the smallest case of $m = 1$. In this instance, A in (5.18) has n copies of $f_{1k}g_{k1}$, so $\sigma_1^2 = \sum_{k=1}^{n} |f_{1k}g_{k1}|^2$. Hence

$$\begin{aligned}
\Pr\left(\sigma_1^2 < \epsilon\right) &= \Pr\left(\sum_{k=1}^{n} |f_{1k}g_{k1}|^2 < \epsilon\right) \\
&\leq \Pr\left(|f_{1k}g_{k1}|^2 < \epsilon,\ k = 1, \ldots, n\right) \\
&= \left(\Pr\left(|f_{1k}g_{k1}|^2 < \epsilon\right)\right)^n \\
&\leq (\epsilon \ln(1/\epsilon))^n \text{ from (5.8).}
\end{aligned}$$

Hence, after incorporating the direct source-destination link, the diversity is $n + 1 = n - m^2 + 2$.

5.4 Summary

Two types of wireless networks were investigated. Firstly we studied a two-user interference channel with and without cooperation and derived the results of table 1, calculating the rate-diversity relationship for well-known transmission schemes. In the case when nodes do not cooperate the diversity is at most one, but rate can be maximal. We have shown a cooperative strategy where this diversity may be increased up to three, but at the cost of a significant reduction in the rate. Such diversity increases require coherent addition at the receivers, which in turn requires channel knowledge at the relays.

Secondly we studied a network with relays and showed how channel knowledge at the relays can enable interference cancellation to occur at the receivers. For a sufficiently large number of relay nodes, this can cause a linear increase in diversity. It would be worth studying further whether the method of analysis given can be applied to other wireless schemes.

5.5 Appendix—Proof of Lemmas

Lemma 7. *For the model described by (5.16) and the text thereafter, for $P \to \infty$ we have*

$$\Pr\left(\log\left(1+\frac{Pc^2}{1+||(DH)_i||^2} < R_i\right)\right) \dot{\leq} \Pr\left(\log\left(1+Pc^2\right) < R_i\right).$$

Proof. Let $\epsilon = (2^{R_i}-1)/P$ which is small for P large. For any $x > 0$ we have

$$\begin{aligned}
& \Pr\left(\frac{c^2}{1+||DH_i||^2} < \epsilon\right) \\
= \ & \Pr\left(\frac{c^2}{1+||DH_i||^2} < \epsilon, ||DH_i||^2 < x)\right) \\
& \quad + \Pr\left(\frac{c^2}{1+||DH_i||^2} < \epsilon \middle| \, ||DH_i||^2 > x\right). \Pr\left(||DH_i||^2 > x\right) \\
\leq \ & \Pr\left(c^2 < \epsilon(1+x), ||DH_i||^2 < x\right) \\
& \quad + \Pr\left(\frac{c^2}{1+||DH_i||^2} < \epsilon \middle| \, ||DH_i||^2 < x\right). \Pr\left(||DH_i||^2 > x\right) \\
\leq \ & \Pr(c^2 < \epsilon(1+x)) + \Pr(||DH_i||^2 > x).
\end{aligned} \tag{5.33}$$

Working on the second term here, we have

$$\begin{aligned}
||DH_i||^2 &= \sum_{j=1} |d_j h_{ij}|^2 \\
&\leq \sum_{j=1}^{n} |d_j|^2 \sum_{j=1}^{n} |h_{ij}|^2 \\
&= n \sum_{j=1}^{n} |h_{ij}|^2 \\
&\leq n \max_j |h_{ij}|^2.
\end{aligned}$$

Hence we find

$$
\begin{aligned}
\Pr\left(||DH_i||^2 > x\right) &\leq \Pr\left(n \max_j |h_{ij}|^2 > x\right) \\
&= 1 - (1 - e^{-x/n})^n \\
&\quad \text{(using the cdf expression for the maximum of } n \\
&\quad \text{independent exponential random variables)} \\
&\leq 1 - (1 - ne^{-x/n}) \\
&= ne^{-x/n}.
\end{aligned}
$$

That is, since $||DH_i||^2$ has an exponential tail, we may set x to $K\ln(1/\epsilon)$, for some constant K independent of ϵ, so that, from the two terms of (5.33) can both be made to be of the same order:

$$
\begin{aligned}
& \Pr\left(\frac{c^2}{1+||DH_i||^2} < \epsilon\right) \\
\leq\ & \Pr(c^2 < \epsilon(1+x)) + \Pr(||DH_i||^2 > x) \\
\doteq\ & \Pr(c^2 < \epsilon(1+x)) \\
\doteq\ & \Pr\left(\log(1+Pc^2) < R_i\right),
\end{aligned}
$$

and we are done. $\square$

Lemma 8. *The density of the $(2m-1)\times n$ matrix Y, having jth column with $(2m-1)$ values described in (5.26) (where the g's and h's are independent $\mathbb{C}\mathcal{N}(0,1)$ variables), is* *given by*

$$p(Y) = \prod_{j=1}^{n} \frac{2}{\pi^{2m-1}|y_{1j}|^{2(m-1)}} \times K_0\left(\frac{2}{|y_{1j}|}\sqrt{\left(\sum_{i=1}^{m}|y_{ij}|^2\right)\left(|y_{1j}|^2 + \sum_{i=m+1}^{2m-1}|y_{ij}|^2)\right)}\right),$$

where $K_0(x) := \int_0^\infty \exp(-t^2 - x^2/4t^2)\, dt$ *is a modified Bessel function of the second kind.*

Proof. The columns of Y are independent so $p(Y) = \prod_{k=1}^{n} p(Y_k)$ where Y_k is the kth column of Y. Hence we wish to show that given $f_1, \ldots, f_m, g_1, \ldots, g_m \sim \mathbb{C}\mathcal{N}(0,1)$, the joint distribution of $u_{ij} := f_i g_j$ is given by

$$p(u_{11}, \ldots, u_{1m}, u_{21}, \ldots, u_{m1}) = \frac{2}{\pi^{2m-1}|u_{11}|^{2(m-1)}} K_0\left(\frac{2}{|u_{11}|}\sqrt{\sum_{i=1}^{m}|u_{i1}|^2 \sum_{j=1}^{m}|u_{1j}|^2}\right).$$

Define the set of random variables

$$Z_1 = |f_1|, Z_2 = |f_2 g_1|, \ldots, Z_m = |f_m g_1|, Z_{m+1} = |f_1 g_1|, Z_{m+2} = |f_1 g_2|, \ldots, Z_{2m} = |f_1 g_m|. \tag{5.34}$$

Then observe that

$$\prod_{i=1}^{2m} Z_i = |f_1|^m |g_1|^{m-1} \prod_{i=1}^{m} |f_i||g_i|. \tag{5.35}$$

The Jacobian of this transformation is formed by taking derivatives of the Z's with respect to the $|f|$'s and $|g|$'s.

$$\frac{\partial(Z_i)}{\partial(|f_j|,|g_j|)} = \begin{bmatrix} 1 & & & & |g_1| & |g_2| & \dots & |g_m| \\ & |g_1| & & & & & & \\ & & \ddots & & & & & \\ & & & |g_1| & & & & \\ 0 & |f_2| & \dots & |f_m| & |f_1| & & & \\ & & & & & |f_1| & & \\ & & & & & & \ddots & \\ & & & & & & & |f_1| \end{bmatrix}.$$

The determinant of this matrix is $(|f_1||g_1|)^{m-1}|f_1|$. Then, by the change of variable formula,

$$\begin{aligned} p(Z_1,\dots,Z_{2m}) &= p(|f_1|,\dots,|f_m|,|g_1|,\dots,|g_m|)\left|\det\frac{\partial(Z_i)}{\partial(|f_j|,|g_j|)}\right|^{-1} \\ &= \prod_{i=1}^{m} 2|f_i|e^{-|f_i|^2}\prod_{i=1}^{m} 2|g_i|e^{-|g_i|^2}[(|f_1||g_1|)^{m-1}|f_1|]^{-1} \\ &\quad \text{since the } |f|s \text{ and } |g|s \text{ are i.i.d. random variables with Rayleigh} \\ &\quad \text{distribution} \\ &= 2^{2m}\frac{\prod_{i=1}^{2m} Z_i}{|f_1|^m|g_1|^{m-1}}e^{-\left(Z_1^2+\sum_{i=2}^{m}\frac{Z_i^2}{|g_1|^2}\right)}e^{-\sum_{i=m+1}^{2m}\frac{Z_i^2}{Z_1^2}}\frac{1}{|f_1|^{m-1}|g_1|^{m-1}|f_1|} \\ &\quad \text{from (5.35)} \\ &= 2^{2m}\frac{\prod_{i=1}^{2m} Z_i}{Z_1 Z_{m+1}^{2(m-1)}}e^{-\left(Z_1^2+\sum_{i=2}^{m}\frac{Z_i^2 Z_1^2}{Z_{m+1}^2}\right)}e^{-\sum_{i=m+1}^{2m}\frac{Z_i^2}{Z_1^2}}. \end{aligned}$$

Hence from (5.34),

$$
\begin{aligned}
& p(|f_1|, |u_{21}|, \ldots, |u_{m1}|, |u_{11}|, \ldots, |u_{1m}|) \\
= \quad & 2^{2m} \frac{\prod_{i=1}^{m} |u_{1i}| \prod_{i=1}^{m} |u_{i1}|}{|f_1||u_{11}|^{2(m-1)}} \exp\left(-|f_1|^2 \left(1 + \sum_{i=2}^{m} \frac{|u_{i1}|^2}{u_{11}|^2}\right)\right) \exp\left(\sum_{i=m+1}^{2m} \frac{|u_{1i}|^2}{|f_1|^2}\right).
\end{aligned}
$$

Now let us consider the phases θ_i of $f_1, f_2 g_1, \ldots, f_m g_1, f_1 g_1, \ldots, f_1 g_m$. Let ϕ_i be the phase of f_i, ψ_i be the phase of g_i, $i = 1, \ldots, m$. Then ϕ_i and ψ_i are independent and have uniform distributions from 0 to 2π. From the equations

$$
\theta_1 = \phi_i, \theta_2 = \phi_2 + \psi_1, \ldots, \theta_m = \phi_m + \psi_1, \theta_{m+1} = \phi_1 + \psi_1, \ldots, \theta_{2m} = \phi_1 + \psi_m,
$$

we have

$$
\begin{aligned}
p(\theta_1, \ldots, \theta_{2m}) &= p(\phi_1, \ldots, \phi_m, \psi_1, \ldots, \psi_m) \left| \det \frac{\partial \theta_i}{\partial(\phi_j, \psi_j)} \right|^{-1} \\
&= \left(\frac{1}{2\pi}\right)^{2m} \left| \begin{array}{ccccccc}
1 & & & & 1 & \cdots & 1 \\
 & 1 & & & & & \\
 & & \ddots & & & & \\
 & & & 1 & & & \\
0 & 1 & \cdots & 1 & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{array} \right|^{-1} \\
&= \left(\frac{1}{2\pi}\right)^{2m}.
\end{aligned}
$$

We conclude that the phases of $Z_1, \ldots, Z_{2m}$ are uniform and so the joint density of

$f_1, u_{21}, \ldots, u_{m1}, u_{11}, \ldots, u_{1m}$ is

$$
\begin{aligned}
& p(f_1, u_{21}, \ldots, u_{m1}, u_{11}, \ldots, u_{1m}) \\
= & \frac{p(|f_1|, |u_{21}|, \ldots, |u_{m1}|, |u_{11}|, \ldots, |u_{1m}|)}{2\pi|f_1| \prod_{i=1}^{m} 2\pi|u_{i1}| \prod_{i=2}^{m} 2\pi|u_{1i}|} \qquad (5.36) \\
= & \frac{1}{\pi^{2m}|f_1|^2|u_{11}|^{2(m-1)}} \exp\left(-|f_1|^2\left(1+\sum_{i=2}^{m}\frac{|u_{i1}|^2}{|u_{11}|^2}\right)\right) \exp\left(\sum_{i=m+1}^{2m}\frac{|u_{1i}|^2}{|f_1|^2}\right). \qquad (5.37)
\end{aligned}
$$

Equation (5.36) is simply the generalization of the result that if $z = re^{j\theta}$ and θ has a uniform distribution, then $p(z) = p(r)/(2\pi r)$.

Integrating (5.37) over f_1 which has uniform phase:

$$
\begin{aligned}
& p(u_{21}, \ldots, u_{m1}, u_{11}, \ldots, u_{1m}) \\
= & \int_0^\infty 2\pi|f_1| p(f_1, u_{21}, \ldots, u_{m1}, u_{11}, \ldots, u_{1m})\ d|f_1| \\
= & \frac{2}{\pi^{2m-1}|u_{11}|^{2(m-1)}} \int_0^\infty \frac{\exp\left(-|f_1|^2\left(1+\sum_{i=2}^{m}\frac{u_{i1}^2}{|f_1|^2}\right)\right)\exp\left(-\sum_{i=1}^{m}\frac{u_{1i}^2}{|f_1|^2}\right)}{|f_1|} d|f_1|. \qquad (5.38)
\end{aligned}
$$

Using the result $\int_0^\infty \frac{\exp(-|f_1|^2\alpha - \beta/|f_1|^2)}{|f_1|}\ d|f_1| = K_0(2\sqrt{\alpha\beta})$ with $\alpha = 1 + \frac{\sum_{i=2}^{m}|u_{i1}|^2}{|f_1|^2}$, $\beta = \sum_{i=1}^{m}|u_{1i}|^2$, from (5.38) we obtain

$$
p(u_{11}, \ldots, u_{1m}, u_{21}, \ldots, u_{m1}) = \frac{2}{\pi^{2m-1}|u_{11}|^{2(m-1)}} K_0\left(\frac{2}{|u_{11}|}\sqrt{\sum_{i=1}^{m}|u_{i1}|^2 \sum_{j=1}^{m}|u_{1j}|^2}\right),
$$

as required.

$\square$

Bibliography

[1] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. The capacity of discrete-time memoryless rayleigh-fading channels. *IEEE Trans. Inform. Theory*, 47:1290–1301, May 2001.

[2] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications, New York, 1974.

[3] K Azarian, H. El Gamal, and P. Schniter. On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels. *IEEE Trans. Inform. Theory*, 51:4152–4172, Dec. 2005.

[4] J.-C. Belfiore, G. Rekaya, and E. Viterbo. The golden code: A 2×2 full-rate space-time code with non-vanishing determinants. *IEEE Trans. Inform. Theory*, 51:1432–1436, Apr. 2005.

[5] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44:2619–2692, Oct. 1998.

[6] H. Bölcskei, R. Nabar, O. Oyman, and A. Paulraj. Capacity scaling laws in MIMO relay networks. *IEEE Trans. Wireless Commun.*, 5:1433–1444, Jun. 2006.

[7] J. Boyer, D. Falconer, and H. Yanikomeroglu. Cooperative connectivity models for wireless relay networks. To appear in *IEEE Trans. Wireless Commun.*, May 2007.

[8] A. B. Carleial. Interference channels. *IEEE Trans. Inform. Theory*, 24:60–70, Jan. 1978.

[9] M. Costa. On the Gaussian interference channel. *IEEE Trans. Inform. Theory*, 31:607–615, Sept. 1985.

[10] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.

[11] A. F. Dana and B. Hassibi. On the power efficiency of sensory and ad hoc wireless networks. *IEEE Trans. Inform. Theory*, 52:2890–2914, Jul. 2006.

[12] P. Dayal and M. Varanasi. An optimal two transmit antenna space-time code and its stacked extensions. In *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2003.

[13] A. Edelman. *Eigenvalues and condition numbers of random matrices.* PhD thesis, Massachusetts Institute of Technology, 1989.

[14] P. Elia, P. Raj Kumar, S. Pawar, P. Vijay Kumar, and H.-F. Lu. Explicit space-time codes achieving the diversity-multiplexing gain tradeoff. Submitted to *IEEE Trans. Inform. Theory*, Feb. 2006.

[15] R. Etkin, D. Tse, and H Wang. Gaussian interference channel capacity to within one bit: The symmetric case. In *Proc. IEEE Information Theory Workshop: ITW'06*, pages 601–605, Chengdu, Oct. 2006.

[16] G. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J.*, 1(2):41–59, Oct. 1996.

[17] H. El Gamal, G. Caire, and M. O. Damen. Lattice coding and decoding achieves the optimal diversity-multiplexing tradeoff of MIMO channels. *IEEE Trans. Inform. Theory*, 50:968–985, Apr. 2004.

[18] H. El Gamal, G.Caire, and M.O. Damen. The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff. *IEEE Trans. Inform. Theory*, 52:3601–3621, Aug. 2006.

[19] M. Godavarti and A. O. Hero-III. Multiple-antenna capacity in a deterministic Rician fading channel. In *Proc. IEEE International Symposium on Information Theory: ISIT'02*, page 251, Lausanne, Jul. 2002.

[20] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[21] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.

[22] M. Gursoy, H. Poor, and S. Verdú. Efficient signaling for low-power Rician fading channels. In *Proc. Allerton Conference on Communication, Control and Computing*, Oct. 2002.

[23] B. Hajek and V. Subramaniam. Capacity and reliability function for small peak signal constraints. *IEEE Trans. Inform. Theory*, 48:828–839, Apr. 2002.

[24] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Trans. Inform. Theory*, 27:49–60, Jan. 1981.

[25] B. Hassibi and B. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inform. Theory*, 49:951–963, Apr. 2003.

[26] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1999.

[27] A. Høst-Madsen. Capacity bounds for cooperative diversity. *IEEE Trans. Inform. Theory*, 52:1522–1544, April 2006.

[28] Y. Jing and B. Hassibi. Cooperative diversity in wireless relay networks with multiple-antenna nodes. In *Proc. IEEE International Symposium on Information Theory: ISIT'05*, pages 815–819, Adelaide, Sep. 2005.

[29] Y. Jing and B. Hassibi. Distributed space-time coding in wireless relay networks. *IEEE Trans. Wireless. Comm.*, 5:3524–3536, Dec. 2006.

[30] Y. Jing and H. Jafarkhani. Distributed differential space-time coding for wireless relay networks. To appear in *IEEE Trans. on Commun.*, 2007.

[31] E. Jorswieck and H. Boche. Multiple-antenna capacity in the low-power regime: Channel knowledge and correlation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, 3:iii/385–iii/388, Mar. 2005.

[32] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inform. Theory*, 51:3037–3063, Sep. 2005.

[33] N. Laneman, D. Tse, and G. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inform. Theory*, 50:3062–3080, Dec. 2004.

[34] N. Laneman and G. Wornell. Exploiting distributed spatial diversity in wireless networks. In *Proc. Allerton Conference on Communication, Control and Computing*, Oct. 2000.

[35] N. Laneman and G. Wornell. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Trans. Inform. Theory*, 49:2415–2425, Oct. 2003.

[36] K. Liu, V. Raghavan, and A. Sayeed. Capacity scaling and spectral efficiency in wide-band correlated MIMO channels. *IEEE Trans. Inform. Theory*, 49:2504–2526, Oct. 2003.

[37] A. Lozano, A. Tulino, and S. Verdú. Multi-antenna capacity in the low-power regime. *IEEE Trans. Inform. Theory*, 49:2527–2544, Oct. 2003.

[38] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in flat fading. *IEEE Trans. Inform. Theory*, 45:139–157, Jan. 1999.

[39] R. McEliece. *The Theory of Information and Coding.* Cambridge University Press, 2004.

[40] M. Médard and R. Gallager. Bandwidth scaling for fading multipath channels. *IEEE Trans. Inform. Theory*, 48:840–852, Apr. 2002.

[41] R. Nabar, H. Bölcskei, and F. Kneubuhler. Fading relay channels: Performance limits and space-time signal design. *IEEE Journ. Select. Areas Commun.*, 22:1099–1109, Aug. 2004.

[42] F. Oggier and B. Hassibi. An algebraic coding scheme for wireless relay networks with multiple-antenna nodes. Submitted to *IEEE Trans. Signal. Proc.*, Mar. 2006.

[43] F. Oggier and B. Hassibi. A coding strategy for wireless networks with no channel information. In *Proc. Allerton Conference on Communication, Control and Computing*, Sep. 2006.

[44] F. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo. Perfect space-time block codes. *IEEE Trans. Inform. Theory*, 52:3885–3902, Sep. 2006.

[45] S. Pawar, K. Kumar, P. Kumar, P. Elia, and B. Sethuraman. Achieving the DMD tradeoff of the MIMO-ARQ channel. In *Proc. IEEE International Symposium on Information Theory: ISIT'05*, pages 901–905, Adelaide, Sep. 2005.

[46] V. Prelov and S. Verdú. Second-order asymptotics of mutual information. *IEEE Trans. Inform. Theory*, 50:1567–1580, Aug. 2004.

[47] J. Proakis. *Digital Communications*. McGraw Hill, New York, 1995.

[48] G. Susinder Rajan and M. Sundar Rajan. A non-orthogonal distributed space-time coded protocol, part i: Signal model and design criteria. In *Proc. IEEE Information Theory Workshop: ITW'06*, Chengdu, Oct. 2006.

[49] C. Rao and B. Hassibi. Analysis of multiple antenna wireless links at low SNR. In *Proc. IEEE International Symposium on Information Theory: ISIT'03*, page 468, Yokohama, Jul. 2003.

[50] C. Rao and B. Hassibi. Analysis of multiple-antenna wireless links at low SNR. *IEEE Trans. Inform. Theory*, 50:2123–2130, Sep. 2004.

[51] C. Rao and B. Hassibi. The Gaussian interference channel at low SNR. In *Proc. IEEE International Symposium on Information Theory: ISIT'04*, page 417, Chicago, Jul. 2004.

[52] C. Rao and B. Hassibi. Rate-diversity trade-offs for interference channels with and without cooperation. In *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2006.

[53] C. Rao and B. Hassibi. Diversity-multiplexing gain trade-off of a MIMO system with relays. Submitted to *IEEE Information Theory Workshop: ITW'07*, Jul. 2007.

[54] C. Rao and B. Hassibi. High diversity scheme for wireless networks based on interference cancellation. To appear in *Proc. IEEE International Symposium on Information Theory: ISIT'07*, Jun. 2007.

[55] S. Ray, M. Médard, and L. Zheng. On non-coherent MIMO capacity in the wideband regime: Interplay of SNR, antennas and coherence. Submitted to *IEEE Trans. Inform. Theory*, Mar. 2005.

[56] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity, part I: System description. *IEEE Trans. Commun.*, 51:1927–1938, Nov. 2003.

[57] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity, part II: Implementation aspects and performance analysis. *IEEE Trans. Inform. Theory*, 51:1939–1948, Nov. 2003.

[58] C. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois, 1998.

[59] H. Shin and J. H. Lee. Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole. *IEEE Trans. Inform. Theory*, 49:2636–2647, Oct. 2003.

[60] M. Simon and M. Alouini. *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. John Wiley & Sons, New York, 2000.

[61] G. Stüber. *Principles of Mobile Communication*. Kluwer Academic Publishers, Norwell, MA, 2002.

[62] V. Subramaniam and B. Hajek. Broad-band fading channels: Signal burstiness and capacity. *IEEE Trans. Inform. Theory*, 48:809–827, Apr. 2002.

[63] Kiran T. and B. Sundar Rajan. Partially-coherent distributed space-time codes with differential encoder and decoder. *IEEE Journ. Select. Areas Commun.*, 25:426–433, Feb. 2007.

[64] V. Tarokh, N. Seshadri, and R. Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Trans. Inform. Theory*, 44:744–765, Mar. 1998.

[65] S. Tavildar and P. Viswanath. Approximately universal codes over slow fading channels. *IEEE Trans. Inform. Theory*, 52:3233–3258, Jul. 2006.

[66] E. Telatar. Capacity of multi-antenna Gaussian channels. *Europ. Trans. on Telecomm.*, 10:585–595, Nov./Dec. 1999.

[67] E. Telatar and D. Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Trans. Inform. Theory*, 46:1384–1400, Jul. 2000.

[68] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication.* Cambridge University Press, 2005.

[69] A. Tulino, A. Lozano, and S. Verdú. Impact of antenna correlation on the capacity of multiantenna channels. *IEEE Trans. Inform. Theory*, 51:2491–2509, Jul. 2005.

[70] A. Tulino and S. Verdú. *Random Matrix Theory and Wireless Communications.* Now Publishers, Hanover, MA, 2004.

[71] R. Vaze and B. Sundar Rajan. On space-time trellis codes achieving optimal diversity-multiplexing tradeoff. *IEEE Trans. Inform. Theory*, 52:5060–5067, Nov. 2006.

[72] S. Verdú. Recent results on the capacity of wideband channels in the low-power regime. *IEEE Wireless Communications Magazine*, 48:40–45, Aug. 2002.

[73] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory*, 48:1319–1343, Jun. 2002.

[74] S. Wei, D. Goeckel, and M. Valenti. Asynchronous cooperative diversity. *IEEE Trans. Wireless Commun.*, 5:1547–1557, Jun. 2006.

[75] X. Wu and R. Srikant. MIMO channels in the low SNR regime: Communication rate, error exponent and signal peakiness. In *Proc. IEEE Information Theory Workshop: ITW'04*, pages 428–433, San Antonio, Oct. 2004.

[76] L. Xie and P. Kumar. A network information theory for wireless communication: Scaling laws and optimal operation. *IEEE Trans. Inform. Theory*, 50:748–767, May 2004.

[77] S. Yang and J.-C. Belfiore. Diversity-multiplexing tradeoff of double scattering MIMO channels. Submitted to *IEEE Trans. Inform. Theory*, Mar. 2006.

[78] S. Yang and J.-C. Belfiore. Optimal space-time codes for the MIMO amplify-and-forward cooperative channel. *IEEE Trans. Inform. Theory*, 53:647–663, Feb. 2007.

[79] H. Yao and G. Wornell. Achieving the full MIMO diversity-multiplexing frontier with rotation-based space-time codes. In *Proc. Allerton Conference on Communication, Control and Computing*, Oct. 2003.

[80] M. Yuksel and E. Erkip. Diversity-multiplexing tradeoff in cooperative wireless systems. In *Proc. 40th Conference on Information Sciences and Systems: CISS*, Princeton, Mar. 2006.

[81] L. Zheng and D. Tse. Communication on the Grassman manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inform. Theory*, 48:359–383, Feb. 2002.

[82] L. Zheng and D. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inform. Theory*, 49:1073–1096, May 2003.

[83] L. Zheng and D. Tse. Diversity-multiplexing tradeoff in multiple-access channels. *IEEE Trans. Inform. Theory*, 50:1859–1874, Sep. 2004.

[84] D. Zwillinger. *Handbook of Integration.* Jones and Bartlett, Boston, 1992.