Information Theoretic Analysis of Multiple-Antenna Transmission Diversity

by

Aradhana Narula

S.M., Massachusetts Institute of Technology (1992)
B.S.E., University of Pennsylvania (1990)

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

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1997

© Massachusetts Institute of Technology, MCMXCVII. All rights reserved.

Author

Department of Electrical Engineering and Computer Science
May 23, 1997

Certified by

Mitchell D. Trott
Professor of Electrical Engineering
Thesis Supervisor

Accepted by

Arthur C. Smith
Chairman, Departmental Committee on Graduate Students
Information Theoretic Analysis of Multiple-Antenna Transmission Diversity
by
Aradhana Narula

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

A framework is developed for the design and evaluation of multiple-antenna transmission diversity methods for slowly varying channels. The channel is characterized by a vector $\alpha$ of gains and phases, perfectly known at the receiver and unknown at the transmitter. The theoretically optimal system is compared to several simpler schemes suggested in the literature. The comparison is based on mutual information as a function of antenna element gains. Using a stochastic model for the antenna coefficients, we show that only the optimal diversity method increases expected mutual information relative to a single antenna system. All schemes do, however, significantly reduce outage probability, which is the potentially more useful performance measure when considering channels with decoding delay constraints. In the limit of an infinite number of antennas, we show that all diversity methods allow communication at the rate of expected mutual information with zero outage probability.

When the channel is known at both transmitter and receiver, transmitter diversity using ideal beamforming achieves performance equal to receiver diversity using ideal beamforming. Beamforming using $M$ antennas results in a factor of $M$ improvement in received signal-to-noise ratio (SNR) over the best transmit diversity scheme when the channel is unknown to the transmitter. We examine the case where the channel coefficients are perfectly known at the receiver and partially known at the transmitter. By varying the amount and kind of side information at the transmitter, we evaluate the gap between perfect channel knowledge in terms of expected SNR and expected mutual information. We show that even a small amount of side information is quite valuable.

We show that the capacity of the majority of the diversity techniques is unaffected by time-variation of the channel. One of the diversity schemes, however, converts the vector input channel into a scalar inter-symbol interference (ISI) channel. We show that time variation decreases the capacity of the ISI channel.

Thesis Supervisor: Mitchell D. Trott
Title: Professor of Electrical Engineering
To Mom, Dad, and Anupam
Acknowledgments

I am sincerely grateful to my thesis advisor Professor Mitchell Trott for his continual guidance, patience, and encouragement. His creativity, analytical ability, and dedication to his students are inspirational. By sharing his insights and encouraging mine, he has helped make my doctoral work at MIT an extremely rewarding and enjoyable experience. His contributions to this thesis and to my professional development are innumerable.

I also wish to thank my thesis committee members, Professor Robert Gallager and Professor Gregory Wornell, for their technical expertise and insights. I would also like to thank Dr. Don Boroson at Lincoln Laboratory for his keen interest in my research. He was always excited to hear about my work and provided many enlightening observations during our discussions.

I would like to thank Professor Amos Lapidoth, Dr. Vivek Borkar, and Professor Sanjoy Mitter for helpful discussions. Dr. Vivek Borkar also provided detailed assistance with Appendix A. I would like to thank the faculty at LIDS for their dedication to research and the educational experience.

I am very thankful to MIT Lincoln Laboratory for supporting this research. I would like to thank Dr. Dave McElroy for taking me on as a research assistant. I would also like to thank my group leader, Dr. Bill Greenberg, and all the members of Group 64 for their kindness and support.

I am also thankful to many friends and officemates over the years for helping make graduate student life such an enjoyable experience. I would especially like to thank Cynara Wu, Rachel Learned, Susie Wee, Kevin Yu, Lon Sunshine, Debra Martin, and Cindy LeBlanc for their friendship and encouragement. I am also very thankful to Kimo Tam for his love and patience and for showing me that research is almost as fun as ice hockey.

Finally, I would like to thank my parents and my brother for their unending faith, love, support, and wisdom.
1 Introduction

1.1 Time-varying multipath channels .................................................. 18
  1.1.1 Channel impulse response ...................................................... 18
  1.1.2 Doppler spread ..................................................................... 19
  1.1.3 Statistical channel characterization ......................................... 20
  1.1.4 Channel model ..................................................................... 22

1.2 Overview of diversity systems .......................................................... 24
  1.2.1 Frequency diversity ............................................................... 25
  1.2.2 Time diversity ..................................................................... 25
  1.2.3 Receive antenna diversity ..................................................... 26
  1.2.4 Transmit antenna diversity ................................................... 27

1.3 Motivation for information-theoretic analysis .................................... 27

1.4 Outline ......................................................................................... 28

1.5 Summary of results ....................................................................... 29

2 Transmit diversity

2.1 Introduction ................................................................................... 33

2.2 Channel model ............................................................................ 34

2.3 Diversity methods and mutual information .................................... 35
  2.3.1 “Unconstrained” signaling ................................................... 37
2.3.2 Time division and frequency division ........................................ 38
2.3.3 Time shifting and frequency shifting ........................................ 40
2.3.4 Randomly time weighted transmit signals .................................. 42
2.4 Outage regions ................................................................. 45
2.5 Probabilistic channel model ..................................................... 47
  2.5.1 Expected mutual information ............................................... 48
  2.5.2 Outage probabilities ......................................................... 50
  2.5.3 Communication at rates up to expected mutual information ........ 52
2.6 Discussion ........................................................................... 53
  2.6.1 Comparison to ideal beamforming ......................................... 53
  2.6.2 Channel coding strategies ..................................................... 55
  2.6.3 Adding diversity to existing systems ....................................... 57
  2.6.4 Diversity vs. directive arrays ............................................... 58

3 Transmit diversity over continuous-time channels .......................... 61
  3.1 Channel model ....................................................................... 62
  3.2 Mutual information calculations .............................................. 63
    3.2.1 “Unconstrained” signaling ................................................ 64
    3.2.2 Time division .................................................................. 65
    3.2.3 Frequency division .......................................................... 66
    3.2.4 Time shifting .................................................................. 66
    3.2.5 Random weighting ............................................................ 67
    3.2.6 Frequency shifting ............................................................ 69

4 Transmit diversity with channel side information ......................... 73
4.1 Receive diversity versus transmit diversity with ideal beamforming ......... 74
4.2 Optimal Gaussian inputs for transmit diversity ............................. 75
4.3 Performance measure: average SNR ........................................... 78
  4.3.1 Side information: Random vector $S$ correlated with $\alpha$ ............. 79
  4.3.2 Side information: $N$ bits of channel information ....................... 80
    4.3.2.1 Results for $M = 2$ transmit antennas .............................. 82
    4.3.2.2 Results for $M > 2$ transmit antennas .............................. 86
  4.3.3 Quantization and rate distortion theory ................................ 88
4.4 Performance measure: average mutual information .......................... 89
  4.4.1 Limiting cases: completely known ($S = \alpha$) and completely unknown
      ($S = \emptyset$) channel ....................................................... 90
  4.4.2 Side information: Random vector $S$ correlated with $\alpha$ ............. 91
  4.4.3 Side information: $N$ bits of channel information ....................... 96

5 Transmit diversity for time-varying channels ................................ 99
  5.1 Channel model ........................................................................ 99
  5.2 Effects of time variation on transmitter diversity ......................... 100
  5.3 Time-varying ISI ................................................................. 102
      5.3.1 Example: Why random time-varying ISI is worse than time-invariant
           ISI ........................................................................... 104
      5.3.2 Effect of time-varying ISI as $M$ increases to three antenna elements 105
      5.3.3 Upper bound for capacity when $\lambda = 0$ .............................. 106
      5.3.4 Mutual information computation: $\lambda = 0$ .......................... 110
          5.3.4.1 Stationary distribution calculation ................................. 110
          5.3.4.2 Ergodicity of the stationary distribution .......................... 112
          5.3.4.3 Mutual information .................................................. 114

- 11 -
Contents

5.3.4.4 Effect of time variation on mutual information 117

6 Conclusion 119

6.1 Conclusions 119

6.2 Future research directions 120

A Time shifting diversity as $M \to \infty$ 123

B Random time weighting as $M \to \infty$ 127

C Justification of independent circular Gaussian codebooks for each antenna input 129

C.1 Minimizing probability of outage 130

C.2 Game theoretic approach 131
List of Figures

1.1 Receiver moving toward fixed transmitter. .......................... 19

2.1 Diversity channel with $M$ transmit antennas. ...................... 34

2.2 Six transmitter diversity schemes, illustrated for $M = 2$ antennas. .... 36

2.3 Values of $|\alpha_1|$ and $|\alpha_2|$ that correspond to a mutual information of 1 nat/complex symbol. ................................................. 46

2.4 Rate in bits per second vs. SNR at outage probability of $\epsilon = 0.01$ with $M = 2$, for time shifting, time division, random weighting, and unconstrained signaling. 52

3.1 Continuous-time transmit diversity schemes illustrated for $M = 2$ antennas. .......................... 63

3.2 Frequency shifting diversity scheme using new continuous-time channel model. 69

4.1 Channel model for $M$-antenna receive diversity. ..................... 74

4.2 Gap between perfect side information and $N$ bits of side information using randomly generated initial partition regions. ......................... 85

4.3 Quantization regions and codebooks resulting from the Lcyd algorithm using a random initial partition. The jagged region boundaries are an artifact of the approximations resulting from the discretization of the $\theta$-$\phi$ sample space. 86

4.4 Sufficient condition for ensuring that the input covariance matrix that maximizes average mutual information has rank 1: when the side information describes each complex channel coefficient to within a Gaussian error with variance $\sigma_x^2$, it is sufficient to require that $\sigma_x^2 \leq \tilde{\sigma}_x^2$. Above we plot $\tilde{\sigma}_x^2$ as a function of $\nu = E||\alpha||^2F_s/N_0$. ................................. 96

5.1 Mutual information as a function of the time correlation $\lambda$ between successive channel coefficients at SNR = $\sigma_x^2 F_s N_0$ of 20 dB. ............................. 103
List of Figures

5.2 Two state Markov chain used to demonstrate the effects of time variation on an ISI channel ........................................ 104

5.3 Difference between the mutual information of a three tap time varying ISI channel and a two tap time varying ISI channel. .................................. 106

5.4 The expected mutual information of the time-invariant ISI channel \( \lambda = 1 \) is compared to the derived upperbound for the capacity of the time-varying ISI channel \( \lambda = 0 \). .................................................. 109

5.5 The capacity of the time-varying ISI channel is compared to the derived upper bound \( (5.35) \). .................................................. 109

5.6 Comparison of capacity for a rapidly time-varying ISI channel, \( C_{TV\text{ ISI}} \) and expected mutual information of the non-time varying ISI channel as a function of SNR, \( \sigma_n^2 \frac{L}{N_0} \). .................................................. 118
List of Tables

2.1 Loss $\delta_{M,e}/M$ in dB with respect to AWGN channel for “unconstrained” transmit diversity. ................................................. 51

4.1 Increase in average SNR given $N$ bits of side information using scalar quantizer based initial partition regions for the Lloyd algorithm. ............... 84

4.2 Increase in average SNR given $N$ bits of side information using randomly generated initial partition regions for the Lloyd algorithm. ............... 85
Chapter 1

Introduction

Wireless mobile communication channels are plagued not only by additive noise, but also by time-varying multipath. The transmitted signal may be reflected, diffracted or scattered by several objects; thus the received signal consists of the sum of several delayed and attenuated copies of the transmit signal. Since the users are presumed to be moving, the amplitudes and delays of the multiple paths change over time. If narrowband signals are transmitted, at the receiver, the multiple copies of the narrowband signal may combine constructively or destructively depending on the path delays. The variation of the resulting received signal’s amplitude due to the time variation of the individual paths’ attenuations and delays is referred to as signal fading.

One way to combat the effects of fading channels is through the use of diversity systems. Diversity combining techniques have been known and utilized for 70 years. The basic idea behind diversity is that errors occur when the channel attenuation is large, i.e., when the channel is in a deep fade. Diversity techniques are used to supply the receiver with several independently faded copies of the message so that the probability that all copies are poor is small. There are many different ways to achieve diversity including frequency diversity, time diversity, and space diversity. In mobile radio communications, space diversity is most widely used since diversity paths can be achieved without an increase in bandwidth or transmission time requirements. Space diversity includes systems with multiple receive antennas, referred to as receiver diversity, and systems with multiple transmit antennas, referred to as transmitter diversity. In a mobile radio system, it is most cost effective to employ multiple antennas at the base station and single or double antennas on the mobile units. Thus, in transmitting from the mobile to the base station, diversity is achieved through multiple receive antennas and in transmitting from the base station to the mobiles, diversity is achieved through multiple transmit antennas. Optimal and suboptimal schemes for exploiting receiver diversity are well-studied in the literature [14]. The study of transmit diversity is far less advanced. Thus our focus is on characterizing the benefits of transmit


Introduction

To understand the applicability and benefits of diversity systems to fading channels, it is important to fully understand the type of channel with which we are dealing. Therefore we first describe and characterize several important aspects of time-varying multipath channels. This introduction to fading channels includes a derivation of the commonly used Rayleigh fading channel model. We then present a brief description of several diversity systems. This is followed by an outline of the thesis and a summary of our results.

1.1 Time-varying multipath channels

There are many references which describe and analyze time-varying channels, including Kennedy [17], Gallager [7, 8], Proakis [26], Medard [19], and Goldsmith [12]. The short summary below is primarily based on analysis presented in Proakis [26], Gallager [9,10] and Medard [19].

1.1.1 Channel impulse response

We begin by considering the effects of transmitting a complex bandpass signal,

$$\tilde{x}(t) = x(t)e^{j2\pi f_c t},$$  \hspace{1cm} (1.1)

over a time varying multipath channel. Denote by $a_n(t)$ and by $\tau_n(t)$ the strength and delay, respectively, of the $n$th path from the transmit antenna to the receive antenna. The received bandpass signal is

$$\tilde{y}(t) = \sum_{n} a_n(t)\tilde{x}(t - \tau_n(t)) + \tilde{v}(t),$$  \hspace{1cm} (1.2)

where $\tilde{v}(t)$ is additive white Gaussian noise (AWGN) over the band of interest. Substituting Equation (1.1) for $\tilde{x}(t)$ yields

$$\tilde{y}(t) = \left[ \sum_{n} a_n(t)e^{-j2\pi f_c \tau_n(t)}x(t - \tau_n(t)) \right] e^{j2\pi f_c t} + \tilde{v}(t).$$  \hspace{1cm} (1.3)
1.1 Time-varying multipath channels

From the equation above, it is apparent that the equivalent baseband received signal is

\[ y(t) = \sum_{n} a_n(t)e^{-j2\pi f_c\tau_n(t)}x(t - \tau_n(t)) + v(t) \]  

(1.4)

and the corresponding baseband time-varying impulse response is

\[ h(\tau, t) = \sum_{n} a_n(t)e^{-j2\pi f_c\tau_n(t)}\delta(\tau - \tau_n(t)). \]  

(1.5)

The impulse response, \( h(\tau, t) \), is the response of the channel at time \( t \) to an impulse applied at time \( t - \tau \). The length of this channel response, equivalently the difference between the largest and smallest propagation delay containing significant energy, is referred to as the multipath spread of the channel and is denoted by \( T_m \). Typical values for multipath spread range from a few hundred nanoseconds indoors to 30\( \mu \)sec in urban environments.

1.1.2 Doppler spread

The time variation of the channel leads to a spreading of the signal bandwidth, or Doppler spread. Consider the simple situation shown in Figure 1.1 where the initial distance between transmitter and receiver is \( d \) and the receiver is moving toward the fixed transmitter at speed \( v_s \).

![Figure 1.1: Receiver moving toward fixed transmitter.](image)

As the receiver moves, the distance the signal must travel decreases with velocity \( v_s \). The transmitter sends \( f_c t \) cycles of the carrier signal from time 0 to \( t \). The first cycle arrives at time \( \frac{d}{c} \), where \( c \) is the speed of light. At time \( t \), the distance to the receiver is \( d - v_s t \). The last cycle arrives at \( t + \frac{d - v_s t}{c} \). Therefore, \( f_c t \) cycles arrive at the receiver in time \( t - \frac{v_s t}{c} \). The
Introduction

frequency seen at the receiver is

\[ f'_c = \frac{f_c t}{t - \frac{v_s t}{c}} = \frac{f_c}{1 - \frac{v_s}{c}}. \]

Assuming that \( v_s \ll c \), we can approximate \( f'_c \) by \( f_c \left(1 + \frac{v_s}{c}\right) \). Therefore the Doppler shift for this scenario is \( B_s = \frac{f_c v_s}{c} \). For a carrier frequency of 1 GHz and a vehicle traveling at 100 km/hour, this corresponds to a Doppler shift of 92.6 Hz. One can of course imagine more complex scenarios in which the transmitted signal bounces off several moving objects before arriving at the receiver. This may lead to even larger Doppler shifts, although typically the Doppler shift is below 100 Hz. In general, the Doppler shift is dependent upon the relative velocity between the transmitter, receiver, and reflecting objects. Therefore different reflective paths give rise to different Doppler shifts. The difference between the largest and smallest Doppler shift is the Doppler spread, \( B_D \), of the channel. With Doppler spread, the bandwidth of the received signal may be larger than the bandwidth of the transmit signal.

Above we described the predominant source of Doppler spread: the time-varying phases of the superimposed components of the received signal. Similar Doppler effects may also be caused by changes in received signal amplitude due to the interposition of RF-opaque obstacles. In most wireless channels, however, the signal phase changes much more rapidly than the amplitude.

1.1.3 Statistical channel characterization

If the number of multipath components in the received signal is small, a ray tracing model can be used to approximate the received signal. Use of this model requires detailed information about the number, location, and characteristics of the reflecting objects. Typically these channel properties are unknown, and hence the channel variations are unpredictable to the user and a statistical model must be used. A detailed description of ray tracing models may be found in Goldsmith [12]. Here, we present only the statistical channel description. We begin by assuming \( h(\tau, t) \) may be modeled as a sample function of a stationary ergodic random process in the \( t \) variable [7] [26]. The stationarity assumption is reasonable over the time interval of interest here.

To describe characteristics of the time-varying multipath channel, we define several
useful correlation and power spectral density functions. These correlation functions lead to definitions of the coherence time and coherence bandwidth of the channel. The development of these definitions is presented in greater detail in Proakis [26], Gallager [7], and Kennedy [17]. The autocorrelation function of the channel impulse response, \( h(\tau, t) \), is defined as

\[
R_{\tau}(\tau_1, \tau_2; \Delta t) = \frac{1}{2} E[h^*(\tau_1, t)h(\tau_2, t + \Delta t)]. \tag{1.6}
\]

Generally, for radio communications, the effect of two different delay paths is uncorrelated, thus

\[
\frac{1}{2} E[h^*(\tau_1, t)h(\tau_2, t + \Delta t)] = R_{\tau}(\tau_1; \Delta t)\delta(\tau_1 - \tau_2). \tag{1.7}
\]

If we assume the difference in observation time, \( \Delta t \), is zero, \( R_{\tau}(\tau; 0) \) gives the average power output as a function of the path delay \( \tau \). The range over which this function is non-zero is the multipath spread of the channel, \( T_m \). In practice \( R_{\tau}(\tau; \Delta t) \) is measured by sending either very narrow pulses or a wideband signal and correlating the received signal with a delayed version of itself.

We can similarly define the autocorrelation function of the Fourier transform of the channel in the \( \tau \) variable:

\[
R_f(f_1, f_2; \Delta t) = \frac{1}{2} E[H^*(f_1, t)H(f_2, t + \Delta t)], \tag{1.8}
\]

where

\[
H(f, t) = \int_{-\infty}^{\infty} h(\tau, t)e^{-j2\pi f \tau}d\tau. \tag{1.9}
\]

Equivalently, \( R_f(f_1, f_2; \Delta t) \) is the Fourier transform of \( R_{\tau}(\tau_1, \tau_2; \Delta t) \). Using the assumption that the response due to different delays is uncorrelated, we observe that \( R_f(f_1, f_2; \Delta t) \) only depends on the frequency difference \( \Delta f = f_2 - f_1 \). This function may be measured by transmitting two sinusoids with frequency difference \( \Delta f \) and cross correlating the two received signals with a relative delay of \( \Delta t \). Letting the difference in observation time be zero, \( R_f(\Delta f; 0) \) describes the frequency coherence of the channel. The coherence bandwidth is defined as the frequency separation beyond which two sinusoids are affected differently by the channel. Since \( R_f(\Delta f; 0) \) is the Fourier transform of \( R_{\tau}(\tau; 0) \), the coherence bandwidth
is proportional to the reciprocal of the multipath spread, $1/T_m$. An equivalent way of
arriving at this conclusion is to realize that since $h(\tau,t)$ is time limited to $(0,T_m)$, its Fourier
transform, $H(f,t)$ may be specified for any given $t$ by samples $1/T_m$ apart in $f$. Therefore,$H(f,t)$ does not vary greatly within a frequency interval $1/T_m$.

The time variation of the channel may be analyzed by looking at either of the auto-
correlation functions, $R_f(\Delta f; \Delta t)$ and $R_r(\tau; \Delta t)$, as a function of $\Delta t$. As mentioned above,
time variations cause a Doppler spreading or shift of the received signal bandwidth. A statistical
description of the time-variation leads to a statistical description of the Doppler spread.
To examine the effects of time variation in the frequency domain, we can take the Fourier
transform of either autocorrelation function with respect to the variable $\Delta t$, for example

$$\sigma(\Delta f; \lambda) = \int_{-\infty}^{\infty} R_f(\Delta f; \Delta t)e^{-j2\pi\lambda\Delta t} d\Delta t. \quad (1.10)$$

Setting $\Delta f$ to zero yields the following Fourier transform relationship:

$$\sigma(0; \lambda) = \int_{-\infty}^{\infty} R_f(0; \Delta t)e^{-j2\pi\lambda\Delta t} d\Delta t. \quad (1.11)$$

The autocorrelation function $R_f(0; \Delta t)$ as a function of $\Delta t$ provides the coherence time ($\Delta t_c$)
of the channel. The range of values of $\lambda$ over which $\sigma(0; \lambda)$ is non-zero provides a measure
of the Doppler spread, $B_D$, of the channel. From the Fourier transform relationship, it is
observed that the coherence time can be approximated by the reciprocal of $B_D$. In other
words, a slowly changing channel has a large coherence time and a small Doppler spread.

### 1.1.4 Channel model

The choice of an appropriate channel model depends upon the transmitted signal’s character-
istics. The effect of the channel upon each symbol depends on the symbol’s duration and
bandwidth as well as upon the characteristics of the channel.

If the bandwidth, $W$, of the transmitted signal, $x(t)$, is larger than the coherence
bandwidth, different frequencies of the signal will be affected dissimilarly. In other words,
the channel is frequency selective. If the symbol duration is much larger than the multipath
spread, $T \gg T_m$, the inter-symbol interference introduced by the channel is negligible. If the
1.1 Time-varying multipath channels

bandwidth of the signal is $W \approx \frac{1}{T}$ and $T \gg T_m$, then the channel is frequency non-selective, $W \ll \frac{1}{T_m} \approx (\Delta f)_c$. In this case all frequency components of $X(f)$ experience the same gain and phase shift as they go through the channel. To see this more concretely, write the baseband output of the channel as

$$y(t) = \int_{-\infty}^{\infty} x(t - \tau) h(\tau, t) d\tau + v(t),$$

or equivalently as

$$y(t) = \int_{-\infty}^{\infty} X(f) H(f, t) e^{j2\pi ft} df + v(t),$$

where

$$H(f, t) = \int_{-\infty}^{\infty} h(\tau, t) e^{-j2\pi ft} d\tau$$

$$= \int_{-\infty}^{\infty} \sum_n a_n(t) e^{j2\pi f \tau_n(t)} e^{-j2\pi ft} \delta(\tau - \tau_n(t)) d\tau$$

$$= \sum_n a_n(t) e^{j2\pi f \tau_n(t)} \int_{-\infty}^{\infty} e^{-j2\pi f \tau} \delta(\tau - \tau_n(t)) d\tau$$

$$= \sum_n a_n(t) e^{j2\pi (f - \tau_n(t))}.$$  

We assume that the phase of each multipath component is independently and uniformly distributed between 0 and $2\pi$. Using this assumption, $h(\tau, t)$ and $H(f, t)$, the sum of a large number of paths, have zero mean. If the frequency content of $X(f)$ is concentrated around zero and if $H(f, t)$ is a constant over the frequency band of interest, i.e., the channel is frequency non-selective, then $H(f, t) = H(0, t)$. Furthermore, $H(0, t)$ is the sum of a large number of independent random variables and, using Central Limit Theorem arguments, may be modeled as zero mean complex Gaussian:

$$H(0, t) = a(t) e^{-j\theta(t)},$$

where $a(t)$ is Rayleigh distributed and $\theta(t)$ is uniformly distributed over the interval $(-\pi, \pi)$. 

\begin{align*}
1.1 & \text{ Time-varying multipath channels} \\
\text{bandwidth of the signal is } W \approx \frac{1}{T} \text{ and } T \gg T_m, \text{ then the channel is frequency non-selective, } W \ll \frac{1}{T_m} \approx (\Delta f)_c. \text{ In this case all frequency components of } X(f) \text{ experience the same gain and phase shift as they go through the channel. To see this more concretely, write the baseband output of the channel as} \\
y(t) &= \int_{-\infty}^{\infty} x(t - \tau) h(\tau, t) d\tau + v(t), \\
\text{or equivalently as} \\
y(t) &= \int_{-\infty}^{\infty} X(f) H(f, t) e^{j2\pi ft} df + v(t),
\end{align*}

where

\begin{align*}
H(f, t) &= \int_{-\infty}^{\infty} h(\tau, t) e^{-j2\pi ft} d\tau \\
&= \int_{-\infty}^{\infty} \sum_n a_n(t) e^{j2\pi f \tau_n(t)} e^{-j2\pi ft} \delta(\tau - \tau_n(t)) d\tau \\
&= \sum_n a_n(t) e^{j2\pi f \tau_n(t)} \int_{-\infty}^{\infty} e^{-j2\pi f \tau} \delta(\tau - \tau_n(t)) d\tau \\
&= \sum_n a_n(t) e^{j2\pi (f - \tau_n(t))}.
\end{align*}

We assume that the phase of each multipath component is independently and uniformly distributed between 0 and $2\pi$. Using this assumption, $h(\tau, t)$ and $H(f, t)$, the sum of a large number of paths, have zero mean. If the frequency content of $X(f)$ is concentrated around zero and if $H(f, t)$ is a constant over the frequency band of interest, i.e., the channel is frequency non-selective, then $H(f, t) = H(0, t)$. Furthermore, $H(0, t)$ is the sum of a large number of independent random variables and, using Central Limit Theorem arguments, may be modeled as zero mean complex Gaussian:

$$H(0, t) = a(t) e^{-j\theta(t)},$$

where $a(t)$ is Rayleigh distributed and $\theta(t)$ is uniformly distributed over the interval $(-\pi, \pi)$. 

\begin{align*}
1.1 & \text{ Time-varying multipath channels} \\
\text{bandwidth of the signal is } W \approx \frac{1}{T} \text{ and } T \gg T_m, \text{ then the channel is frequency non-selective, } W \ll \frac{1}{T_m} \approx (\Delta f)_c. \text{ In this case all frequency components of } X(f) \text{ experience the same gain and phase shift as they go through the channel. To see this more concretely, write the baseband output of the channel as} \\
y(t) &= \int_{-\infty}^{\infty} x(t - \tau) h(\tau, t) d\tau + v(t), \\
\text{or equivalently as} \\
y(t) &= \int_{-\infty}^{\infty} X(f) H(f, t) e^{j2\pi ft} df + v(t),
\end{align*}

where

\begin{align*}
H(f, t) &= \int_{-\infty}^{\infty} h(\tau, t) e^{-j2\pi ft} d\tau \\
&= \int_{-\infty}^{\infty} \sum_n a_n(t) e^{j2\pi f \tau_n(t)} e^{-j2\pi ft} \delta(\tau - \tau_n(t)) d\tau \\
&= \sum_n a_n(t) e^{j2\pi f \tau_n(t)} \int_{-\infty}^{\infty} e^{-j2\pi f \tau} \delta(\tau - \tau_n(t)) d\tau \\
&= \sum_n a_n(t) e^{j2\pi (f - \tau_n(t))}.
\end{align*}

We assume that the phase of each multipath component is independently and uniformly distributed between 0 and $2\pi$. Using this assumption, $h(\tau, t)$ and $H(f, t)$, the sum of a large number of paths, have zero mean. If the frequency content of $X(f)$ is concentrated around zero and if $H(f, t)$ is a constant over the frequency band of interest, i.e., the channel is frequency non-selective, then $H(f, t) = H(0, t)$. Furthermore, $H(0, t)$ is the sum of a large number of independent random variables and, using Central Limit Theorem arguments, may be modeled as zero mean complex Gaussian:

$$H(0, t) = a(t) e^{-j\theta(t)},$$

where $a(t)$ is Rayleigh distributed and $\theta(t)$ is uniformly distributed over the interval $(-\pi, \pi)$. 

for any value of \( t \). The output of the channel reduces to

\[
y(t) = H(0, t) \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df + v(t) \tag{1.19}
\]

\[
= H(0, t)x(t) + v(t). \tag{1.20}
\]

Thus, when the signal bandwidth is less than the coherence bandwidth, the effect of the channel is a multiplication by a complex Gaussian random process \( H(0, t) \). This is sometimes referred to as narrowband signaling and the time-varying channel model is the well known Rayleigh fading model.

Recall that the time variation of the channel is characterized by the coherence time or Doppler spread. If it is possible to choose both the signal bandwidth to be less than the coherence bandwidth, \( W \ll (\Delta f)_c \), and the symbol duration to be smaller than the coherence time, \( T \ll (\Delta t)_c \), the input signal will be multiplied by a fixed complex Gaussian. A channel which satisfies these conditions is said to be slowly time-varying or slowly fading. Assuming \( W \approx \frac{1}{T} \), the spread factor of this frequency non-selective channel, \( T_m B_d \), is less than 1. When \( T_m B_d \) is small, nothing is varying too rapidly and the channel may be measured. Most wireless communication channels fall into this category. For our transmitter diversity analyses, we start with an extremely slowly varying (essentially time-invariant) frequency non-selective channel model that is assumed to be measured perfectly at the receiver. We later analyze more rapidly varying channels.

### 1.2 Overview of diversity systems

Over the years, many types of diversity systems have been developed and implemented. The various systems differ in complexity, cost, efficiency, and suitability for different types of channels. In this section we present a brief description of several diversity systems; greater detail may be found in Proakis [26] and Jakes [14].
1.2 Overview of diversity systems

1.2.1 Frequency diversity

As mentioned above, the fading characteristics of signal frequencies separated by a distance larger than the coherence bandwidth, $(\Delta f)_c$, are independent. In frequency diversity, $M$ copies of a message are sent on $M$ carrier frequencies sufficiently separated to be independent. The $M$ received copies are thus independently faded. Frequency diversity is a special case of broadband signaling. Suppose a bandwidth $W$ much larger than the coherence bandwidth $(\Delta f)_c$ is available to the user and a signal covering the full band bandwidth $W$ is transmitted. When $W \gg (\Delta f)_c$, the multipath components in the received signal are resolvable with time resolution $\frac{1}{W}$, and thus $T_m W \approx \frac{W}{(\Delta f)_c}$ copies of the message signal are received. With a suitable choice of signal set, the received signals may be combined optimally using a RAKE receiver [25]. Unfortunately, frequency diversity/broadband signaling requires a bandwidth expansion factor of $M$ or equivalently $\frac{W}{(\Delta f)_c}$. Furthermore, if the coherence bandwidth is large, frequency diversity schemes will be difficult to implement. Mobile radio channels generally have a coherence bandwidth of about 1–2MHz [14] while the coherence bandwidth of personal communication systems is much larger.

1.2.2 Time diversity

The $M$ copies of the message may alternatively be separated in time. The separation between the multiple message copies must be larger than the coherence time of the channel, $(\Delta t)_c$. This becomes difficult when the channel is too slowly varying. On the other hand, if the channel is varying too rapidly, channel measurement and signaling become difficult since the channel varies within very short intervals. For a vehicle traveling at 100 km/hour and a carrier frequency of 1–10 GHz, the coherence time ranges from about 10 msec to 1 msec. With time diversity, it takes $M$ times as long to transmit the same amount of information.

A fading channel may be modeled as a channel with burst noise. A multitude of errors occur when the channel is in a deep fade. Time and frequency division are analogous to using a repetition code with block interleaving. Repetition codes however are generally wasteful of bandwidth compared to more complex methods of coding.
1.2.3 Receive antenna diversity

Space diversity is often used in mobile radio communications since it requires no additional bandwidth or time requirements. In receive diversity, multiple copies of the transmit signal are received on multiple omni-directional receive antennas. There are a variety of techniques that may be used to combine the multiple received signals. These techniques vary in complexity, effectiveness, and in channel measurement requirements.

The receive antennas must be sufficiently separated in space so that the received signals have independent fading characteristics. One method for approximating the required antenna spacing is to assume that the mobile is moving in a straight line and that there are a large number of incoming paths with uniformly distributed angles of arrival. Under this assumption, there is a one-to-one correspondence between the temporal and spatial autocorrelations. In this case, the received signal as a function of spatial location (or time) may be modeled as a stationary complex Gaussian random process. The envelope of this stationary process \( r(x) \) has a Rayleigh distribution:

\[
    f_r(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}.  \tag{1.21}
\]

The spatial autocovariance of the signal envelope is approximated by [14]:

\[
    R(\zeta) = E[(r(x) - E[r])(r(x + \zeta) - E[r])]
    \approx \frac{\pi}{8} \frac{\sigma^2 J_0^2 \left( \frac{2\pi \zeta}{\lambda} \right)}{J_0^2 \left( \frac{2\pi}{\lambda} \right)}, \tag{1.23}
\]

where \( J_0 \) denotes the 0th order Bessel function and \( \lambda \) is the carrier wavelength. The first null of the autocorrelation occurs near \( \lambda/2 \). Thus, for signals at the multiple antennas to be uncorrelated, the antennas are generally spaced at least \( \lambda/2 \) apart. In practice, vehicle motion may result in a non-uniform distribution of the angles of arrivals. However, the approximation above yields a reasonable estimate for the order of magnitude of required antenna spacing. If the number of paths is small, the antennas must be sufficiently separated so that the pathlengths of the multiple paths of the received signal are separated by \( \lambda/2 \). This may require spacing the antennas several wavelengths apart.

An alternative to using omni-directional receive antennas is to use directive receive antennas that point beams in different directions. If the fading characteristics from the
1.3 Motivation for information-theoretic analysis

individual beams are uncorrelated, the received signals may be used in a diversity system.

1.2.4 Transmit antenna diversity

Transmit diversity, as the name suggests, involves the use of multiple transmit antennas and a single receive antenna. Since the multiple transmit signals are already combined when they reach the receiver, simply sending multiple copies of the same message is not useful. In this research we analyze several techniques for exploiting transmit diversity and determine how to optimally use multiple transmit antennas. We also compare transmit diversity to the more widely studied receive diversity case.

1.3 Motivation for information-theoretic analysis

This research was motivated by transmit diversity schemes suggested in the literature. Multiple transmit antennas were probably first used to send multiple copies of a signal over orthogonal time or frequency slices. This of course incurs a bandwidth expansion factor equal to the number of antennas. A transmit diversity technique without bandwidth expansion was first suggested by Wittenben [34] in 1991. Wittenben's diversity technique of sending time-delayed copies of a common input signal over the multiple antennas was also independently discovered by Seshadri and Winters [28] and by Weerackody [32] in 1993. In 1992, a diversity scheme in which frequency modulated versions of a common input signal are transmitted over the multiple antennas was devised by Hiroike, Adachi, and Nakajama [13].

The performance of these schemes was generally assessed by choosing a particular modulation scheme and determining the benefits of multiple transmit antennas in terms of signal-to-noise ratio or probability of error. In an effort to understand the similarities and differences between the diversity methods and to determine the fundamental limitations of these diversity methods, we analyze transmit diversity from an information-theoretic viewpoint. Our information theoretic analysis is primarily based on the ideas and considerations presented by Ozarow, Shamai, and Wyner [22] for evaluating the performance of time division multiple access protocols. Ozarow et al. [22] first suggested using probability of outage as the performance measure for channels with decoding delay constraints.


Introduction

This information-theoretic approach to analyzing transmit diversity allows us to compare the suggested diversity schemes under a common framework independent of the choice of modulation scheme or error correction code. Furthermore, we are able to determine the optimal method for using multiple transmit antennas to maximize the rate of reliable communication. In independent and concurrent work, Foschini [6] also presents some information theoretic analysis of the multiple transmit antenna channel as a special case of multiple transmit and multiple receive antenna systems.

1.4 Outline

Chapter 2

In this chapter we introduce a framework for the comparison of several transmit diversity schemes proposed in the literature. A discrete-time frequency non-selective time-invariant channel model is assumed. The transmit signal from each antenna is multiplied by a complex channel coefficient and corrupted by additive white Gaussian noise (AWGN) as it travels through the channel. The channel coefficients are assumed to be perfectly known at the receiver and unknown at the transmitter. We first compare the proposed schemes to an "optimal" scheme without assuming any particular probability distribution for the channel coefficients. We use mutual information as a function of the channel coefficients as our performance measure. Further comparisons of the transmit diversity schemes require assuming a statistical description of the channel coefficients. We assume a Rayleigh fading channel model. This statistical description also allows a comparison of multiple antenna systems to single antenna systems. We consider two performance measures: expected mutual information and probability of outage.

Chapter 3

The transmit diversity schemes may be implemented in continuous-time at passband. In the previous chapter we assumed a discrete-time channel model. In this chapter we partially extend the mutual information results of the previous chapter to a continuous-time channel model. The resulting mutual informations are consistent with the discrete-time channel model calculations.

Chapter 4

We assume, as before, that the channel is known perfectly at the receiver. We now
assume that the transmitter also has some side information describing the channel coefficients. This side information may result from reverse path channel measurements or a dedicated feedback path from the receiver to the transmitter. We show that transmit diversity with complete channel knowledge at the transmitter achieves performance equal to that of receive diversity. We then evaluate the gap between perfect channel knowledge and no channel knowledge at the transmitter.

Chapter 5
In this chapter, we examine the effects of time variation on the transmit diversity schemes. We show that time variation only affects the time-shifting diversity scheme. This scheme implemented on a time-varying channel leads to randomly time-varying inter-symbol interference (ISI). We show that random time-variation of the ISI taps degrades communication over the ISI channel, even when channel measurement is perfect.

1.5 Summary of results

- For a multiple transmit antenna channel with channel coefficients known at the receiver and unknown at the transmitter, we
  - calculated the mutual information of six transmit diversity schemes as a function of the channel coefficients. We determined the “optimal” scheme and defined a new diversity method with properties similar to the schemes suggested in the literature, but with robustness and analysis benefits.
  - showed that the “optimal” scheme, unconstrained signaling, performed uniformly better than the remaining linear diversity methods.
  - showed the size of the gap between unconstrained signaling and the linear diversity methods increased to approximately .833 bits per symbol in the limit of a large number of antennas and high SNR.
- Using a stochastic model for the channel coefficients, we
  - showed that the expected mutual information of the unconstrained signaling scheme is equal to the capacity of an additive white Gaussian noise channel.
Introduction

- showed that the linear diversity schemes do not increase expected mutual information over the single antenna channel.
- showed that all the diversity schemes significantly reduce probability of outage.
- proved that in the limit of an infinite number of antennas, the mutual information of the diversity schemes converges to its expected value.

- Using a continuous-time (rather than discrete-time) channel model, we calculated the mutual information of the transmit diversity methods.

- We showed that if the channel is known to both transmitter and receiver, then ideal beamforming using multiple transmit antennas achieves performance equal to that of ideal beamforming using multiple receive antennas. Furthermore, we showed that ideal beamforming provides a factor of $M$ increase in expected SNR over the best transmit diversity method when the channel is unknown to the transmitter.

- We evaluated the factor of $M$ gap between perfect and zero channel side information at the transmitter. Specifically, we

  - showed that, without loss of generality, the input vector can be restricted to be circular Gaussian when maximizing mutual information.
  - showed that a rank 1 input covariance matrix maximizes average SNR for any given form of channel side information. The optimal covariance matrix “points” in the direction suggested by the side information.
  - showed that if the side information is in the form of random variables $S_i$ correlated with the channel coefficients $\alpha_i$, the expected SNR increases to $M$ quadratically in the correlation coefficient $|\rho|$.
  - determined locally optimal quantization regions and covariance matrices to maximize expected SNR for $M = 2$ antennas and given $N$ bits of side information.
  - showed that for the $M$-antenna case, a scheme using $\log_2 M$ bits of side information increases expected SNR logarithmically in $M$, whereas a scheme using $M - 1$ bits of side information can provide an increase in expected SNR that is linear in $M$.
  - showed that for large $M$ and large $N$, the gap between perfect and zero side information decreases exponentially in the number of bits of side information $N$. 

- 30 -
1.5 Summary of results

- showed that unconstrained signaling, which corresponds to a full rank input covariance matrix, maximizes expected SNR when channel side information is unavailable to the transmitter.

- determined, given side information in the form of a channel measurement with zero mean Gaussian error, conditions on the variance of the error that ensure the input covariance matrix that maximizes expected mutual information is rank 1.

• We showed that time variation adversely affects the mutual information of a channel employing the time shifting diversity scheme. In other words, we showed that random time variation of the ISI coefficients degrades the rate of reliable communication over the ISI channel.
Chapter 2

Transmit diversity

This chapter is a slightly modified version of “Information Theoretic Analysis of Multiple Antenna Transmission” submitted to the IEEE Transactions on Information Theory by Aradhana Narula, Mitchell D. Trott, and Gregory W. Wornell.

2.1 Introduction

Transmitter diversity is generally viewed as more difficult to exploit than receiver diversity, in part because the transmitter is assumed to know less about the channel than the receiver, and in part because of the challenging signal design problem: the transmitter is permitted to generate a different signal at each antenna. Unlike the receiver diversity case, where independently faded copies of a single transmitted signal may be combined optimally to achieve a performance gain, for transmitter diversity the many transmitted signals are already combined when they reach the receiver. How, then, should the transmitted signals be selected to either achieve capacity, or, more practically, to simplify the receiver while maintaining performance near capacity?

A number of simple transmitter diversity methods have been suggested in the literature [13, 14, 28, 32, 34]. We put these methods into a common framework and introduce two improved methods within the same general class. Following the approach of Ozarow, Shamai, and Wyner [22], our measure of performance is the mutual information between input and output over a long block, which corresponds in an approximate sense to the maximum achievable rate of reliable communication. The diversity methods are compared based on outage probabilities, i.e., the probability that mutual information falls below a threshold. A variety of other aspects of the transmit diversity problem are developed in the companion paper [35], which follows a signal processing approach.
The chapter is organized as follows. In Section 2.2 we describe our channel model and assumptions. In Section 2.3 we introduce the diversity methods used in the remainder of the chapter and compute the mutual information achieved by each. Several of the methods are seen to be time-frequency duals. The mutual informations are plotted in the form of outage regions in Section 2.4. To compare the schemes more fully requires a stochastic model for the channel parameters. In Section 2.5 a Rayleigh model is used to compute outage probabilities and expected mutual informations. Not surprisingly, antenna diversity allows a given outage probability to be achieved with substantially lower signal power. Power savings are computed in the limit of high signal-to-noise ratio (SNR) and in the limit of an array with a large number of independent elements. Implementation issues and alternative interpretations are discussed Section 2.6.

2.2 Channel model

![Diagram](image)

Figure 2.1: Diversity channel with $M$ transmit antennas.

We model the $M$-antenna transmitter diversity channel as shown in Figure 2.1. The complex baseband received signal

$$y_k = \sum_{i=1}^{M} \alpha_i x_{i,k} + v_k$$  \hspace{1cm} (2.1)

at time $k$ is the superposition of the $M$ transmitted symbols $x_{1,k}, \ldots, x_{M,k}$, each scaled and phase-shifted by a complex coefficient $\alpha_i$ which represents the aggregate effect of the channel encountered by antenna $i$. The channel is frequency nonselective, i.e., the delay spread of the channel is small compared to the reciprocal of the signal bandwidth. The additive noise $v_k$ is assumed to be white circular Gaussian with variance (for each real and imaginary component) $N_0/2$, and the transmitted energy over a block of $N$ symbols is limited to $\frac{1}{N} \sum_{k=1}^{N} \sum_{i} |x_{i,k}|^2 \leq \mathcal{E}_s$ per symbol. As $M$ increases, the fixed energy $\mathcal{E}_s$ must be distributed
more thinly among the antenna elements; this allows a fair comparison of single and multiple antenna systems.

The channel is assumed to be varying slowly enough that the coefficient vector $\mathbf{\alpha} = [\alpha_1 \alpha_2 \ldots \alpha_M]^T$ is effectively constant over as long a block of symbols as desired. This model is appropriate when the transmitter, receiver, and all reflecting surfaces are either unmoving or slowly moving relative to the carrier wavelength and symbol rate.

The transmitter is assumed to have no knowledge of the coefficients, while the receiver is assumed to have perfect knowledge. The lack of knowledge of $\mathbf{\alpha}$ at the transmitter represents either a lack of feedback from the receiver to the transmitter, or a broadcast scenario where the transmitter must send the same information to many receivers with different locations and hence different $\mathbf{\alpha}$'s. Variations on these assumptions are discussed in Chapters 4 and 6.

Our initial results in Sections 2.3 and 2.4 assume no stochastic model for $\mathbf{\alpha}$. In later sections we adopt a Rayleigh model, where the components of $\mathbf{\alpha}$ are independent, identically distributed, zero-mean, complex Gaussian random variables. The performance of all of our diversity schemes depends only on the distribution of the coefficient magnitudes $|\alpha_1|, \ldots, |\alpha_M|$, not on their relative phases. Furthermore, the performance of some schemes depends only on the distribution of $||\mathbf{\alpha}||$, the magnitude of the coefficient vector. This simplification allows the same analysis methods to be applied to any model for which the tail behavior of $||\mathbf{\alpha}||$ is known.

2.3 Diversity methods and mutual information

There are a variety of ways to exploit multiple antennas at the transmitter. We analyze the six representative schemes shown in Figure 2.2: "unconstrained" signaling, time division, frequency division, the time-shift technique proposed by Wittenben [34] and concurrently by Seshadri and Winters [28,29,33] and by Weerackody [32], a discrete-time version of the frequency-shift technique proposed by Hiroiike et al. [13], and a random weighting technique which makes the channel appear to vary in time. By unconstrained signaling we mean that the system is evaluated as a vector-input scalar-output Gaussian channel. The other five schemes use linear processing to convert the vector-input channel into a scalar-input channel.
As developed in [35], frequency division and time shifting convert antenna diversity into frequency diversity; the memoryless vector-input channel becomes a scalar-input channel with intersymbol interference. Likewise, time division, frequency shifting, and random weighting convert antenna diversity into time diversity; the time invariant vector-input channel becomes a periodically or randomly time-varying scalar-input channel. A procedure analogous to random weighting could be carried out in the frequency domain. Such a channel is not analyzed here, but would have properties similar to random (time) weighting.

To compute the mutual information $I$ achieved by each scheme, we define

$$I = \frac{1}{N} \lim_{N \to \infty} I_N; \quad (2.2)$$
where $I_N$ is the mutual information between a block of $N$ input and output symbols. The existence of this limit is straightforward to establish in all cases we consider. The diversity methods which create time-varying channels are analyzed in the time domain; those which create intersymbol interference channels are analyzed in the frequency domain.

We assume that the input codebooks are derived from complex circular white Gaussian random processes. The use of Gaussian codebooks follows from the assumption that the receiver knows $\alpha$, which reduces the multiple antenna channel to an additive white Gaussian noise channel; the use of codebooks with independent and identically distributed (i.i.d.) components follows from standard arguments. Note that beamforming and waterfilling methods cannot be applied because the transmitter is assumed to have no knowledge of the channel parameters.

All analysis is done using a discrete-time channel model rather than the more physically correct continuous-time, strictly time-limited, approximately bandlimited model adopted in Gallager [8]. The continuous-time model resolves a number of conceptual problems in what follows, such as the use of a finite block length channel code with infinite impulse response bandpass filters. Many of the results shown here are proven for a continuous-time channel in the following chapter. The continuous-time analysis is considerably more cumbersome than the discrete-time counterpart, while the conclusions are effectively the same.

An important aspect of the analysis that follows is the normalization needed to maintain the transmitted energy at $E_s$ per symbol. Unlike a multiple access problem, where the total power increases with the number of users, for multiple transmit antennas the total power is held constant, independent of the number of antennas.

### 2.3.1 "Unconstrained" signaling

The unconstrained multiple transmit antenna system is a memoryless vector-input scalar-output power-limited Gaussian channel. With $M$ transmit antennas, the complex baseband received signal at time $k$ is

$$Y_k = \alpha^T X_k + V_k,$$

(2.3)
where $X_k = [X_{1,k}, \ldots, X_{M,k}]^T$ is the input vector and $V_k$ is complex white Gaussian noise with variance $N_0$.

If the components of $X_k$ are independent zero-mean complex circular Gaussian random variables with variance $E|X_{i,M}|^2 = \mathcal{E}_s/M$ each, the output $Y_k$ is zero-mean Gaussian with variance $||\alpha||^2 \mathcal{E}_s/M + N_0$. The mutual information is then

$$I_{\text{OPT}} = h(Y) - h(Y|X) = \log \left( 1 + \frac{||\alpha||^2 \mathcal{E}_s}{MN_0} \right).$$

Note that the mutual information depends only on the antenna gain $||\alpha||$.

Choosing a different joint distribution on the components of $X$ will increase the mutual information for some values of $\alpha$ and decrease it for others. Selecting the components of $X$ to be i.i.d. is therefore not “optimal” in some global sense. However, we will see that $I_{\text{OPT}}$ is greater than or equal to the mutual information achieved by the other diversity methods considered here for all $\alpha$. Furthermore, in Appendix C we show that this approach is the saddle point solution to a max-min problem in which nature chooses a distribution on $\alpha$ to minimize the rate of reliable communication.

### 2.3.2 Time division and frequency division

Time and frequency division exploit transmit antenna diversity by using linear processing to generate orthogonal signals for each antenna. In the absence of intersymbol interference or Doppler spread, the signals remain orthogonal at the receiver. The multiple antenna channel can then be analyzed as a set of independent parallel channels. As in the case of the multiple-access Gaussian channel, the mutual informations achieved by time division and frequency division are equal. Note that our discrete-time channel model implicitly keeps the bandwidth fixed; as the number of antennas increases each antenna must use a smaller slice of spectrum or time.

The time division approach is shown in Figure 2.2b. The input to the antenna array is formed by time-multiplexing a scalar input sequence $\{X_k\}$ across the antenna elements, so that $X_k$ is transmitted using antenna element $i$ when $k \equiv i \pmod{M}$. For example, when
$M = 2$, odd-time inputs are transmitted on antenna 1 and even-time inputs are transmitted on antenna 2. The output of this channel is then

$$
Y_k = \begin{cases} 
\alpha_1 X_k + V_k & k \equiv 1 \pmod{M}, \\
\alpha_2 X_k + V_k & k \equiv 2 \pmod{M}, \\
\vdots \\
\alpha_M X_k + V_k & k \equiv M \pmod{M}.
\end{cases} 
$$

(2.6)

Let the input sequence $\{X_k\}$ be i.i.d. complex circular Gaussian with energy $E|X_k|^2 = \mathcal{E}_s$ per symbol. The mutual information between input and output is the average of the mutual informations achieved by each antenna:

$$
I_{TD} = \frac{1}{M} \sum_{i=1}^{M} \log \left( 1 + \frac{|\alpha_i|^2 \mathcal{E}_s}{N_0} \right). 
$$

(2.7)

The frequency division approach is shown in Figure 2.2c. The input process $\{X_k\}$ is frequency multiplexed across the antennas by bandpass filtering into $M$ disjoint bands of width $\pi/M$ each. The Fourier transform $H(\omega)$ of the unit-sample response of the resulting scalar-input channel is

$$
H(\omega) = \begin{cases} 
\alpha_1 & 0 \leq |\omega| < \pi/M, \\
\alpha_2 & \pi/M \leq |\omega| < 2\pi/M, \\
\vdots \\
\alpha_M & (M - 1)\pi/M \leq |\omega| < \pi.
\end{cases} 
$$

(2.8)

The mutual information between input and output is

$$
I_{FD} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left( 1 + \frac{\mathcal{E}_s |H(\omega)|^2}{N_0} \right) d\omega 
$$

(2.9)

$$
= \frac{1}{M} \sum_{i=1}^{M} \log \left( 1 + \frac{|\alpha_i|^2 \mathcal{E}_s}{N_0} \right). 
$$

(2.10)

Comparing (2.7) with (2.10) reveals that time division and frequency division yield
the same mutual information. Indeed, any linear method that generates orthogonal signals (that remain orthogonal after passing through the channel) will have the same characteristic behavior as time and frequency division.

Applying Jensen’s inequality to (2.7) or (2.10) shows that the mutual information achieved by time or frequency division is always less than or equal to that achieved by the unconstrained channel (2.5):

\[ I_{TD} = I_{FD} \leq I_{OPT}, \]

with equality if and only if all antenna gains \(|\alpha_1|, |\alpha_2|, \ldots, |\alpha_M|\) are equal.

2.3.3 Time shifting and frequency shifting

Like time division and frequency division, time shifting and frequency shifting are duals. The time shifting diversity method sends delayed versions of a common input signal over the multiple transmit antennas: the \(i\)th antenna carries the input signal \(\{X_k\}\) delayed by \(i - 1\) time steps. The output of the channel is

\[ Y_k = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \alpha_i X_{k-i+1} + V_k. \]  

(2.11)

The Fourier transform of the unit-sample response of the time shift channel is

\[ H(\omega) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} e^{-j(i-1)\omega} \alpha_i. \]  

(2.12)

The frequency shifting diversity method sends modulated versions of a common input signal over the multiple transmit antennas: the \(i\)th antenna carries the signal \(\{e^{j\pi \delta(i-1)k}X_k\}\). The modulation parameter \(\delta > 0\) is arbitrary. The output of the channel is

\[ Y_k = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} e^{j\pi \delta(i-1)k} \alpha_i X_k + V_k. \]  

(2.13)

The magnitude of the frequency response (2.12) of the time shift channel and the
2.3 Diversity methods and mutual information

The magnitude of the time response (2.13) of the frequency shift channel have local maxima corresponding to partial coherent combining and local minima corresponding to partial destructive interference between the antenna elements.

When the input process \{X_k\} is i.i.d. complex circular Gaussian with per symbol variance \(E|X_k|^2 = \mathcal{E}_s\), the mutual information for the time shifting diversity method may be computed from the frequency response (2.12) as

\[
I_{TS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left( 1 + \frac{\mathcal{E}_s}{MN_0} \left| \sum_{i=1}^{M} e^{-j(i-1)\omega} \alpha_i \right|^2 \right) d\omega. \tag{2.14}
\]

The computation for the frequency shifting diversity method is slightly more involved. From (2.13), over a block of \(N\) symbols, the mutual information is

\[
I_N = \sum_{k=1}^{N} \log \left( 1 + \frac{\mathcal{E}_s}{MN_0} \left| \sum_{i=1}^{M} \alpha_i e^{j\pi \delta (i-1) k} \right|^2 \right). \tag{2.15}\]

Under the mild condition that \(\delta\) is irrational [23], the average mutual information \(\frac{1}{N} I_N\) converges to

\[
I_{FS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left( 1 + \frac{\mathcal{E}_s}{MN_0} \left| \sum_{i=1}^{M} \alpha_i e^{j(i-1)\omega} \right|^2 \right) d\omega, \tag{2.16}\]

which is the same as the mutual information (2.14) achieved by the time shifting method.

By Jensen’s inequality, the mutual information (2.14) and (2.16) achieved by time and frequency shifting is less than or equal to that achieved by unconstrained signaling:

\[
I_{TS} = I_{FS} \leq I_{OPT},
\]

with equality if and only if a single component \(\alpha_i\) of \(\alpha\) is nonzero.
2.3.4 Randomly time weighted transmit signals

The randomly time weighted diversity scheme is similar to time division and frequency shifting, in that a scalar input $X$ is mapped linearly into a vector output $[X_1 X_2 \ldots X_M]^T$. For time division and frequency shifting the linear map varies periodically or quasi-periodically in time; for random weighting the linear map is random. The motivation behind random time weighting is the construction of a linear diversity method whose performance, like unconstrained signaling, depends only on $||\alpha||$.

In the random time weighting technique, the vector input is generated by multiplying the scalar input by a complex unit-magnitude random vector $\beta_k$:

$$X_k = \beta_k X_k,$$  \hspace{1cm} (2.17)

where $\beta_k = [\beta_{1,k} \ldots \beta_{M,k}]^T$ is chosen randomly and uniformly over the surface of the unit sphere. The output of the channel is then

$$Y_k = \alpha^T \beta_k X_k + V_k.$$  \hspace{1cm} (2.18)

The weighting vector $\beta_k$ is assumed to be known at the receiver; this is easily achieved by selecting $\beta_k$ pseudo-randomly according to a prearranged scheme.

To simplify (2.18), it is useful to interpret $\beta_k$ as the first column

$$\beta_k = U_k \left[ \begin{array}{c} 1 \\ 0 \\ \vdots \\ 0 \end{array} \right]^T$$  \hspace{1cm} (2.19)

of a random unitary matrix $U_k$ drawn from the circular unitary ensemble. The defining property of the circular unitary ensemble is invariance under unitary transformation. That is, each column of the unitary matrix drawn from the circular unitary ensemble has unit gain and points in a random direction on the surface of the unit sphere.

The vector of coefficients $\alpha$ may similarly be written as an appropriately normalized
2.3 Diversity methods and mutual information

unit vector multiplied by a unitary matrix $\hat{U}(\alpha)$. This rearrangement of (2.18) yields

$$Y_k = \begin{bmatrix} ||\alpha|| & 0 & \ldots & 0 \end{bmatrix} \hat{U}(\alpha)U_k \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} X_k + V_k. \quad (2.20)$$

Because the circular unitary ensemble is invariant under unitary transformation, the unitary matrix $\hat{U}(\alpha)U_k$ may also be interpreted as a random matrix drawn from the circular unitary ensemble. The channel output may therefore be written as

$$Y_k = ||\alpha||\mu_k X_k + V_k, \quad (2.21)$$

where $\mu_k$ denotes the upper left entry of a matrix drawn from the circular unitary ensemble.

When the input $X_k$ is i.i.d. Gaussian with energy $E|X_k|^2 = \mathcal{E}_s$ per symbol, the mutual information achieved by random time weighting is

$$I_{\text{RAN}} = \int_{0}^{1} \log \left( 1 + \frac{||\alpha||^2 \mathcal{E}_s \eta}{N_0} \right) f_{|\mu|^2}(\eta) \, d\eta, \quad (2.22)$$

where $f_{|\mu|^2}(\cdot)$ is the probability density function of the squared magnitude of any entry $\mu$ of a matrix drawn from the circular unitary ensemble.

To evaluate $f_{|\mu|^2}(\cdot)$, note that $\mu$ is a component of a vector chosen randomly and uniformly over the surface of the complex unit sphere. Such a vector can be generated taking a vector with zero-mean i.i.d. complex Gaussian components and scaling it by its norm. Thus, $\mu$ may be written as

$$\mu = \frac{G_1}{(|G_1|^2 + \cdots + |G_M|^2)^{1/2}}, \quad (2.23)$$

where $G_1, \ldots, G_M$ are zero-mean i.i.d. complex circular Gaussian random variables. The squared magnitude of a complex circular Gaussian is exponentially distributed, hence $|\mu|^2$ may be written as

$$|\mu|^2 = E_1/(E_1 + \cdots + E_M), \quad (2.24)$$
where $E_1, \ldots, E_M$ are i.i.d. exponential random variables. It is now straightforward to derive the required probability density function:

$$f_{|\mathbf{\mu}|^2}(\eta) = \begin{cases} (M - 1)(1 - \eta)^{M-2} & 0 < \eta \leq 1, \\ 0 & \text{otherwise,} \end{cases} \quad (2.25)$$

for $M \geq 2$.

By combining (2.22) and (2.25) and applying Jensen’s inequality, it can be shown that the mutual information is again less than that achieved by unconstrained signaling:

$$I_{\text{RAN}} \leq I_{\text{OPT}}.$$ 

Equality occurs only in the trivial case $||\alpha|| = 0$.

Like unconstrained signaling, but in contrast to the other linear methods, the mutual information achieved by random weighting depends only on the antenna gain $||\alpha||$. Choosing $\beta$ according to any other distribution that is invariant under unitary transformation also gives a mutual information that depends only on $||\alpha||$. However, the circular unitary ensemble is the only such distribution with $||\beta||$ identically 1, and thus it can be verified using Jensen’s inequality that all other such ensembles—which necessarily allow $||\beta||$ to vary—achieve strictly lower mutual information.

The effective scalar-input channel created by the random weighting technique may be reinterpreted by dividing and multiplying by $\sqrt{M}$ in (2.21):

$$Y_k = \frac{||\alpha||}{\sqrt{M}}Z_k \mathbf{X}_k + V_k, \quad (2.26)$$

where $Z_k = \sqrt{M} \mu_k$. From (2.23), as $M \to \infty$, $Z_k$ converges almost surely to a zero mean circular complex Gaussian random variable. In other words, as the number of antennas becomes large, the random time weighting channel approaches an i.i.d. Rayleigh faded channel with signal to noise ratio determined by the average antenna gain $||\alpha||^2/M$. It is important to emphasize, however, that the resulting (approximately) Rayleigh fading channel does not have an effectively infinite diversity order. Rather, the $M$-fold diversity in the antenna array appears in the term $||\alpha^2||/M$, which, for reasonable stochastic models on $\alpha$, exhibits less and less variability as $M$ becomes large. In particular, if the components of $\alpha$ are i.i.d.,
then $||\alpha^2||/M \to E|\alpha_1|^2$ almost surely as $M \to \infty$. This property will be explored further in Section 2.5.1.

While time division and frequency shifting are periodically time varying, they can be easily recast in a random framework. The randomized counterpart of frequency shifting chooses the phases of the antennas independently and uniformly over $[0, 2\pi)$ for each symbol, while the randomized version of time division selects an antenna at random for each symbol. The randomized and deterministic versions of a given diversity method have the same mutual information.

### 2.4 Outage regions

A useful way to interpret the mutual information formulas derived above for the case of $M = 2$ antennas is to plot lines of constant mutual information on the $|\alpha_1|, |\alpha_2|$ quarter plane. These curves delimit outage regions, in the sense that one can devise a codebook at any rate $R$ so that reliable communication will occur if the mutual information $I$ exceeds $R$.

As a representative example, the curves for $I = 1$ nat are shown in Figure 2.3: the solid innermost quarter circle is unconstrained diversity $I_{\text{OPT}}$; the dashed curve tangent to the quarter circle at $|\alpha_1| = |\alpha_2|$ is time division $I_{\text{TD}}$ and frequency division $I_{\text{FD}}$; the dotted curve tangent to the quarter circle at $|\alpha_1| = 0$ and $|\alpha_2| = 0$ is time shifting $I_{\text{TS}}$ and frequency shifting $I_{\text{FS}}$; the dash-dot outer quarter circle is randomly weighted transmit diversity $I_{\text{RAN}}$. For values of $(\alpha_1, \alpha_2)$ outside these curves, reliable communication is possible at rate 1; for values inside, attempting communication at rate 1 will result in outage.

The outage region for unconstrained signaling is strictly smaller than the outage regions for any of the five linear methods. This effect is caused in part by the symmetry in the diversity methods we have selected. For example, a time division scheme which uses one antenna element more often than another will have an outage region that contains only part of the outage region for unconstrained signaling.

None of the linear diversity methods dominates the others for all $\alpha$. To better distinguish the methods, a comparison in terms of outage probabilities will be done in Section 2.5 using a stochastic model for $\alpha$. 

---

- 45 -
Figure 2.3: Values of $|\alpha_1|$ and $|\alpha_2|$ that correspond to a mutual information of 1 nat/complex symbol.

The gap between the outage regions of random weighting and unconstrained signaling depend only on $||\alpha||$, and may therefore be evaluated even in the absence of a stochastic model for $\alpha$. Subtracting (2.22) from (2.5), inserting (2.25), and setting $z = M\eta$ yields

$$I_{OPT} - I_{RAN} = \int_0^M \log \left( \frac{1 + \text{SNR}}{1 + z\text{SNR}} \right) \frac{M - 1}{M} \left( 1 - \frac{z}{M} \right)^{M-2} dz,$$

(2.27)

where $\text{SNR} = \frac{||\alpha||^2 \xi}{M N_0}$ is the signal to noise ratio per antenna element. For large $M$ and $\text{SNR} \gg 1$, the difference is approximately

$$I_{OPT} - I_{RAN} \approx \int_0^\infty \log \left( \frac{1}{z} \right) e^{-z} dz$$

(2.28)

$$= \gamma,$$

(2.29)

where $\gamma$ is Euler's constant, which is about 0.577 nats or 0.833 bits. Hence, when the SNR per antenna element is large, random weighting needs roughly $10\gamma \log_{10} e \approx 2.51$ dB more signal power to achieve the same outage region as unconstrained signaling. The gap shrinks when the number of antennas is small. For example, at $M = 2$ the high-SNR gap is 0.307 nats (1.33 dB), at $M = 4$ it is 0.447 (1.94 dB), and at $M = 8$ it is 0.513 (2.23 dB).
2.5 Probabilistic channel model

A statistical model for the coefficient vector $\alpha$ allows the computation of outage probabilities and the comparison of multiple antenna systems to single antenna systems. The effectiveness of diversity is, of course, based on the idea that as the number of antenna elements increases, the probability of having poor gain for all the elements becomes small. Phrased differently, even when diversity does not improve the average SNR of the channel, it does significantly reduce the variance.

We assume the components of $\alpha$ are independent and identically distributed with uniform phase and Rayleigh magnitude. We continue to assume that $\alpha$ is effectively constant over long blocks of symbols. The Rayleigh model is generally viewed as appropriate when there is no line of sight to the receiver, when there are a large number of reflected paths, and when the antenna elements are at least 1/2 wavelength apart. If the antenna elements are extremely widely distributed relative to the propagation distance to the receiver, as would be the case if they were placed throughout a building, then a more complex model must be adopted that accounts for the strong attenuation of more distant antennas. We do not consider this case here. While it might appear that a stochastic model for $\alpha$ conflicts with the assumption of no time variation over a long block of symbols, the assumption is reasonable, for example, in a slow frequency hopped system where each block of symbols is transmitted on a different band with a different coefficient vector $\alpha$.

Since mutual information is a function of the random variable $\alpha$, it too is a random variable. We first compare the transmit diversity schemes in terms of expected mutual information. If the number of blocks $K$ used to transmit the message is large, and if the sequence of channel coefficient vectors $\alpha_k$, $k = 1, \ldots, K$, is ergodic, then the expected mutual information is the Shannon capacity in the usual sense. On the other hand, practical decoding constraints generally prevent $K$ from being large, so that with any finite number of antennas $M$ there is always some nontrivial probability of decoding error regardless of rate. The channel in this case therefore has no capacity in the usual sense; see Ozarow, Shamai, and Wyner [22] and references therein for a more thorough discussion. Among the diversity schemes considered here, only unconstrained signaling improves expected mutual information compared to a single antenna system.

A more useful measure of diversity is the probability that mutual information of a
Transmit diversity

single block drops below some critical value. Such an event is termed an outage. To reduce the probability of outage to a reasonable level, the signal to noise ratio must be boosted significantly above the value that would be needed if the channel were corrupted only by additive white Gaussian noise. We compute this loss in performance for the single antenna case, and measure the improvement afforded by diversity as $M$ increases. We then compare the probabilities of outage associated with the various diversity schemes for a two antenna system. Finally, we show that in the limit of infinitely many antennas, it is possible to transmit at rates up to the expected mutual information without errors using any of the transmit diversity schemes. This result was conjectured to be true for a channel equivalent to the time-shift diversity method in [22].

2.5.1 Expected mutual information

Consider first the baseline case of a channel with a single transmit and receive antenna:

$$Y_k = \alpha_1 X_k + V_k.$$  \hfill (2.30)

Assuming the input satisfies the energy constraint $E[|X_k|^2] \leq \mathcal{E}_s$, the mutual information of the channel as a function of $\alpha_1$ is

$$I(\alpha_1) = \log \left( 1 + \frac{|\alpha_1|^2 \mathcal{E}_s}{N_0} \right).$$ \hfill (2.31)

The expected mutual information

$$E[I(\alpha_1)] = E \left[ \log \left( 1 + \frac{|\alpha_1|^2 \mathcal{E}_s}{N_0} \right) \right]$$ \hfill (2.32)

is found by averaging over the squared magnitude $|\alpha_1|^2$, which is exponentially distributed. The expected mutual information of a single antenna channel may be interpreted as the capacity of an i.i.d. Rayleigh fading channel with perfect channel state information at the receiver \cite{4}. Applying Jensen's inequality to (2.32) shows that the expected mutual information is less than the capacity of an additive white Gaussian noise (AWGN) channel with
the same average SNR:

\[ E[I(\alpha_1)] < \log \left( 1 + \frac{E[|\alpha_1|^2]E_s}{N_0} \right) = C_{AWGN}. \]  

(2.33)

(2.34)

The mutual information of an \( M \)-antenna channel using unconstrained signaling (c.f. (2.5)) is

\[ I_{OPT}(\alpha) = \log \left( 1 + \frac{||\alpha||^2E_s}{MN_0} \right). \]  

(2.35)

As \( M \rightarrow \infty, ||\alpha||^2/M \) converges almost surely to its expected value, hence \( I_{OPT}(\alpha) \) converges almost surely to \( C_{AWGN} \). It can also be shown that the expected value of \( I_{OPT}(\alpha) \) increases monotonically to \( C_{AWGN} \). Thus, for unconstrained signaling, as the number of antennas become large the variability in \( \alpha \) effectively disappears.

The expected mutual information of the remaining diversity schemes—time division, frequency division, time shifting, frequency shifting, and random weighting—are all equal to the single-antenna case (2.32), as we now show.

The mutual informations achieved by time division and frequency division are equal for all \( \alpha \), hence their expected mutual informations are equal. Similarly, the expected mutual informations achieved by time shifting and frequency shifting are equal. To show that the expected value of the respective mutual informations (2.7), (2.16), and (2.22) of time division, frequency shifting, and random weighting are equal, we make a direct argument from the channel model.

As discussed in Section 2.3.4, time division, frequency shifting, and random weighting all have the same general form. The channel output is

\[ Y_k = \alpha^T \beta_k X_k + V_k, \]  

(2.36)

where \( \beta_k \) is chosen either randomly or according to some deterministic pattern. Since in all
cases $\|\beta_k\| = 1$, (2.36) may be rewritten as

$$Y_k = \alpha^T U_k \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T X_k + V_k$$  \hspace{1cm} (2.37)$$

for some suitably chosen sequence of unitary matrices $U_k$. The components of $\alpha$ are i.i.d. circular complex Gaussian random variables, hence the distribution of $\alpha$ is invariant under unitary transformation. This implies that $Y_k$ is equal in distribution to

$$\tilde{Y}_k = \alpha^T \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T X_k + V_k$$  \hspace{1cm} (2.38)$$

$$= \alpha_1 X_k + V_k.$$  \hspace{1cm} (2.39)$$

The expected mutual information

$$E[I(X_k; Y_k)] = E[I(X_k; \tilde{Y}_k)]$$  \hspace{1cm} (2.40)$$

is therefore equal to that of the single-antenna channel (2.32).

2.5.2 Outage probabilities

From Section 2.4, the outage region of the diversity system using unconstrained signaling lies inside the outage regions of the simpler diversity schemes, hence its outage probability is smaller for any stochastic model on $\alpha$. We now quantify this observation when $\alpha$ has a Rayleigh model, and also compare the simpler diversity schemes, namely time division/frequency division, time shifting/frequency shifting, and random weighting.

All mutual informations and outage probabilities may be parameterized by the expected signal to noise ratio

$$\text{SNR} = \frac{E[|\alpha_1|^2]}{N_0} \mathcal{E}_s.$$  \hspace{1cm} (2.41)$$

Without loss of generality we assume that $E[|\alpha_1|^2] = 1$.

The outage probability for unconstrained signaling is straightforward to evaluate. The squared magnitude $||\alpha||^2$, a sum of $M$ independent exponential random variables, has
2.5 Probabilistic channel model

Table 2.1: Loss $\delta_{M,\epsilon}/M$ in dB with respect to AWGN channel for “unconstrained” transmit diversity.

<table>
<thead>
<tr>
<th>$M$</th>
<th>$\epsilon = 10^{-2}$</th>
<th>$\epsilon = 10^{-4}$</th>
<th>$\epsilon = 10^{-6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-19.98</td>
<td>-40.00</td>
<td>-60.00</td>
</tr>
<tr>
<td>2</td>
<td>-11.29</td>
<td>-21.48</td>
<td>-31.50</td>
</tr>
<tr>
<td>4</td>
<td>-6.87</td>
<td>-12.37</td>
<td>-17.51</td>
</tr>
<tr>
<td>8</td>
<td>-4.40</td>
<td>-7.61</td>
<td>-10.43</td>
</tr>
<tr>
<td>16</td>
<td>-2.91</td>
<td>-4.91</td>
<td>-6.57</td>
</tr>
<tr>
<td>256</td>
<td>-0.65</td>
<td>-1.06</td>
<td>-1.36</td>
</tr>
</tbody>
</table>

cumulative distribution function

$$P(||\alpha||^2 \leq \delta) = \int_0^{\delta} \frac{x^{M-1}}{(M-1)!} e^{-x} \, dx$$  \hspace{1cm} (2.42)

$$= e^{-\delta} \sum_{r=M}^{\infty} \frac{\delta^r}{r!}$$  \hspace{1cm} (2.43)

$$\approx \frac{1}{M!} \delta^M e^{-\delta}.$$  \hspace{1cm} (2.44)

Let $\delta_{M,\epsilon}$ be the threshold at which outage occurs with probability $\epsilon$ for $M$ antennas, i.e.,

$$P(||\alpha||^2 \leq \delta_{M,\epsilon}) = \epsilon.$$  \hspace{1cm} (2.45)

Using this notation, the unconstrained diversity method with $M$ antennas achieves an outage probability of $\epsilon$ at rate

$$R = \log \left(1 + \frac{\delta_{M,\epsilon} E_{\text{sec}}}{MN_0}\right).$$  \hspace{1cm} (2.46)

The dB loss in signal power of the diversity channel with respect to an AWGN channel is therefore $10 \log_{10}(\delta_{M,\epsilon}/M)$. The loss is tabulated for various values of $M$ and $\epsilon$ in Table 2.1. Even a small number of antennas dramatically reduces the power needed to achieve reliable communication. As is typical for $M$-fold diversity, the loss approximately halves when $M$ is doubled. In accordance with the results in Section 2.5.1 the loss approaches 0 dB as $M \to \infty$. 
Transmit diversity

The outage probabilities of the suboptimal methods are rather similar, hence we evaluate them only for \( M = 2 \). Ozarow et al. [22] calculate the probabilities of outage—or equivalently, the cumulative distribution functions of mutual information—which correspond to time shifting [22, (2.26b)] and time division [22, (3.4)]. We compute the cumulative distribution function of the mutual information achieved by the randomly weighted method using Monte-Carlo simulation. These curves are used to plot the maximum rate of reliable communication as a function of signal to noise ratio when the outage probability is held at 1\%. The gap between random weighting and time shifting is nearly zero. Unconstrained diversity is somewhat better than random weighting at high SNR; as shown in Section 2.4, with \( M = 2 \) antennas the gap is about 0.307 nats or 1.33 dB.

2.5.3 Communication at rates up to expected mutual information

For a diversity scheme employing only a finite number of antennas there is a nonzero outage probability at any nonzero communication rate. In the limit of infinitely many transmit antennas, however, we saw in Section 2.5.1 that the mutual information achieved by unconstrained signaling converges almost surely to the capacity of an AWGN channel with the same average SNR. In other words, the outage probability can be driven to zero by using sufficiently many antennas. In this section we argue that the same holds for any of the simpler transmit diversity schemes, albeit at the lower communication rate of a Rayleigh fading channel rather than an AWGN channel.
2.6 Discussion

The result is straightforward to prove for time division diversity. By the strong law of large numbers, the mutual information achieved by $M$-antenna time division (2.7) converges almost surely to its expected value

$$
\lim_{M \to \infty} I_{TD} = E \left[ \log \left( 1 + \frac{|\alpha_1|^2 \mathcal{E}_s}{N_0} \right) \right].
$$

(2.47)

The proof for time shifting diversity, given in Appendix A, is based on the idea that as $M$ gets large, $H(\omega)$ in (2.12) becomes statistically independent at distinct values of $\omega$. Integrating $\log(1+|H(\omega)|^2 \mathcal{E}_s/N_0)$ over $\omega$ then corresponds to adding up an infinite number of independent random variables, which by the law of large numbers converges to the expected mutual information.

Since the mutual information of the channels using frequency division and time shifting have a form equivalent to time division and frequency shifting, respectively, the proofs above also show that the mutual information achieved by these methods converges to its expected value as $M \to \infty$. The proof for the random weighting technique is given in Appendix B.

2.6 Discussion

2.6.1 Comparison to ideal beamforming

The transmit diversity methods studied above are appropriate for the scenario where there is a single receiver with unknown coefficient vector $\alpha$, or for a broadcast scenario where the number of receivers is much larger than $M$. It is important to contrast these two fundamentally different problems: with a single receiver, performance can be dramatically improved using feedback to send information about the channel to the transmitter, but when the goal is to ensure a uniform SNR among a large number of receivers the achievable performance is inherently more limited. To make this comparison more precise, we determine the mutual information for the case of a single receiver with $\alpha$ known perfectly at the transmitter.

If the transmitter knows the coefficient vector $\alpha$, mutual information is maximized
Transmit diversity

using beamforming. Beamforming is the $M' = 1$ case of a system with $M$ transmit antennas and $M'$ receive antennas [27]. Mutual information is maximized by performing a singular value decomposition of the antenna coefficient matrix to separate the channel into a set of \( \min\{M, M'\} \) independent Gaussian channels with noise variances determined by the singular values. Water pouring then applies. When $M$ or $M'$ equals 1, there is only 1 nonzero singular value, and water pouring becomes trivial.

For beamforming at the transmitter, mutual information is maximized by setting

$$X = \frac{\alpha^*}{||\alpha||} X,$$  \hspace{1cm} (2.48)

where $X$ is a zero-mean complex Gaussian random variable with variance $E|X|^2 = \mathcal{E}_s$. The channel output is then

$$Y = ||\alpha||X + V,$$  \hspace{1cm} (2.49)

and the mutual information is

$$I_{BMP} = \log \left(1 + \frac{||\alpha||^2\mathcal{E}_s}{N_0}\right).$$  \hspace{1cm} (2.50)

Comparison to (2.5) reveals that the signal-to-noise ratio achieved by beamforming is a factor of $M$ better than that achieved by unconstrained diversity signaling. This is quite natural, and can be interpreted as follows.

Frequency shifting, time division, and random weighting choose for each symbol a rank 1 covariance matrix $\Lambda$ for $X$, where the principal component of $\Lambda$ follows a pattern that depends on the method. If the principal component lies in nearly the same direction as (or is a phase-shifted version of) $\alpha$ then performance is as good as beamforming. If not, the signal is highly attenuated. The probability of a random vector in $M$-space aligning with $\alpha$ is roughly $1/M$, hence the loss in performance compared to ideal beamforming. In effect, the linear diversity methods create a beam pattern in the antenna field that changes from symbol to symbol; the density of peaks in this field is roughly $1/M$ times the density of nulls or near-nulls.

Unconstrained signaling performs better because it reduces variability. The covariance
matrix of $X$ is a scaled identity matrix, hence the curves of constant likelihood are spheres, and the fraction of energy in $X$ in the direction of $\alpha$ is always $1/M$. Rather than aiming the antenna in a different direction for each symbol, unconstrained signaling creates a field that is spatially white.

2.6.2 Channel coding strategies

Among the diversity methods considered here, unconstrained signaling and random weighting seem most appealing in terms of robustness and performance. As discussed in Section 2.3.4, the random weighting method creates a channel that approximates i.i.d. Rayleigh fading as the number of antennas increases. A codebook designed for a Rayleigh fading channel will therefore be effective.

Practical codes for unconstrained signaling require more effort. Using, for example, standard QPSK signal sets on each antenna results in an undecodable tangle after the signal sets are rotated, scaled, and summed. One solution to this problem is to convert unconstrained signaling from a vector-input time-invariant channel into $M$ (approximately) i.i.d. Rayleigh fading channels driven by $M$ virtual users. Standard coding methods for Rayleigh channels can then be applied.

These i.i.d. Rayleigh fading channels can be generated by premultiplying by a sequence of matrices $U_k$ drawn pseudorandomly from the circular unitary ensemble. The output of the channel is then

$$Y_k = \alpha^T U_k X_k + V_k$$
$$= ||\alpha|| \beta_k^T X_k + V_k,$$

where $\beta$ is the first column of a matrix from the circular unitary ensemble. The channel remains i.i.d. memoryless, but, because the components of $\beta$ have the same marginal densities, each virtual antenna element now "looks" the same. The transmitter can split itself into $M$ virtual users, time-synchronized but otherwise noncooperative, and use multiple access coding. The $i$th virtual user controls the $i$th component of $X_k$; transmit energy is shared evenly.

Consider the case of two antennas, $M = 2$. If the receiver uses successive decoding
(stripping), the first and second virtual users achieve mutual informations of

$$I_{OPT1} = E_{\beta_{1,k}, \beta_{2,k}} \left[ \log \left( 1 + \frac{||\alpha||^2 ||\beta_{1,k}||^2 \varepsilon_s / 2}{N_0 + ||\alpha||^2 ||\beta_{2,k}||^2} \right) \right]$$

(2.53)

and

$$I_{OPT2} = E_{\beta_{2,k}} \left[ \log \left( 1 + \frac{||\alpha||^2 ||\beta_{2,k}||^2 \varepsilon_s / 2}{N_0} \right) \right]$$

(2.54)

respectively, where the first user has lower mutual information due to interference from the second. This method achieves an extremum point of the multiaccess achievable rate region. Furthermore, the mutual informations depend on $\alpha$ only through its magnitude $||\alpha||$; this critical property allows the rates of the two virtual users to be selected so that decoding simultaneously fails when $||\alpha||$ drops below a design threshold. Since both decodings must succeed to allow the virtual users to be recombined at the receiver, simultaneous failure is necessary for efficient operation.

As a consistency check, the total mutual information of the channel is

$$I_{OPT} = I_{OPT1} + I_{OPT2}$$

(2.55)

$$= E_{\beta_{1,k}, \beta_{2,k}} \left[ \log \left( 1 + \frac{||\alpha||^2 ||\beta||^2 \varepsilon_s}{2N_0} \right) \right]$$

(2.56)

$$= \log \left( 1 + \frac{||\alpha||^2 \varepsilon_s}{2N_0} \right),$$

(2.57)

since $||\beta|| \equiv 1$. Given the convexity of mutual information it seems paradoxical that converting a time-invariant channel into a time-varying one does not degrade performance. The resolution to this conundrum is that the two virtual channels do not fade independently: when one is good the other is bad, and the full antenna gain is used at all times.

Implementation complexity for this virtual $M$-user system is roughly $M$ times larger than that of a single-antenna system. The channel identification and power-control problems normally associated with stripping are not as severe as with a true $M$-user system because the multiple virtual users arise from the coordinated action of a single user. However, given that the performance gap of roughly 2.51 dB between unconstrained diversity and the simpler random weighting method does not change significantly with the number of antennas, this implementation of unconstrained diversity is probably worthwhile only for small or moderate
2.6.3 Adding diversity to existing systems

Any of the linear transmit diversity methods can be easily implemented in continuous time at passband. This is a simple way to upgrade a system designed for a single antenna to use transmitter diversity, though of course the full performance gain will accrue only if the transmitter and receiver processing is redesigned for the (artificially created) time-varying channel. The time variation follows a pattern known to the receiver, so channel identification is no harder than learning the slowly varying coefficient vector \( \alpha \). If the error-correction portion of the existing system cannot be changed then the linear signal processing methods developed in [35] are a good near-optimal alternative.

Perhaps the most obvious approach to adding diversity to an existing system is to send identical copies of the signal over each antenna using orthogonal frequency bands; see, e.g., Jakes [14]. Unlike the frequency division method described in Section 2.3, the bandwidth expands by a factor of \( M \). A simple computation shows that, interestingly, the mutual information achieved by this method is equal to that of unconstrained signaling. From this perspective, such an approach is particularly inefficient, as no benefit is obtained from the \( M \)-fold increase in bandwidth. On the other hand, the complexity of unconstrained “minimum bandwidth” signaling is comparatively high: it requires the use of error-correction codes designed for time-varying channels and an \( M \)-user decoder. One may conclude that if bandwidth is plentiful then naive repeated-transmission antenna diversity is almost certainly preferable to any of the diversity methods described in this chapter.

In fact, repetition diversity is frequently used in practical systems. Qualcomm, for example, uses this approach to adapt their direct sequence spread spectrum cellular system to indoor use. Multiple antennas are arranged around the perimeter of a building, each carrying a delayed copy of the transmit signal. The delays vary by a fraction of a symbol duration. The pseudonoise sequence and the high signal bandwidth ensure that the delayed copies are nearly orthogonal, hence (unlike time shifting) there is little intersymbol interference. A RAKE receiver in the Qualcomm handset combines the delayed paths in a near-optimal manner.
2.6.4 Diversity vs. directive arrays

The communication and signal processing literature loosely uses terms such as “diversity,” “directive arrays,” and “beamforming” to describe the ways in which multiple antennas—at either the transmitter or receiver—can be used to improve communication. There appears to be no widely accepted definitions of these terms, and, interestingly, the literature on diversity is often conceptually orthogonal to that on beamforming. Compare, for example, Jakes [14] to Johnson and Dudgeon [15]. In order to place our results in the context of this literature, we argue that the difference between diversity and beamforming is not so much in their methods, but rather in their modeling assumptions for the channel.

In the basic directional array or beamforming scenario, there is no stochastic model for the channel parameters $\alpha$, there is a dominant direction of arrival for the signal, implying that the components of $\alpha$ have a particular structure, and the antenna is assumed to know the arrival angle almost perfectly. The array pattern is adjusted by manipulating the gains and phases of the antenna elements, or, more generally, by combining the array with a multiple-input or multiple-output linear filter. For a transmit array, the channel knowledge is used to focus as much energy in the direction of the receiver as possible. For a receive array, the gain of the antenna is maximized in the direction of the transmitter. By reciprocity, transmitter and receiver beamforming have the same performance, e.g., the same SNR or mutual information.

Diversity, on the other hand, inherently assumes a stochastic model for the channel parameters $\alpha$. The stochastic model is derived assuming a large number of multipath reflections, so that signals do not have a meaningful notion of direction of arrival and the components of $\alpha$ are comparatively much less correlated. As with beamforming, the array pattern is adjusted by manipulating the gains and phases of the antenna elements. However, the array does not have full knowledge of the channel: the phases are generally unknown while the gains may be unknown or partly known through a statistical characterization. In other words, the array is free to choose an antenna directivity pattern, but it does not know where the peaks and nulls will fall. The receiver usually has more information about the channel than the transmitter, which means that transmit and receive diversity may behave quite differently.

The distinction between diversity and directional arrays blurs in many applications. For example, in a diversity setting, as bandwidth increases, multiple paths may begin to
2.6 Discussion

resolve, rendering a stochastic model invalid. In a beamforming setting, as the number of transmit antenna elements increases and the power per antenna element decreases, channel identification becomes more difficult, making the assumption of perfect channel knowledge unrealistic. The achievable performance in such cases depends critically on the channel model. We view the results in this chapter as a step towards understanding the "diversity" endpoint of the continuum between beamforming and diversity applications.
Chapter 3

Transmit diversity over continuous-time channels

In the previous chapter we analyzed several transmit diversity schemes using a discrete-time channel model. When digital signal processing hardware is used to implement the antenna diversity systems, a discrete-time analysis is appropriate. The diversity methods can also be implemented in continuous-time at passband. The delays, switches, modulators and filters can be implemented directly at the transmitting antennas. This is a simple way to upgrade an existing single antenna system to use multiple antennas.

A possible approach for analyzing the continuous-time system is to assume that the input signals are bandlimited to $W$ Hz and then represent the input signals with $2W$ samples per second. This approach leads to mathematical difficulties since a strictly bandlimited signal can not be strictly limited in time and since capacity is defined [8] as the largest mutual information that can be transmitted over a single channel use.

In this chapter, we partially extend the mutual information results of the previous chapter to the continuous-time channel model. The mutual information of several of the diversity methods is calculated following the approach in Ozarow, Shamai, and Wyner [22, App. A] in which the continuous-time complex waveforms are represented as a series expansion of a set of orthonormal waveforms. The mutual information is derived by treating the waveform channel as an infinite set of parallel discrete channels parameterized by the eigenvalues and coefficients of the series expansion. For unconstrained signaling, time-division, frequency-division, time-shifting, and random weighting, we show that the resulting mutual informations are equivalent to those derived for the discrete-time channel model. For frequency shifting diversity, the mutual information is calculated using an analogous channel model based on the frequency (rather than time) variable. For this calculation, we assume the bandwidth $W \rightarrow \infty$. 

- 61 -
3.1 Channel model

The $M$-antenna transmitter diversity channel was shown in Figure 2.1. The continuous-time channel model differs from the discrete-time version as follows. For the continuous-time channel model, the complex baseband received signal is

$$ y(t) = \sum_{i=1}^{M} \alpha_i x_i(t) + v(t) $$

$$ = u(t) + v(t). $$

The $M$ input signals, $x_1(t), x_2(t), \ldots, x_M(t)$, are assumed to be time limited sample functions, constrained to the interval $[-T/2, T/2]$, of complex stationary random processes with respective power spectral densities $S_{X_i}(f)$ strictly bandlimited to $W$ Hz. The total transmit power is limited to $\sum_{i=1}^{M} \int S_{X_i}(f) \, df \leq P$. The additive noise, $v(t)$, is assumed to be a sample function of white circular Gaussian with two-sided power spectral density $N_0$ ($N_0/2$ per real and imaginary component).

Since we can represent the continuous-time channel with approximately $WT$ complex discrete channels, it can be shown that the discrete-time energy constraint $E_s$ is related to the continuous-time power constraint by $E_s = P/W$. This follows from the relations:

$$ \int S_{X_1}(f) \, df = E[|x(t)|^2] $$

and

$$ \int E[|x(t)|^2] \, dt = \sum_{n=1}^{WT} E[|X_n|^2], $$

where $\{X_n\}$ are coefficients of the expansion, $X_n = \int x(t) \phi_n(t) \, dt$, of $x(t)$ in terms of the prolate spheroidal wave functions $\phi_n(t)$ that are strictly time limited to $T$ and approximately bandlimited to $W$. 

- 62 -
3.2 Mutual information calculations

Continuous-time versions of the diversity methods analyzed in Chapter 2 are shown in Figure 3.1 for the case of \( M = 2 \) transmit antennas. The mutual information of five of these diversity methods, unconstrained signaling, time-division, frequency-division, and time shifting, and random weighting is easily calculated under the continuous-time channel model. The mutual information calculation is considerably more cumbersome for frequency shifting.

![Diagram](image)

(a) Unconstrained signaling  
(b) Time division

![Diagram](image)

(c) Frequency division  
(d) Time shifting

![Diagram](image)

(e) Frequency shifting  
(f) Random weighting

Figure 3.1: Continuous-time transmit diversity schemes illustrated for \( M = 2 \) antennas.

To calculate the mutual information of the unconstrained signaling, time-division, frequency-division, time shifting, and random weighting diversity schemes, we use the approach described in [22, App. A] for continuous-time channels. We compute the mutual information between the channel input and output as a function of the channel coefficients,
Transmit diversity over continuous-time channels

\( \alpha \). We assume \( U(t) = \alpha^T \mathbf{X}(t) \) is a finite duration replica of a bandlimited complex process with power spectral density (psd) \( S_U(f) \). The time restriction follows directly from the definition of capacity as the maximum mutual information over a single channel use. The output process (psd \( S_Y(f) \)) is truncated to the same time interval \( T \) and we observe the mutual information per unit time in the limit as \( T \) becomes large. To compute this mutual information, the Karhunen-Loeve expansion is used to convert the waveform channel into an infinite set of parallel channels parameterized by expansion eigenvalues. The formula for the mutual information of this set of parallel channels is lower bounded using an application of the theorem by Kac, Murdock, and Szego which characterizes the behavior of the eigenvalues in the limit as \( T \to \infty \). From Ozarow [22, App. A], the lower bound is approached with equality for \( WT \to \infty \) as

\[
\lim_{WT \to \infty} S_U(f) \ast \frac{T \sin^2(\pi f T)}{(\pi f)^2} = S_U(f).
\]

Under this model, the channel inputs are strictly time limited, but only approximately bandlimited. A more precise continuous-time channel model is described in Gallager [8] where the time limited inputs are strictly bandlimited by the channel. The mutual information analysis of several of the diversity systems is considerably more difficult using this model.

Analogous to the discrete-time analysis in which the input codebooks were derived from complex circular white Gaussian random processes, we assume the input signals are derived from complex circular Gaussian random processes whose power spectral density is uniform over their bandwidth. Recall that beamforming and waterfilling methods cannot be applied since the transmitter lacks knowledge of the channel coefficients.

3.2.1 “Unconstrained” signaling

Unconstrained, the multiple transmit antenna channel is a memoryless, vector-input, scalar-output, bandwidth and power limited, additive white Gaussian noise channel. The complex baseband received signal is

\[
Y(t) = \alpha^T \mathbf{X}(t) + V(t)
\]
3.2 Mutual information calculations

where \( \mathbf{X}(t) = [X_1(t), \ldots, X_M(t)] \) and \( V(t) \) is complex white Gaussian noise with power spectral density \( N_0 \). If the antenna input signal \( X_i(t) \) is a time-limited replica of a zero-mean complex Gaussian random processes with uniform power spectral density \( S_{X_i}(f) = \frac{P}{M_W} \) for \( |f| \leq W/2 \) and the antenna inputs are independent, then the channel output \( Y(t) \) is a time limited slice of a zero-mean complex Gaussian random process with power spectral density \( S_Y(f) = ||\alpha||^2 \frac{P}{M_W} + N_0 \). The mutual information in bits per second per Hertz is then

\[
I_{\text{opt}} = \frac{1}{W} \int_{-W/2}^{W/2} \log \left( \frac{S_Y(f)}{N_0} \right) df
\]

(3.7)

\[
= \log \left( 1 + \frac{||\alpha||^2 P}{M_W N_0} \right).
\]

(3.8)

Using the relation \( \mathcal{E}_s = P/W \) shows the mutual information formulas for the discrete-time and continuous-time channels are equivalent as expected.

3.2.2 Time division

For time-division diversity, the input is time-multiplexed over the \( M \) transmit antennas. The \( i \)th antenna input is a finite duration replica, time limited to \( \frac{t}{2} - \frac{(i-1)T}{M} \leq t < \frac{t}{2} + \frac{iT}{M} \), of a complex Gaussian random process with psd \( S_{X_i}(f) = \frac{P}{W} \). The channel output for the \( i \)th time slice, \( [\frac{t}{2} - \frac{(i-1)T}{M}, \frac{t}{2} + \frac{iT}{M}] \), is a function of the \( i \)th transmitted signal and the \( i \)th antenna coefficient. The mutual information between input and output is just the sum of \( M \) mutual informations

\[
I_{\text{TD}} = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{W} \int_{-W/2}^{W/2} \log \left( 1 + |\alpha_i|^2 \frac{S_{X_i}(f)}{N_0} \right) df
\]

(3.9)

\[
= \frac{1}{M} \sum_{i=1}^{M} \log \left( 1 + \frac{|\alpha_i|^2 P}{WN_0} \right).
\]

(3.10)

As expected, (3.10) is equivalent to the mutual information of the frequency-division scheme over the discrete-time channel if \( P/W \) is replaced by \( \mathcal{E}_s \).
3.2.3 Frequency division

As shown in Figure 3.1c, for frequency-division diversity, each input antenna is used over a fraction 1/M of the total channel bandwidth W Hz. The input for the i-th antenna, \( X_i(t) \), is a finite duration replica of a complex Gaussian random process bandlimited to \( \frac{(i-1)W}{2M} \leq |f| < \frac{iW}{2M} \) and with psd \( S_{X_i}(f) = \frac{P}{W} \). Thus, \( U(t) = \sum_{i=1}^{M} \alpha_i X_i(t) \) is a time limited version of a complex Gaussian random process with psd

\[
S_U(f) = \begin{cases} 
|\alpha_1|^2 \frac{P}{W} & 0 \leq |f| < \frac{W}{2M}, \\
|\alpha_2|^2 \frac{P}{W} & \frac{W}{2M} \leq |f| < \frac{W}{M}, \\
& \vdots \\
|\alpha_M|^2 \frac{P}{W} & \frac{(M-1)W}{2M} \leq |f| < \frac{W}{2}. 
\end{cases} 
\tag{3.11}
\]

The mutual information between input and output is

\[
I_{FD} = \frac{1}{W} \int_{-W/2}^{W/2} \log \left( 1 + \frac{S_U(f)}{N_0} \right) df 
\tag{3.12}
\]

\[
= \frac{1}{M} \sum_{i=1}^{M} \log \left( 1 + \frac{|\alpha_i|^2 P}{WN_0} \right). \tag{3.13}
\]

As in the discrete-time channel case, both time-division and frequency division yield the same mutual information. If \( P/W \) in (3.13) is replaced by \( \mathcal{E}_s \), the continuous-time and discrete-time mutual informations are equivalent.

3.2.4 Time shifting

The continuous-time version of Winters’ time shifting diversity scheme involves sending delayed versions of a common input signal over the multiple transmit antennas. The i-th antenna carries \( X(t - (i-1)\Delta) \), where the delay \( \Delta \) is arbitrary. We assume the delayed signals are present during the entire interval \([ -\frac{T}{2}, \frac{T}{2} ]\). Let the input \( X(t) \) be a finite duration replica of a complex Gaussian process with psd \( S_X(f) = \frac{P}{MW} \) for \( f \in [-W/2, W/2] \). Then \( U(t) = \sum_{i=1}^{M} \alpha_i X(t - (i-1)\Delta) \) is a time limited replica of a complex Gaussian process with psd \( S_U(f) = S_X(f)|H(f)|^2 \) where \( H(f) = \sum_{i=1}^{M} \alpha_i e^{-j(i-1)2\pi f\Delta} \). The mutual information of
the time shift diversity method is

\[ I_{TS} = \frac{1}{W} \int_{-W/2}^{W/2} \log \left( 1 + \frac{S_X(f)|H(f)|^2}{N_0} \right) df \]  

(3.14)

\[ = \frac{1}{W} \int_{-W/2}^{W/2} \log \left( 1 + \frac{P}{WMN_0} \left| \sum_{i=1}^{M} \alpha_i e^{-j(i-1)2\pi f \Delta} \right|^2 \right) df. \]  

(3.15)

Time shift diversity creates a channel that is varying periodically in frequency. If \( \Delta \) is chosen to be an integer multiple of \( \frac{1}{W} \) then the integration in (3.15) is over an integer number of periods and the mutual information is

\[ I_{TS} = \int_{-1/2}^{1/2} \log \left( 1 + \frac{P}{WMN_0} \left| \sum_{i=1}^{M} \alpha_i e^{-j(i-1)2\pi f} \right|^2 \right) df. \]  

(3.16)

Alternatively, in the limit of large values of \( \Delta \), where \( T \) is assumed to be increasing faster than \( \Delta \), the effect of integrating over a non-integer number of periods becomes negligible. Replacing \( P/W \) in (3.16) by \( E_s \) yields the mutual information formula for the discrete-time time shift diversity channel.

### 3.2.5 Random weighting

The channel output for the continuous-time version of the random weighting diversity scheme has the form:

\[ y(t) = \alpha^T \beta(t)x(t) + v(t), \]

(3.17)

where \( \beta(t) = [\beta_1(t) \cdots \beta_M(t)]^T \) is chosen to vary randomly and uniformly over the surface of the unit sphere as a function of time. As in Chapter 2, we can use unitary matrices to
represent $\beta(t)$ and $\alpha$ in (3.17) more compactly. This yields

$$Y(t) = \begin{bmatrix} ||\alpha|| & 0 & \ldots & 0 \end{bmatrix} \hat{U}(\alpha) U(t) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} X(t) + V(t)$$  \hspace{1cm} (3.18)

$$= ||\alpha|| \mu(t) X(t) + V(t),$$  \hspace{1cm} (3.19)

where $U(t)$ is a random matrix process. At each time $t$, $U(t)$ is a matrix drawn from the circular unitary ensemble. The random process $\mu(t)$ denotes the upper left entry of the matrices generated by $U(t)$. We define $\beta(t)$ and consequently $\mu(t)$ to be constant over intervals of length $T'$. We then compute the mutual information between the input and output over time interval $T >> T'$.

Define $\mu(t) = \mu_i$ for $t \in \left[-\frac{T}{2} + (i-1)T', -\frac{T}{2} + iT' \right]$. Defined in this way, random weighting is similar to time division: the channel gain, $||\alpha|| \mu(t)$, is constant over time slices of duration $T'$. Let the input for the $i$th time slice $X_i(t)$ be a finite duration replica of a complex Gaussian process with psd $S_{X_i}(f) = \frac{P}{W}$ restricted to bandwidth $W$. Also, define $L = T/T'$. Then the mutual information in the limit as $T' \to \infty$ is

$$I_{\text{RAN}} = \frac{1}{L} \sum_{i=1}^{L} \frac{1}{W} \int_{-W/2}^{W/2} \log \left( 1 + ||\alpha||^2 \mu_i^2 \frac{S_{X_i}(f)}{N_0} \right) df$$  \hspace{1cm} (3.20)

$$= \frac{1}{L} \sum_{i=1}^{L} \log \left( 1 + ||\alpha||^2 \mu_i^2 \frac{P}{WN_0} \right).$$  \hspace{1cm} (3.21)

In the limit as $L \to \infty$, by the strong law of large numbers, this mutual information converges to

$$I_{\text{RAN}} = \int_0^1 \log \left( 1 + \frac{||\alpha||^2 \eta P}{WN_0} \right) f_{\mu^2}(\eta) d\eta,$$  \hspace{1cm} (3.22)

where $f_{\mu^2}(\cdot)$ is the probability density function of the squared magnitude of any entry $\mu$ of a matrix drawn from the circular unitary ensemble. Replacing $P/W$ by $E_s$ in (3.22) shows that the continuous-time mutual information is equivalent to the discrete-time mutual information (2.22).
3.2.6 Frequency shifting

In frequency shifting diversity, modulated versions of a common input signal are sent over the multiple antennas: the $i$th antenna carries the signal $e^{j(\xi - 1)\delta} X(t)$, where $\delta > 0$ is arbitrary. This method converts the vector input channel into a periodically time varying scalar input channel. The Kharunen-Loeve (K-L) expansion of this is difficult. Specifically, we were unable to calculate the K-L eigenvalues of this time-varying channel and no theorem analogous to the Szego theorem is available to characterize the behavior of these eigenvalues. For time-division and random weighting diversity, we were able to circumvent the problem of computing $K - L$ eigenvalues of a time-varying channel by assuming the channel was constant over long blocklengths. For frequency shifting diversity we follow an alternative approach.

Frequency shifting is simply a dual of time shifting. Therefore we suspect this channel is easier to analyze in the frequency domain. To facilitate this analysis, we define a continuous-time channel model parallel to the model used in [22, App. A]. For our parallel model, the signals are defined in the frequency domain. We assume $X(f), U(f)$ and $Y(f)$ are finite bandwidth (limited to $W$) replica of complex Gaussian random processes with power spectral densities, $S_X(t), S_U(t)$, and $S_Y(t)$ respectively, limited to time $T$.

![Frequency shifting diversity scheme using new continuous-time channel model.](image)

The parallel frequency-domain channel model implementing the frequency shift diversity method is shown in Figure 3.2. In this model, the signals are convolved in frequency
and multiplied in time. Thus the filters become Doppler shifters and vice versa. The validity of this model also requires looking at white noise in a slightly different way.

In Gallager [8], the definition of white noise is physically based on the assumption that the noise at two time instants separated by a small interval \( \epsilon \) is approximately statistically independent. It is further assumed that this \( \epsilon \) is so small that from Central Limit Theorem arguments, the output from any linear filter can be modeled as a Gaussian random variable. For a white complex Gaussian noise random process \( V(t) \) and any set of \( L_2 \) functions \( \{ \phi_i(t) \} \), the \( V_i = \int \phi_i(t)V(t) \, dt \) are zero mean, circular complex, jointly Gaussian random variables with variance \( N_0 \). In decomposing the continuous-time channel into an infinite set of parallel discrete channels, the Kahruren-Loeve eigenfunctions \( \{ \phi_i(t) \} \) are strictly time limited and approximately band limited.

For the parallel frequency-domain channel model, white noise is defined in an analogous way. We begin with the assumption that the noise at two different frequencies separated by a small bandwidth is approximately statistically independent. From Central Limit Theorem arguments, the output from passing the noise through any linear Doppler shifter can be modeled as a Gaussian random variable. Given such a white Gaussian noise process \( V(f) \) and any set of \( L_2 \) functions \( \{ \phi_i(f) \} \), the outputs \( V_i = \int \phi_i(f)V(f) \, df \) are zero mean circular complex jointly Gaussian random variables with variance \( N_0 \). Here, the Kahruren-Loeve output eigenfunctions \( \{ \phi_i(f) \} \) are strictly limited in frequency and approximately time limited.

Comparison of the continuous-time models of time and frequency shifting, shows that the two techniques differ only in that one is based on the time variable \( t \) and the other is based on frequency variable \( f \). For a completely parallel analysis, the total input power constraint is also redefined as a function of time: \( \sum_{i=1}^{M} \int S_X(t) \, dt \leq P \). Let \( X(f) \) be a band limited replica of a complex Gaussian process with uniform psd \( S_X(t) = \frac{P}{MT} \) over the time interval \( [-T/2, T/2] \). Then \( U(f) \) is a finite bandwidth replica of a complex Gaussian process with psd \( S_U(t) = S_X(t)|H(t)|^2 \), where \( H(t) = \sum_{i=1}^{M} \alpha_i e^{j(i-1)\delta t} \) The mutual information per
3.2 Mutual information calculations

unit time and bandwidth is

\[
I_{FS} = \frac{1}{T} \int_{-T/2}^{T/2} \log \left( 1 + \frac{S_X(t)|H(t)|^2}{N_0} \right) dt
\]

\[
= \frac{1}{T} \int_{-T/2}^{T/2} \log \left( 1 + \frac{P}{TMN_0} \left| \sum_{i=1}^{M} \alpha_i e^{-j(i-1)t\delta} \right|^2 \right) dt. \tag{3.24}
\]

Frequency shift diversity creates a channel that is varying periodically in time. If \( \delta \) is chosen to be an integer multiple of \( \frac{2\pi}{T} \) then the integration in (3.24) is over an integer number of periods and the mutual information is

\[
I_{FS} = \int_{-1/2}^{1/2} \log \left( 1 + \frac{P}{TMN_0} \left| \sum_{i=1}^{M} \alpha_i e^{-j(i-1)2\pi t} \right|^2 \right) dt. \tag{3.25}
\]

Analogous to the time shift diversity case, in the limit of large values of \( \delta \), where \( T \) is assumed to be increasing faster than \( \delta \), the effect of integrating over a non-integer number of periods becomes negligible.

Equation (3.23) requires assuming \( W \) approaches \( \infty \). In particular, analogous to (3.5), we expect the result to be true for \( WT \to \infty \) as

\[
\lim_{WT \to \infty} S_U(t) * W \frac{\sin^2(\pi t W)}{(\pi t)^2} = S_U(t). \tag{3.26}
\]

The frequency shift scheme introduces inter-symbol interference in the frequency domain, thus the received signal bandwidth is larger than the transmitted signal bandwidth. Only in the limit of large \( W \) does the effect of ignoring the received signal outside bandwidth \( W \) become negligible. Comparing this mutual information formula (3.25) to the discrete-time counterpart (2.16), we find the two are equivalent if \( P/T \) is exchanged with \( \mathcal{E}_s \).
Chapter 4

Transmit diversity with channel side information

The transmit diversity schemes examined thus far all assume that the channel coefficients are known at the receiver but unknown at the transmitter. Both transmitter and receiver, however, are assumed to have an accurate statistical model for the channel. In many point-to-point channels the transmitter may obtain additional information about the channel by measuring the reverse channel signal or through a dedicated feedback path from the receiver.

In Section 2.6, it was shown that with complete channel knowledge at both transmitter and receiver, beamforming at the transmitter provides a factor of $M$ increase in SNR over the unconstrained signaling scheme which assumes the channel is unknown to the transmitter. Below we show that beamforming using $M$ receive antennas achieves the same factor of $M$ improvement. By varying the amount and kind of channel side information available to the transmitter, we then evaluate the gap in performance between perfect channel knowledge and no channel knowledge.

The $M$-antenna transmitter diversity channel model was presented in Chapter 2. The complex baseband received signal is $Y = \sum_{i=1}^{M} \alpha_i X_i + V$, where the channel coefficients $\alpha_i$ are independent, zero mean, complex Gaussian random variables. The transmitted symbols are assumed to be power limited to $\sum_i E|X_i|^2 \leq \mathcal{E}_s$. The receiver is assumed to have perfect knowledge of the channel coefficients $\alpha = [\alpha_1 \alpha_2 \ldots \alpha_M]^T$. The transmitter is now assumed to have access to a random variable $S$ correlated with $\alpha$. The variable $S$ represents side information which might be obtained via feedback from the receiver, reverse path signal measurements, or approximate multipath directional information. We consider two special cases: (i) $S$ is a vector $S = [S_1 S_2 \ldots S_M]^T$ with i.i.d. components $S_i$ each correlated with $\alpha_i$, and (ii) $S$ is $N$ bits of information describing a subset of the $M$-dimensional complex space containing $\alpha$. The former case may be viewed as an approximation to the latter when
Transmit diversity with channel side information

$N$ is large.

Since the receiver knows $\alpha$, the multiple antenna channel reduces to an additive white Gaussian noise channel for which a Gaussian input distribution is optimal. We determine the optimal covariance matrix $D_X$ of the Gaussian array input vector, $X = [X_1 X_2 \ldots X_M]^T$, as a function of the side information. We use two measures of performance: average mutual information and average received SNR. The latter is a simplification based on an upper bound of the former.

### 4.1 Receive diversity versus transmit diversity with ideal beamforming

The $M$-antenna receiver diversity channel is shown in Figure 4.1. The channel model is essentially equivalent to the transmitter diversity channel model except that the channel coefficients now correspond to signals transmitted from a common transmit antenna and received by multiple receive antennas. Note that the total received signal power increases linearly with the number of receive antennas, as does the noise power. The receiver is assumed to have perfect knowledge of the channel coefficients.

![Figure 4.1: Channel model for $M$-antenna receive diversity.](image)

Since the channel coefficients are known at the receiver, the $M$ received signals can be combined optimally using Maximal Ratio Combining [14]. This optimal combination is a matched filter; the signal received at antenna $i$ is weighted by $\alpha_i^*$ and the $M$ weighted signals are summed. The mutual information of this channel is easily derived by decomposing the $M$ receive antenna channel into an equivalent AWGN channel [27]. The mutual information
of the channel with $M$ receive antennas is

$$I_{RCV} = \log \left( 1 + \sum_{i=1}^{M} |a_i|^2 \frac{\mathcal{E}_s}{N_0} \right). \quad (4.1)$$

For a multiple transmit antenna diversity system with channel known at the transmitter, the optimal signaling scheme of ideal beamforming is essentially a matched filter at the input. The antenna input signals are weighted so that, after going through the channel, they combine coherently and optimally at the receiver. The mutual information of the $M$-antenna transmit diversity channel using ideal beamforming computed in (2.50) is equal to the mutual information of the $M$-antenna receive diversity channel, (4.1).

### 4.2 Optimal Gaussian inputs for transmit diversity

As mentioned above, the random input vector that maximizes mutual information for the multiple transmit antenna diversity channel, with or without side information, is a zero-mean complex Gaussian array input vector $\mathbf{X}$. The probability density function (pdf) of this $M$-dimensional complex random vector is completely specified by its $2M \times 2M$ real covariance matrix, $D_X$, or alternatively by its complex covariance $\Lambda_X = E[\mathbf{X}\mathbf{X}^H]$ and pseudo-covariance $\tilde{\Lambda}_X = E[\mathbf{X}\mathbf{X}^T]$. Determining the best input for the transmitter array is therefore equivalent to selecting the optimal covariance matrix. If the pseudo-covariance $\tilde{\Lambda}_X$ equals zero, the pdf of $\mathbf{X}$ is specified completely in terms of its complex covariance matrix $\Lambda_X$ and the Gaussian random vector $\mathbf{X}$ is referred to as circular or proper [21,24]. We show that the input vector that maximizes mutual information can be restricted to the class of circular Gaussian random vectors without loss of generality. This result is not surprising: the additive white Gaussian noise is circular, and proper complex Gaussian random vectors maximize entropy [21].

To allow for the possibility of non-circular Gaussian inputs, we rewrite the $M$ dimensional complex vector channel as a $2M$ dimensional real channel:

$$\begin{bmatrix} Y_R \\ Y_I \end{bmatrix} = A^T \begin{bmatrix} X_R \\ X_I \end{bmatrix} + \begin{bmatrix} V_R \\ V_I \end{bmatrix}, \quad (4.2)$$

- 75 -
where

\[
A^T = \begin{bmatrix}
\alpha_R^T & -\alpha_I^T \\
\alpha_I^T & \alpha_R^T
\end{bmatrix},
\tag{4.3}
\]

and each $M$ dimensional complex vector has been written in terms of its $2M$ dimensional real and imaginary parts. The covariance matrix of the input vector $X$ is

\[
D_X = \begin{bmatrix}
\Lambda_R & \Lambda_{R,I} \\
\Lambda_{R,I}^T & \Lambda_I
\end{bmatrix} = \begin{bmatrix}
E[X_RX_R^T] & E[X_RX_I^T] \\
E[X_IX_R^T] & E[X_IX_I^T]
\end{bmatrix}.
\tag{4.4}
\]

The total power constraint becomes a constraint on the trace of the covariance matrix: $\text{tr}(D_X) \leq \mathcal{E}_s$. The mutual information of this channel is

\[
I(X;Y) = h(Y) - h(Y|X) = \frac{1}{2} \log \frac{\det(\frac{N_0}{2}I + A^T D_X A)}{(\frac{N_0}{2})^2}.
\tag{4.6}
\]

The covariance matrix $D_X$ is chosen to maximize mutual information given side information $S$. We show below that this maximization depends on $D_X$ only through the sum $\Lambda_R + \Lambda_I$ and difference $\Lambda_{R,I}^T - \Lambda_{R,I}$. Therefore we can choose $\Lambda_R = \Lambda_I$ and $\Lambda_{R,I}^T = -\Lambda_{R,I}$ without loss of generality. This is equivalent to restricting $X$ to be circular. The proof begins by diagonalizing the $2 \times 2$ matrix $A^T D_X A = Q \Lambda Q^T$ where $Q$ is orthogonal and

\[
\Lambda = \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix}.
\tag{4.7}
\]
4.2 Optimal Gaussian inputs for transmit diversity

Then,

\[
I(X; Y) = \left[ \frac{1}{2} \log \frac{\det(\frac{N_0}{2} I + A^T D_X A)}{(\frac{N_0}{2})^2} \right]
\]

\[
= \frac{1}{2} \log \frac{\det(\frac{N_0}{2} I + Q\Lambda Q^T)}{(\frac{N_0}{2})^2}
\]

\[
= \frac{1}{2} \log \frac{\det Q(\frac{N_0}{2} I + \Lambda)Q^T}{(\frac{N_0}{2})^2}
\]

\[
= \frac{1}{2} \log \frac{(\frac{N_0}{2} + \lambda_1)(\frac{N_0}{2} + \lambda_2)}{(\frac{N_0}{2})^2}.
\]  

This formula is further simplified by recognizing that the eigenvalues of \( A^T D_X A \) have the form \( \lambda_1 = t + w \) and \( \lambda_2 = t - w \), where

\[
t = \frac{1}{2} [\alpha_R^T(\Lambda_R + \Lambda_I)\alpha_R + \alpha_I^T(\Lambda_R + \Lambda_I)\alpha_I + 2\alpha_R^T(\Lambda_{RI}^T - \Lambda_{RI})\alpha_I]
\]  

and

\[
w = \frac{1}{2} [\alpha_R^T(\Lambda_R - \Lambda_I)\alpha_R + \alpha_I^T(\Lambda_R - \Lambda_I)\alpha_I + 2\alpha_R^T(\Lambda_{RI}^T + \Lambda_{RI})\alpha_I]^2
\]

\[-2[\alpha_I^T(\Lambda_R - \Lambda_I)\alpha_R]^2.
\]

Substituting \( \lambda = t \pm w \) into (4.11) yields

\[
I(X; Y) = \frac{1}{2} \log \frac{(\frac{N_0}{2} + t + w)(\frac{N_0}{2} + t - w)}{(\frac{N_0}{2})^2}
\]

\[
\leq \frac{1}{2} \log \frac{(\frac{N_0}{2} + t)^2}{(\frac{N_0}{2})^2},
\]

where equality holds in (4.15) if and only if \( w = 0 \). Therefore, the maximum mutual information is upper bounded by

\[
\max_{D_X : w(D_X) \leq \varepsilon_s} I(X; Y) \leq \max_{D_X : \text{tr}(D_X) \leq \varepsilon_s} \log \left( 1 + \frac{2t}{N_0} \right),
\]  

Note that random variable \( t \) depends on \( D_X \) only through the sum \( \Lambda_R + \Lambda_I \) and difference \( \Lambda_{RI}^T - \Lambda_{RI} \). Hence, the maximization of the upperbound can be limited to consider only circular covariance matrices where \( \Lambda_R = \Lambda_I \) and \( \Lambda_{RI}^T = -\Lambda_{RI} \). Furthermore, this choice of
circular covariance matrices ensures that $w = 0$ and thus this upperbound (4.16) holds with equality.

We have shown that the maximum mutual information can be achieved by a circular Gaussian input vector. A similar proof shows that the SNR, $\frac{\det A^T D_A A}{(N_0/2)^2}$, can always be maximized by a circular Gaussian input vector. Therefore, for both performance measures SNR and mutual information, the input vector can be restricted to circular Gaussian without loss of generality. The pdf of a circular Gaussian random vector is completely specified by its complex covariance matrix $\Lambda_X = E[XX^H]$. Therefore in subsequent analyses we search for the optimal complex covariance matrix $\Lambda_X$ of the circular complex Gaussian array input vector subject to power constraint $\text{tr}(\Lambda_X) \leq \mathcal{E}_s$.

### 4.3 Performance measure: average SNR

Given that the transmitter is equipped with side information $S$ about the channel, we are interested in determining the channel input that maximizes the rate of reliable communication. Maximizing mutual information is an analytically difficult problem since the mutual information is logarithmically related to the channel and input. In this section we hope to get some insight by evaluating a simpler performance measure, the average SNR. For a complex circular Gaussian input, the received SNR is $\alpha^T \Lambda_X \alpha^*/N_0$. By Jensen's Inequality, we can upper bound the average mutual information $E_\alpha[\log(1 + \text{SNR})]$ by $\log(1 + E_\alpha[\text{SNR}])$, where the upper bound holds with equality if and only if the SNR has variance zero. Therefore, maximizing average SNR corresponds to maximizing an upper bound on the rate of reliable communication.

We begin by showing that regardless of the form of side information, $S$, the average SNR is maximized by a rank 1 input covariance matrix that is fully determined by the conditional correlation matrix of the channel vector $\Gamma_{\alpha|S} = E[\alpha^T \alpha^T | S = s]$. Note that $\alpha$ given side information $S$ may no longer have mean zero. We determine the probability distribution of the input vector $X$ that maximizes the average SNR given side information $S = s$:

$$\max_{p(X): \text{tr}(\Lambda_X) \leq \mathcal{E}_s} \frac{E[\alpha^T \Lambda_X \alpha^* | S = s]}{N_0}.$$  

(4.17)
As mentioned above, without loss of generality, \( \mathbf{X} \) can be restricted to circular Gaussian random vectors, thus its pdf \( p(\mathbf{X}) \) is completely specified by its complex covariance matrix \( \Lambda_X \). Expanding \( \Lambda_X \) as \( \Lambda_X = E[\mathbf{X}\mathbf{X}^H] \) and interchanging the order of expectations yields

\[
\max_{p(\mathbf{X}) : \text{tr}(\Lambda_X) \leq \varepsilon_s} \frac{E[\mathbf{X}^H E[\mathbf{\alpha}^*\mathbf{\alpha}^T | S = s] \mathbf{X}]}{N_0}.
\]

(4.18)

Let \( \hat{\lambda} \) be the principal eigenvalue of \( \Gamma_{\alpha|S} \), i.e., the eigenvalue of largest magnitude. The corresponding eigenvector, \( \hat{\mathbf{e}} \) is referred to as the principal eigenvector. The maximizing input distribution points \( \mathbf{X} \) in the direction of the principal eigenvector of the channel correlation matrix: \( \Gamma_{\alpha|S} = E[\mathbf{\alpha}^*\mathbf{\alpha}^T | S = s] \). The maximizing input covariance matrix is therefore a rank 1 matrix: \( \Lambda_X = \hat{\mathbf{e}}\hat{\mathbf{e}}^H\varepsilon_s \).

### 4.3.1 Side information: Random vector \( S \) correlated with \( \alpha \)

Now suppose the side information available at the transmitter is in the form of a zero mean random vector \( S = [S_1 S_2 \ldots S_M]^T \), with i.i.d. components and with each \( S_i \) jointly Gaussian and correlated with \( \alpha_i \). The jointly Gaussian restriction is stronger than necessary as described below. This side information may, for example, be the result of measuring the reverse channel or receiving a noisy version of the forward channel measurement. The degree of correlation between the side information and channel vector is parameterized by the correlation coefficient \( \rho = \frac{E[\alpha_i S_i]}{\sigma_{\alpha} \sigma_S} \) which is in general complex. We denote the variance of \( \alpha_i \) as \( \sigma_{\alpha}^2 \) and the variance of \( S_i \) as \( \sigma_S^2 \).

Straightforward computations of the conditional channel covariance matrix \( \Lambda_{\alpha|S} \), its principal eigenvalue \( \hat{\lambda} \), and its principal eigenvector \( \hat{\mathbf{e}} \) yield:

\[
\Lambda_{\alpha|S} = \sigma_{\alpha}^2 (1 - |\rho|^2) I_M + |\rho|^2 \frac{\sigma_S^2}{\sigma_{\alpha}^2} \mathbf{s}^T \mathbf{s},
\]

(4.19)

\[
\hat{\lambda} = \sigma_{\alpha}^2 (1 - |\rho|^2) + |\rho|^2 \frac{\sigma_S^2}{\sigma_{\alpha}^2} ||\mathbf{s}||^2,
\]

(4.20)

\[
\hat{\mathbf{e}} = \frac{\mathbf{s}^*}{||\mathbf{s}||},
\]

(4.21)

where \( I_M \) is an \( M \)-dimensional identity matrix. The input vector maximizing SNR, \( \hat{\mathbf{X}} \), points in the direction of the principal eigenvector: i.e., \( \hat{\mathbf{X}} = X \hat{\mathbf{e}} \) where \( X \) is a zero-mean
complex Gaussian random variable with variance $E_s$. The resulting maximum average SNR is

$$E[\text{SNR}] = E[\lambda | \mathbf{S} = \mathbf{s}] \frac{E_s}{N_0}$$

$$= \sigma^2_\alpha (1 + |\rho|^2(M - 1)) \frac{E_s}{N_0}.$$  \hspace{1cm} (4.22)

(4.23)

The assumption that $\alpha_i$ and $S_i$ are jointly Gaussian was used to compute the conditional channel correlation matrix:

$$E[\alpha^* \alpha^T | \mathbf{S}] = \Lambda_{\alpha|S} + |E[\alpha|\mathbf{S}]|^2.$$  \hspace{1cm} (4.24)

Identical results are achieved for $\mathbf{S}$ with an arbitrary distribution if the side information results in conditional mean proportional to $\mathbf{s}$, $E[\alpha|\mathbf{S} = \mathbf{s}] = \frac{\alpha}{\sigma_s} \mathbf{s}$, and a scaled identity conditional covariance matrix, $\Lambda_{\alpha|S} = \sigma^2_\alpha (1 - |\rho|^2)I_M$.

Examining (4.23) in the limiting cases shows that when $|\rho| = 1$, i.e., when the transmitter has perfect information about the channel, as expected, the average SNR is $M$ times larger than the case when $\rho = 0$ and the transmitter has no channel information. When $\rho = 0$, the conditional channel correlation matrix is $\Gamma_{\alpha|S} = \sigma^2_\alpha I_M$. Therefore any rank 1 input covariance matrix with energy $E_s$ will yield equivalent performance in terms of average SNR. If the side information $\mathbf{S}$ provides relevant information about the channel, we arrive at the intuitively appealing result that we should point our antenna array in the direction suggested by the side information. The average SNR increases quadratically in $|\rho|$ as the side information becomes more precise, $|\rho| \to 1$.

### 4.3.2 Side information: $N$ bits of channel information

Now suppose the side information at the transmitter consists of $N$ bits of information used to describe the channel vector $\alpha$. This is representative of a scenario in which a dedicated feedback path is available from receiver to transmitter. A feedback path generally requires an allocation of resources that could instead have been used for transmission. Therefore it is of interest to minimize the number of bits relayed to the transmitter. In this section we show that even a small amount of channel knowledge at the transmitter is quite valuable.

---

- 80 -
4.3 Performance measure: average SNR

Given $N$ bits of side information, the transmitter selects the optimal input probability distribution function (pdf) from $2^N$ possible input pdfs. Since the optimal input distribution is circularly Gaussian, these $2^N$ pdfs are completely specified by $2^N$ covariance matrices $\{\Lambda_1, \ldots, \Lambda_{2^N}\}$. The $2^N$ covariance matrices are selected to best represent the space of all possible received channel vectors $\alpha$. The expected SNR (or mutual information) is maximized by associating each received channel vector $\alpha$ with a specific covariance matrix. Thus the problem is in many ways similar to vector quantization. The quantizer divides the space of all possible channel vectors into $2^N$ regions and specifies the input covariance matrix associated with each region. Both the regions and the covariance matrices must be selected to maximize expected SNR.

A local maximum can be found by using a slightly modified Lloyd iterative algorithm [11]. The algorithm begins by dividing the coefficient vector space into $2^N$ regions, $\{R_1, R_2, \ldots, R_{2^N}\}$ and then determines the input covariance matrix, $\Lambda_i$, that maximizes average SNR for each region, $R_i$. Given the set of input covariance matrices $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_{2^N}\}$, a new partition of the coefficient vector space is formed by associating each possible $\alpha$ with the covariance matrix that produces the largest SNR:

$$R_i = \{\alpha : \alpha^T \Lambda_i \alpha^* > \alpha^T \Lambda_j \alpha^*; \text{for all } j \neq i\} \quad (4.25)$$

where ties (vectors $\alpha$ which produce equal SNR for multiple $\Lambda_i$) are placed in the region with the lowest index. Given the new partition, the algorithm again determines the optimal set of input covariance matrices. Each iteration of the algorithm is guaranteed to increase or leave unchanged the average SNR. The algorithm repeats until the partition regions stabilize or the average SNR converges. The local maximum achieved is of course dependent on the initial partition. We investigate both random initial partitions as well as several intuitive initial configurations in an attempt to achieve performance close to the globally optimal solution.

The unknown channel vector $\alpha$ has $2M$ components, $M$ magnitudes and $M$ phases. Since the input is power constrained, only the relative magnitudes of the vector components are relevant in determining the optimal input distribution. The number of degrees of freedom is further reduced by one, since only the phases relative to some reference phase are of consequence. If we assume the first channel coefficient is real and positive (zero phase), the problem reduces to a $2M-1$ dimensional space. The quantization problem may be visualized as quantizing the surface of a $2M-2$ dimensional half-sphere in a $2M-1$ dimensional
space. Unfortunately, the distance rule for partitioning the space (4.25) does not lead to easily visualized regions on the surface of the sphere.

4.3.2.1 Results for $M = 2$ transmit antennas

We begin by analyzing the case of $M = 2$ transmit antennas for which the channel vector has only $2M - 2 = 2$ relevant degrees of freedom. Thus $\alpha$ can be completely described by two angles: the relative magnitude, $\phi = \tan^{-1} \frac{|\alpha_2|}{|\alpha_1|}$ and the relative phase $\theta = \angle \alpha_2 - \angle \alpha_1$. For this case the problem reduces to finding the optimal partition and set of corresponding input covariance matrices on a two dimensional space. The two dimensional problem is computationally tractable and, in this format, the resulting locally optimal partition of the space is easily visualized. We denote by $\gamma$ the increase in average SNR provided by the channel side information:

$$\gamma = \frac{\text{average SNR given } N \text{ bits of side information}}{\text{average SNR without side information}}.$$  

(4.26)

Recall that if the transmitter has complete knowledge of the channel coefficients, the optimal beamforming scheme leads to a factor $\gamma = 2$ (or equivalently $M$) improvement in average SNR.

Our simulations of the Lloyd Algorithm begin by discretizing the two dimensional $\theta-\phi$ space corresponding to the channel vectors $\alpha$. The covariance matrix associated with each initial region is determined by maximizing the average SNR on this space. A new set of regions is formed using (4.25) and associating points in the discretized $\theta-\phi$ sample space with the appropriate covariance matrix. For these new regions, covariance matrices that maximize expected SNR are determined. These iterations are repeated until the partition regions (and associated covariance matrices) converge.

Scalar quantizer based initial partition regions

We begin by exploring quantizers resulting from the Lloyd algorithm using intuitive initial partition regions based on scalar quantizers for $\phi$ and $\theta$. With $N = 1$ bit of side information, two reasonable strategies for dividing up the channel coefficient vector space are (i) specifying which antenna coefficient has the largest magnitude or (ii) providing one bit of phase information which signifies whether the relative phase of the antenna coefficients
4.3 Performance measure: average SNR

is between 0 and $\pi$ or between $-\pi$ and 0. In the first case, the transmitter uses only the antenna with the largest magnitude. In the second case, the transmitter sends signals with the appropriate relative phase of $\pm \frac{\pi}{2}$ on the two antennas. Both quantization schemes are local maxima of the iterative Lloyd algorithm. Interestingly, both forms of side information provide equal benefit. Computing the expected SNR shows both schemes lead yield a performance improvement $\gamma = 1.5$ times larger than the optimal scheme when no channel information is available to the transmitter.

For $N > 1$, an initial estimate of the vector space partition is formed by choosing uniform scalar quantizers for each of the two vector components. This corresponds to dividing the space into rectangular partition regions. When $N = 2$, for example, we can use one bit to describe the relative magnitude and the other bit to specify the relative phase. Alternatively, we could use both bits for phase or both bits for magnitude. Given only magnitude information, the optimal strategy is for the transmitter to use the higher gain antenna. Therefore, in the two transmit antenna system, multiple bits of magnitude information are not useful, since only one bit is required to specify the antenna with larger gain. Additional magnitude information only becomes useful if some phase information is available. Given any phase information, the optimal input corresponds to “pointing” the array in the proper direction. Using two bits for phase or one bit each for phase and magnitude both result in an increase in average SNR by a factor $\gamma = 1.707$. These uniform partitioning schemes are both locally optimal solutions of the Lloyd Algorithm. However, we show that these local maxima are not global maxima in the following section.

Given the best initial partition for $N$ bits of side information, one might guess that the best initial partition for the case of $N+1$ bits of side information can be formed by using the $N$ bit partition and examining the two cases of adding an additional bit of phase information or adding an additional bit of magnitude information. The results of using 1 to 6 bits of side information with various initial partition configurations are shown in Table 4.1. The sizes of the final partition regions differ from the initial partition, but remain rectangular. The final partition and increase in average SNR correspond to local maxima of the Lloyd Algorithm.

Random initial partition regions
We now examine the results of randomly generating initial partition regions for the Lloyd vector quantization algorithm. For $N$ bits of side information, the two-dimensional space is divided into $2^N$ random connected sets. For each $N$, the Lloyd algorithm was used with several randomly generated initial partitions. Each initial partition led to the same increase
Transmit diversity with channel side information

Table 4.1: Increase in average SNR given $N$ bits of side information using scalar quantizer based initial partition regions for the Lloyd algorithm.

<table>
<thead>
<tr>
<th>$N$</th>
<th># of phase bits</th>
<th># of magnitude bits</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1.500</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1.500</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1.707</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1.707</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1.765</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1.866</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1.766</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1.916</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1.915</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1.927</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1.964</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1.930</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1.976</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1.978</td>
</tr>
<tr>
<td>$\infty$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>2.000</td>
</tr>
</tbody>
</table>

in expected SNR, $\gamma$, and furthermore each $\gamma$ was greater than or equal to the $\gamma$ achieved with the scalar quantizer based initial regions suggested in the section above. When scalar quantizer based partition regions are used to initialize the Lloyd algorithm, the algorithm converges to a local maximum more rapidly, but randomly generated partition regions yield better performance in terms of expected SNR.

For $N = 1$, all randomly generated initial partition regions yielded an increase in expected SNR of $\gamma = 1.5$ after the Lloyd algorithm was applied. The resulting quantization regions were often equivalent to the one bit of phase information scheme described above. Other local maxima solutions corresponded to a sinusoidal division of the $\theta$-$\phi$ space.

The resulting increase in expected SNR from $N$ bits of side information is shown in Table 4.2. For $N > 1$, the randomly generated initial partition regions led to better quantizers in terms of maximizing expected SNR. Each additional bit of side information roughly halves the distance to $\gamma = 2$, the increase in expected SNR achieved with complete side information. In other words, the gap in performance between perfect and zero side
4.3 Performance measure: average SNR

Table 4.2: Increase in average SNR given $N$ bits of side information using randomly generated initial partition regions for the Lloyd algorithm.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1.500</td>
</tr>
<tr>
<td>2</td>
<td>1.745</td>
</tr>
<tr>
<td>3</td>
<td>1.871</td>
</tr>
<tr>
<td>4</td>
<td>1.936</td>
</tr>
<tr>
<td>5</td>
<td>1.968</td>
</tr>
<tr>
<td>6</td>
<td>1.984</td>
</tr>
<tr>
<td>$\infty$</td>
<td>2.000</td>
</tr>
</tbody>
</table>

information falls exponentially ($\sim 2^{-N}$) in the number of bits of side information as shown in Figure 4.2. The resulting vector quantizer for $N = 6$ is shown in Figure 4.3. As mentioned above, the Lloyd Algorithm was implemented on a discretized $\theta$-$\phi$ sample space. The jagged region boundaries are an artifact of the approximations resulting from this discretization. In this figure, the quantization regions are smaller in the middle than near the edges. This can be partially explained by examining the pdf's of $\theta$ and $\phi$ which show that points in the middle of the $\theta$-$\phi$ space as drawn in Figure 4.3 are more probable than those near the edges.

Figure 4.2: Gap between perfect side information and $N$ bits of side information using randomly generated initial partition regions.
Transmit diversity with channel side information

Figure 4.3: Quantization regions and codebooks resulting from the Lloyd algorithm using a random initial partition. The jagged region boundaries are an artifact of the approximations resulting from the discretization of the $\theta$-$\phi$ sample space.

Specifically, the pdf of $\theta$ is uniform from $-\pi$ to $\pi$ while the pdf of $\phi$ is

$$f_\phi(\phi) = \frac{2\tan(\phi)}{\tan^2(\phi) + 1},$$

and is greatest at $\phi = \pi/4$.

4.3.2.2 Results for $M > 2$ transmit antennas

If $M$ transmit antennas are used, the unknown channel vector has $2M - 2$ relevant degrees of freedom: $M - 1$ relative phases and $M - 1$ relative magnitudes. This $2M - 2$ dimensional space can be quantized into $2^N$ regions with $2^N$ associated input covariance matrices using the ideas above. However, the iterative quantization and maximization process becomes more complex and more computationally intensive as the number of antennas increases, since the size of the quantization space increases exponentially with the number of antennas.

We do, however, discuss in detail two special cases of side information for $M > 2$ antenna systems. The first method, selection diversity for $M$ transmit antennas, requires $\log_2 M$ bits to specify the antenna associated with the largest gain. The transmitter then uses only the largest gain antenna. The pdf of the antenna gain of the largest gain antenna is easy to calculate as the maximum of $M$ Rayleigh gains, $\max\{ |\alpha_1|, |\alpha_2|, \ldots, |\alpha_M| \}$. The

- 86 -
resultant increase in average SNR provided by selection diversity is [14]

$$\gamma = \sum_{k=1}^{M} \frac{1}{k} \approx \log M. \quad (4.28)$$

The number of bits of side information for selection diversity is logarithmic in the number of transmit antennas. The increase in average SNR provided by this scheme is logarithmic as well. Note that as $M$ increases, this scheme achieves a decreasing fraction of the $M$-fold improvement possible with complete channel knowledge. This is not surprising since the number of bits used to describe the channel, $\log_2 M$, is also a decreasing fraction of the total number of degrees of freedom $2M - 2$.

An alternative scheme is to use $M - 1$ bits of side information to specify one bit of information about each of the relative phases of the channel coefficients and then apply beamforming. Earlier we showed that the optimal input covariance matrix points $X$ in the direction of the principal eigenvector of the conditional channel correlation matrix $E[\alpha^* \alpha^T | S]$. The increase in average SNR is determined by the principal eigenvalue of this correlation matrix. The $M - 1$ bits of phase information specify whether each of the relative angles $\angle \alpha_i - \angle \alpha_1$, for $i = 2 \ldots M$, is uniformly distributed between 0 and $\pi$ or between $-\pi$ and 0. The conditional correlation matrix $E[\alpha^* \alpha^T | S]$ depends on terms of the form $E[|\alpha_i|^2 | S]$, $E[\alpha_i \alpha_j^* | S]$, and $E[\alpha_i \alpha_j^* | S]$ for $j \neq i \neq 1$. The pdf of the angles $\Psi = \angle \alpha_i - \angle \alpha_j$ conditional on $S$ is either

$$f_\Psi(\psi) = \frac{1}{\pi} - \frac{1}{\pi^2} |\psi| \quad (4.29)$$

or

$$f_\Psi(\psi) = \frac{1}{\pi^2} |\psi| \quad (4.30)$$

for $-\pi \leq \psi \leq \pi$, depending on whether, given the side information, $\angle \alpha_i$ and $\angle \alpha_j$ are uniformly distributed over the same interval or over differing intervals. The expected value of $\Psi$ is the of course the same in either case. Assume for simplicity $E[|\alpha_i|^2] = 1$; then $E[|\alpha_i|^2 | S] = \frac{1}{2} \sqrt{-1}$, and $E[\alpha_i \alpha_j^* | S] = \frac{1}{\pi}$. The principal eigenvalue of the conditional
correlation matrix and equivalently the increase in average SNR for the $M$ antenna case is

$$\gamma = 1 + \frac{M - 1}{2\pi} + \frac{\sqrt{(M - 1)^2\pi^2 + (M - 2)^2}}{2\pi}. \tag{4.31}$$

This increase is linear in the number of antennas, as is the increase in the number of bits of side information.

### 4.3.3 Quantization and rate distortion theory

Using rate distortion theory, we can gain insight on achievable performance for large $N$ and large $M$ without resorting to vector quantization design.

Assume $R$ bits are used to describe each component of $\alpha$ using a rate-distortion codebook that minimizes mean squared error:

$$D = \frac{1}{2M} \sum_{i=1}^{M} E[|\alpha_i - \hat{\alpha}_i|^2]. \tag{4.32}$$

Since the components of $\alpha$ are independent identically distributed complex Gaussians with variance $\frac{\sigma^2}{2}$ per real and imaginary part, the distortion $D$ is lower-bounded by

$$D > D(R) = \frac{\sigma^2}{2} 2^{-R}. \tag{4.33}$$

This lower bound can be approached arbitrarily closely as the number of antennas $M$ (as well as the number of descriptive bits $N = RM$) approaches $\infty$.

If the number of bits of side information $N$ is large, $\alpha$ may be modeled as conditionally Gaussian given side information $S$:

$$\alpha = \hat{\alpha} + \epsilon, \tag{4.34}$$

where $\hat{\alpha} = E[\alpha|S]$ and $\epsilon = [\epsilon_1 \epsilon_2 \ldots \epsilon_M]^T$ is a zero mean complex Gaussian vector with variance $\frac{\sigma^2}{4}$ per each real and imaginary component. In terms of (4.33),

$$\sigma_\epsilon^2 \approx \sigma_\alpha^2 2^{-R} = \sigma_\alpha^2 2^{-N/M}. \tag{4.35}$$
4.4 Performance measure: average mutual information

The above description of the side information is equivalent to assuming $\alpha_i$ and $S_i$ are jointly Gaussian with correlation $\rho$ as defined in Section 4.3.1. The two descriptions are related by

$$\sigma_i^2 = \sigma_\alpha^2 (1 - |\rho|^2).$$

(4.36)

Therefore, for large $M$, from (4.23) the increase in expected SNR achieved by using $N$ bits to describe the channel coefficients to minimize mean square error is roughly

$$\gamma = \sigma_\alpha^2 [1 + (1 - 2^{-N/M})(M - 1)].$$

(4.37)

The factor of $M$ gap between perfect and zero side information decreases exponentially with $N$. Recall from Section 4.3.2.1 that vector quantization for $M = 2$ also reduced the gap between perfect and zero side information exponentially with $N$. In particular, the reduction was proportional to $2^{-N}$. The estimate (4.37) for $M = 2$ instead gives a reduction proportional to $2^{-N/2}$. In this section, each of the $2M$ real and imaginary channel coefficient vector components were quantized uniformly to minimize mean square error. This result is overly pessimistic for two reasons: first, minimum MSE is not our true performance measure and second, from our previous results we know that to maximize average SNR, it is only necessary to provide a description of $2M - 2$ degrees of freedom. Thus, using $N$ bits to uniformly describe $2M$ components is wasteful. We expect the difference in performance of the two approaches to become small as $M$ becomes large.

An alternative (and more efficient) approach is to use $N$ bits to quantize the $M - 1$ dimensional complex vector $[\frac{\alpha_2}{\alpha_1}, \frac{\alpha_3}{\alpha_1}, \ldots, \frac{\alpha_M}{\alpha_1}]$ to minimize mean square error. Unfortunately, the conditional expected SNR is difficult to evaluate with this form of side information.

4.4 Performance measure: average mutual information

In the absence of decoding delay constraints, the average mutual information is the Shannon capacity in the usual sense; it is the maximum rate of reliable communication without errors. In this section, the optimal input distribution is chosen to maximize average mutual information. It was shown in Section 4.2 that we can consider only circular Gaussian input distributions without loss of generality. Thus specifying the optimal input distribution to
maximize average mutual information reduces to determining the optimal complex covariance matrix $\Lambda_X = E[XX^H]$ under energy constraint, $\text{tr}(\Lambda_X) \leq \mathcal{E}_s$.

We begin by examining the two limiting cases of complete channel knowledge at the transmitter, $S = \alpha$, and no channel knowledge at the transmitter, $S = \emptyset$. We then consider $S$ to be a complex Gaussian vector correlated with $\alpha$. Finally, we analyze the case where $S$ is $N$ bits of side information used to describe $\alpha$.

### 4.4.1 Limiting cases: completely known ($S = \alpha$) and completely unknown ($S = \emptyset$) channel

We choose the input distribution to maximize average mutual information conditional on side information $S$:

$$\max_{\Lambda_X: \text{tr}(\Lambda_X) \leq \mathcal{E}_s} E_{\alpha|S} \left[ \log \left( 1 + \frac{\alpha^T \Lambda_X \alpha^*}{N_0} \right) \right].$$

Diagonalizing the complex covariance matrix $\Lambda_X = UDU^H$ yields

$$\max_{U, D: \text{tr}(D) \leq \mathcal{E}_s} E_{\alpha|S} \left[ \log \left( 1 + \frac{\alpha^T UDU^H \alpha^*}{N_0} \right) \right].$$

Thus, choosing the optimal covariance matrix is equivalent to choosing a unitary matrix $U$ and diagonal matrix $D$. When the channel coefficient vector $\alpha$ is completely known at the transmitter, we showed in Section 2.6.1 that the mutual information given any particular $\alpha$ is maximized by beamforming. Therefore, for $\alpha$ known at the transmitter, beamforming maximizes the average mutual information as well, where the average is over the set of all possible coefficient vectors $\alpha$. Recall that beamforming corresponds to setting

$$X = \frac{\alpha^*}{||\alpha||} X,$$

where $X$ is a zero-mean complex Gaussian random variable with variance $E|X|^2 = \mathcal{E}_s$. Beamforming is equivalent to choosing unitary matrix $U$ to align the coefficient vector into a vector with one non-zero component, $U^H \alpha = [||\alpha||0 \ldots 0]^T$, and choosing the diagonal matrix $D$ to have only a single component $\mathcal{E}_s$ in the upper left hand corner. The resulting covariance matrix $\Lambda_X$ has rank 1.
4.4 Performance measure: average mutual information

In the absence of channel knowledge, \( \alpha \) is a zero-mean complex Gaussian random vector. Multiplying a zero-mean complex Gaussian random vector by a unitary matrix does not effect its distribution. Therefore the maximization in (4.39) only depends on the diagonal matrix \( D \)

\[
\max_{D : \text{tr}(D) \leq \mathcal{E}_s} E_{\alpha|S} \left[ \log \left( 1 + \frac{\alpha^T D \alpha^*}{N_0} \right) \right].
\]  

(4.41)

Let \( \lambda_1, \ldots, \lambda_M \) be the diagonal components of \( D \). The maximization above can be rewritten as

\[
\max_{\{\lambda_i\} : \sum_{i=1}^M \lambda_i = 1} E_{\alpha|S} \left[ \log \left( 1 + \sum_{i=1}^M \lambda_i |\alpha_i|^2 \frac{\mathcal{E}_s}{N_0} \right) \right].
\]  

(4.42)

Since the expected mutual information is concave in \( \lambda_i \), setting the derivatives with respect to \( \lambda_i \) to zero shows that the maximum is achieved if \( \lambda_i = \frac{1}{M} \) for all \( i \). Therefore, in the absence of side information, the optimal \( \Lambda_X \) is a scaled identity matrix. As expected, this covariance matrix corresponds to the unconstrained diversity signaling method analyzed in the previous chapters.

4.4.2 Side information: Random vector \( S \) correlated with \( \alpha \)

Now suppose the transmitter is provided with side information \( S \) consisting of a zero-mean Gaussian random vector correlated with the coefficient vector \( \alpha \). As above, the variance of \( \alpha_i \) is \( \sigma_\alpha^2 \), the variance of \( S_i \) is \( \sigma_S^2 \), and the correlation coefficient is \( \rho = \frac{E[\alpha_i S_i]}{\sigma_\alpha \sigma_S} \). The conditional random vector \( \alpha \) given \( S = s \) is complex Gaussian with mean \( \mu_\alpha s \) and covariance \( \sigma_\alpha^2 (1 - |\rho|^2) I_M \). To facilitate analysis, we rewrite \( \alpha \) as the sum of a constant \( \hat{\alpha} = \rho_{\alpha s} s \) plus a zero-mean complex Gaussian random vector \( \epsilon = [\epsilon_1 \epsilon_2 \ldots \epsilon_M]^T \) with variance \( \sigma_\epsilon^2 = \sigma_\alpha^2 (1 - |\rho|^2) \). In this form, the side information is suggestive of a noisy measurement of the true channel vector \( \alpha \). Rewriting the maximization problem in terms of \( \alpha = \hat{\alpha} + \epsilon \) yields

\[
\max_{u, D : \text{tr}(D) \leq \mathcal{E}_s} E \left[ \log \left( 1 + \frac{(\hat{\alpha} + \epsilon)^T U D U^H (\hat{\alpha} + \epsilon)^*}{N_0} \right) \right].
\]  

(4.43)

As \( \sigma_\epsilon^2 \) goes to zero, the channel estimate available at the transmitter becomes better and better, and thus we expect the optimal input covariance matrix to tend toward a rank
Transmit diversity with channel side information

1 matrix with principal direction $\hat{\alpha}^\ast$. If the channel estimate is poor and $\sigma^2_\epsilon$ is large, we expect the optimal input covariance matrix to tend toward a scaled identity matrix as is appropriate when no channel information is available. We present specific results for the case of $M = 2$ transmit antennas.

Expanding the matrix terms in (4.43) and using the fact that $U_e$ has the same distribution as $\epsilon$, since $\epsilon$ is zero-mean complex Gaussian, yields

$$\max_{u, D, \text{tr}(D) \leq \mathcal{E}_s} E \left[ \log \left( 1 + \frac{\hat{\alpha}^T U D U^H \hat{\alpha}^* + \epsilon^T D \epsilon^* + \hat{\alpha}^T U D \epsilon^* + \epsilon^T D U^H \epsilon^*}{\mathcal{N}_0} \right) \right].$$ (4.44)

Without loss of generality, $\hat{\alpha}^T U$ can be written as $[\sqrt{\beta}||\hat{\alpha}|| \quad \sqrt{1-\beta}||\hat{\alpha}||]^T$, where $0 \leq \beta \leq 1$, since $\hat{\alpha}^T U D U^H \hat{\alpha}^*$ is independent of the phases of the components of $\hat{\alpha}^T U$ and since the probability distributions of $\hat{\alpha}^T U D \epsilon^*$ and $\epsilon^T D U^H \epsilon^*$ are also independent of the phase components of $\hat{\alpha}^T U$. The $2 \times 2$ diagonal matrix, $D$, can be parameterized in terms of a single parameter $\lambda \in [0, 1]$:

$$D = \begin{bmatrix} \lambda & 0 \\ 0 & 1-\lambda \end{bmatrix} \mathcal{E}_s.$$ (4.45)

The optimization problem reduces to choosing $\lambda$ and $\beta$ to maximize the average mutual information:

$$\max_{0 \leq \lambda, \beta \leq 1} E_{(\epsilon_1, \epsilon_2)} [\log(1 + \{[\lambda \beta + (1-\lambda)(1-\beta)]||\hat{\alpha}||^2 + \lambda |\epsilon_1|^2 + (1-\lambda) |\epsilon_2|^2 \\
+ 2\sqrt{\beta} \lambda ||\hat{\alpha}|| |\epsilon_{1R} + 2\sqrt{1-\beta}(1-\lambda)||\hat{\alpha}|| |\epsilon_{2R}|/\mathcal{E}_s)] ,$$ (4.46)

where $\epsilon_{iR}$ denotes the real part of $\epsilon_i$. Note, the average mutual information as a function of $\lambda$ and $\beta$ is symmetric around the line $\lambda = \beta$. For example, if $\lambda = 1, \beta = 1$ is a maximum, then $\lambda = 0, \beta = 0$ is a maximum also.

We assume $\beta = 1$ and maximize the average mutual information as a function of $\lambda$. Choosing $\beta = 1$ corresponds to pointing the covariance matrix in the direction of $\hat{\alpha}$. Although we have not proven this value for $\beta$ always maximizes average mutual information, Monte Carlo simulations of the average mutual information as a function of $\lambda$ and $\beta$ show that the maximum occurs at $\beta = 1$ with different values for $\lambda$. Geometrically, $\lambda$ specifies the relative widths of the axes of an ellipse. The shape of this ellipse corresponds to the energy
4.4 Performance measure: average mutual information

spread in directions other than $\hat{\alpha}^*$. When $\lambda = 1$, the rank of the covariance matrix is one and the ellipse becomes a line. We determine a sufficient condition on the variance of our estimate, $\sigma_e^2$ that ensure a rank 1 covariance matrix.

We begin by introducing some simplifying notation. Let $R = ||\hat{\alpha}||^2$ and $\nu = E_s/N_0$. With $\beta = 1$, the maximum average mutual information is

$$\max_{0 \leq \lambda \leq 1} E[\log(1 + \lambda A + (1 - \lambda)B)],$$

where

$$A = \nu(\epsilon_{1R} + \sqrt{R})^2 + \nu \epsilon_{1I}^2$$

(4.48)

and

$$B = \nu(\epsilon_{2R}^2 + \epsilon_{2I}^2).$$

(4.49)

Finally, we define $J$ as the average mutual information:

$$J = E[\log(1 + \lambda A + (1 - \lambda)B)].$$

(4.50)

We first show that $J$ is strictly concave in $\lambda$ since the second derivative is strictly negative:

$$\frac{d^2J}{d\lambda} = E \left[ \frac{-(A - B)^2}{(1 + \lambda A + (1 - \lambda)B)^2} \right]$$

(4.51)

$$< 0.$$  

(4.52)

The inequality is strict because $A$ and $B$ are independent random variables with nonzero variance. Consequently, the maximum of $J$ occurs at $\lambda = 1$ if the first derivative $\frac{dJ}{d\lambda} \geq 0$ at
Transmit diversity with channel side information

\[ \lambda = 1. \text{ Evaluating the first derivative at } \lambda = 1 \text{ we find} \]

\[
\left. \frac{dJ}{d\lambda} \right|_{\lambda=1} = E \left[ \frac{A - B}{1 + A} \right] \tag{4.53}
\]

\[
= E \left[ 1 - \frac{1 + B}{1 + A} \right] \tag{4.54}
\]

\[
= 1 - E[1 + B]E \left[ \frac{1}{1 + A} \right]. \tag{4.55}
\]

Thus, the maximum occurs at \( \lambda = 1 \) if and only if

\[
E[1 + B]E \left[ \frac{1}{1 + A} \right] \leq 1. \tag{4.56}
\]

Although we are unable to compute \( E \left[ \frac{1}{1 + A} \right] \), we can upper bound this quantity using methods suggested in [30, Lemma 1]. Because the third derivative of \( \frac{1}{1 + A} \) is negative, the maximum of \( E \left[ \frac{1}{1 + A} \right] \) over all pdfs \( A \) with \( A \geq 0 \), mean \( E[A] \), and second moment \( E[A^2] \) is achieved by a two mass point distribution:

\[
p_A(x) = \begin{cases} 
1 - \frac{(E[A])^2}{E[A^2]} & x = 0, \\
\frac{(E[A])^2}{E[A^2]} & x = \frac{E[A^2]}{E[A]}.
\end{cases} \tag{4.57}
\]

Therefore,

\[
E \left[ \frac{1}{1 + A} \right] \leq \frac{(E[A])^2}{E[A^2]} \left( \frac{1}{1 + \frac{E[A^2]}{E[A]}} \right) + \left( 1 - \frac{(E[A])^2}{E[A^2]} \right) \tag{4.58}
\]

\[
= \frac{E[A] + E[A^2] - (E[A])^2}{E[A] + E[A^2]} \tag{4.59}
\]

Using this upper bound in (4.56) shows that a sufficient condition for achieving the maximum average mutual information with \( \lambda = 1 \) is

\[
E[1 + B] \left( \frac{E[A] + E[A^2] - (E[A])^2}{E[A] + E[A^2]} \right) \leq 1. \tag{4.60}
\]
4.4 Performance measure: average mutual information

Simple algebraic manipulation leads to the following equivalent condition:

\[
\frac{E[A]}{1 + \frac{\sigma_A^2}{E[A]}} \geq E[B],
\]

(4.61)

where the variance of \( A \) is \( \sigma_A^2 = E[A^2] - (E[A])^2 \). In other words, if the mean of \( A \) is greater than the mean of \( B \) (\( R > 0 \)) and if the variance of \( A \) is not too large, a rank 1 covariance matrix maximizes average mutual information.

Substituting the means and variances of \( A \) and \( B \) into (4.61) yields the sufficient condition in terms of \( \sigma_c^2 \). From (4.48) and (4.49), we know that \( A \) is a non-central \( \chi^2 \) distribution with two degrees of freedom. Its mean and variance [16] are

\[
E[A] = \nu(R + \sigma_c^2)
\]

(4.62)

\[
\sigma_A^2 = \nu^2 \sigma_c^2(\sigma_c^2 + 2R).
\]

(4.63)

Random variable \( B \) is also \( \chi^2 \) with two degrees of freedom. The expected value of \( B \) is

\[
E[B] = \nu \sigma_c^2.
\]

(4.64)

Without loss of generality we can assume \( E||\alpha||^2 = 1 \) since (if non-zero) values of this moment can be incorporated into the signal-to-noise ratio \( \nu \). Since \( E||\alpha||^2 = R + 2\sigma_c^2 \), we can eliminate \( R \) from our equations using \( R = 1 - 2\sigma_c^2 \) and restricting \( \sigma_c^2 \) to \( 0 \leq \sigma_c^2 \leq 1/2 \). Following some algebraic manipulation, we arrive at the following sufficient condition for ensuring the maximum average mutual information occurs at \( \lambda = 1 \):

\[
3\nu(\sigma_c^2)^3 + 2(1 - \nu)(\sigma_c^2)^2 - 3\sigma_c^2 + 1 \geq 0.
\]

(4.65)

For \( \nu > 0 \), this third order polynomial is positive between 0 and its smallest positive root, \( \hat{\sigma}_c^2 \). For \( 0 \leq \sigma_c^2 \leq \hat{\sigma}_c^2 \), the sufficient condition above is satisfied: a rank 1 covariance matrix maximizes average mutual information. In Figure 4.4, we plot \( \hat{\sigma}_c^2 \) as a function of \( \nu = E||\alpha||^2 \mathcal{E}_s/N_0 \).
Figure 4.4: Sufficient condition for ensuring that the input covariance matrix that maximizes average mutual information has rank 1: when the side information describes each complex channel coefficient to within a Gaussian error with variance $\sigma_e^2$, it is sufficient to require that $\sigma_e^2 \leq \delta_e^2$. Above we plot $\delta_e^2$ as a function of $\nu = E||\alpha||^2E_s/N_0$.

4.4.3 Side information: $N$ bits of channel information

We now consider the problem of using $N$ bits of side information to maximize mutual information (rather than SNR). As in Section 4.3.2, we use the $N$ bits of side information to divide the set of possible coefficient vectors $\alpha$ into $2^N$ regions, and, for each region, we select an input covariance matrix $\Lambda_X$ to maximize the average mutual information.

Unlike the case of average SNR, for average mutual information the covariance matrix assigned to each region may need rank greater than 1. There appears to be no simple expression for expected mutual information as a function of region shape and the $M^2 - 1$ unknown parameters of $\Lambda_X$, nor does there appear a way to compute the maximizing $\Lambda_X$ directly. Numerically, a Monte Carlo simulation for computing the average mutual information could be combined with a search for the maximizing covariance matrix over the $M^2 - 1$ dimensional space, but the iterative procedure of choosing regions and then optimal covariance matrices becomes computationally prohibitive as $M$ and $N$ increase. We do not pursue this idea here.

Some analytic results can however be obtained for the case of $M = 2$ antennas. The
4.4 **Performance measure: average mutual information**

The covariance matrix can be parameterized as

$$\Lambda_X = \left[ \begin{array}{cc} \frac{\lambda}{\sqrt{\lambda(1-\lambda)}} e^{j\vartheta} & \frac{\sqrt{\lambda(1-\lambda)} re^{-j\vartheta}}{(1-\lambda)} \\ \sqrt{\lambda(1-\lambda)} re^{j\vartheta} & (1-\lambda) \end{array} \right] \mathcal{E}_s,$$

(4.66)

where $0 \leq \lambda, r \leq 1$ and $-\pi \leq \vartheta \leq \pi$. Recall that for $M = 2$, the space of all possible $\alpha$ can be completely described in terms of a relative magnitude and a relative phase. Suppose $N = 1$ bit of side of information is available to the transmitter. Two natural possibilities for the side information are (i) designating the antenna with the larger gain and (ii) specifying whether the relative phase is between $-\pi$ and 0 or between 0 and $\pi$. We show that rank 1 covariance matrices are optimal in both cases.

Expanding $\alpha^T \Lambda_X \alpha^*$ yields the maximum average mutual information

$$\max_{\Lambda_X: \text{tr}(\Lambda_X) \leq \mathcal{E}_s} \mathbb{E}[\log(1 + \alpha^T \Lambda_X \alpha^* \frac{\mathcal{E}_s}{N_0})] = \max_{0 \leq \lambda, r \leq 1; \vartheta} \mathbb{E}_{\alpha_1|S}[\log(1 + \frac{\lambda|\alpha_1|^2 + (1-\lambda)|\alpha_2|^2 + 2|\alpha_1||\alpha_2|r \sqrt{\lambda(1-\lambda)} \cos(\varphi)}{N_0})],$$

(4.67)

where $\varphi = \angle \alpha_1 - \angle \alpha_2 - \vartheta$. If $S$ is chosen to specify the larger gain antenna, $\varphi$ is uniformly distributed between $-\pi$ and $\pi$ and is independent of $|\alpha_1|, |\alpha_2|$. By Jensen’s inequality, the average over $\varphi$ in (4.67) is maximized by choosing $r = 0$. The problem thus reduces to

$$\max_{0 \leq \lambda \leq 1} \mathbb{E}_{\{\alpha_1, |\alpha_2|\}|S}[\log(1 + \frac{\lambda|\alpha_1|^2 + (1-\lambda)|\alpha_2|^2}{N_0})].$$

(4.68)

If $S$ specifies $|\alpha_1| > |\alpha_2|$, the average mutual information is monotonically increasing with $\lambda$ and is thus maximized by $\lambda = 1$. For $|\alpha_1| < |\alpha_2|$, the maximizing $\lambda$ is 0. This solution is a stationary point of the Lloyd iteration since the optimal partition associated with the two covariance matrices above divides the coefficient vector space according to the largest gain antenna.

If instead $S$ is used to provide one bit of information about the relative phase, $\vartheta$ can be chosen to keep $\varphi$ between $-\pi/2$ and $\pi/2$, thus the cosine term in (4.67) will always be positive. Therefore, to maximize average mutual information, $r$ should be as large as possible: $r = 1$. Assume without loss of generality that $\mathbb{E}|\alpha_i|^2 = 1$. Averaging (4.67) over
Transmit diversity with channel side information

\( a_1 = |\alpha_1| \) and \( a_2 = |\alpha_2| \), which are Rayleigh distributed, yields

\[
\max_{0 \leq \lambda \leq 1} E_{\Theta|S} \int_{a_2=0}^{\infty} \int_{a_1=0}^{\infty} \log(1 + \{\lambda a_1^2 + (1 - \lambda)a_2^2
\]
\[
+ 2a_1a_2 \sqrt{\lambda(1 - \lambda)} \cos(\Theta)\}) \frac{e}{N_0} \cdot 4a_1a_2e^{-a_1^2}e^{-a_2^2} da_1 da_2. \tag{4.69}
\]

The maximum is found by setting the derivative of (4.69) with respect to \( \lambda \) equal to zero:

\[
\int_{a_2=0}^{\infty} \int_{a_1=0}^{\infty} \frac{(a_1^2 - a_2^2) e^{-a_1^2}e^{-a_2^2}}{\lambda a_1^2 + (1 - \lambda)a_2^2 + 2a_1a_2 \sqrt{\lambda(1 - \lambda)} \cos(\Theta)} \frac{e}{N_0} da_1 da_2 = 0. \tag{4.70}
\]

By symmetry, (4.70) is satisfied at \( \lambda = \frac{1}{2} \). Evaluating the second derivative confirms that this is indeed a maximum. The resulting covariance matrix is again rank 1.

We have shown that with \( N = 1 \) bit of side information, identical partitioning schemes and rank 1 covariance matrices yield local maxima for both average mutual information and average SNR. However, maximum average mutual information is not guaranteed to be achieved with a rank 1 covariance matrix. Simulations show that with \( N = 2 \) bits of side information, the maximizing covariance matrix may have a full rank of 2. This is indeed the case if one bit of side information specifies magnitude and the other represents phase information.

When the performance measure is expected mutual information rather than expected SNR, the optimal covariance matrix \( \Lambda_X \) in general has rank greater than 1. If no side information is available to the transmitter, we showed in Chapter 2 that unconstrained signaling, which corresponds to an optimal full rank covariance matrix, improves mutual information by at most \( \gamma \approx 0.833 \) bits over the best rank 1 method. This maximal gain is achieved in the limit of large number of transmit antennas and high SNR. We conjecture that this gap decreases to zero as the amount of side information increases, so that SNR-based design and mutual information-based design become equivalent.
Chapter 5

Transmit diversity for time-varying channels

In previous sections we have assumed that the channel is sufficiently slowly time-varying so that the channel is constant over the length of the codeword. Many information theoretic results rely, however, on assuming the codeword length is infinitely long. Unfortunately, most radio channels are not constant over all time. In mobile communications, the paths from the transmitter to the receiver are constantly changing. Smaller variations are due to the changing physical characteristics of the media. Since the received signal is the sum of several paths, changes in path lengths and amplitudes determine whether the paths add constructively or destructively and consequently vary the amplitude and phase of the received signal. As the channel varies more rapidly, it becomes increasingly difficult to measure the channel [19]. We ignore this effect for now and assume the channel can be perfectly measured. In this way we can separate the effects of time-variation from effects of poor channel measurement.

5.1 Channel model

The $M$-antenna transmitter diversity channel is shown in Figure 2.1 and described in Chapter 2. We now assume the channel coefficients, $\alpha_i$, are varying probabilistically with time. The channel coefficients are modeled as a Gauss-Markov process. The channel coefficient for antenna $i$ at time $k$ is given by

$$\alpha_{i,k} = \lambda \alpha_{i,k-1} + w_k,$$  \hspace{1cm} (5.1)
Transmit diversity for time-varying channels

where $w_k$ at each time $k$ is an i.i.d. complex Gaussian random variable with mean zero and variance $(1 - \lambda^2)\sigma_w^2$, and where $\lambda \in [0, 1]$ parameterizes how quickly the channel is varying in time. The channel coefficients are thus marginally distributed under the Rayleigh fading model, i.e., they are identically distributed, zero mean, complex Gaussian random variables with variance $E|\alpha_i|^2 = \sigma_a^2$. The transmitter is assumed to know the model (5.1) but not the values of the channel coefficients, while the receiver is assumed to have complete channel knowledge.

5.2 Effects of time variation on transmitter diversity

Much of our previous analysis assumed that the channel was effectively constant over long blocklengths, i.e., $\lambda = 1$. If $\lambda < 1$, the channel is time-varying and the channel coefficients vary as an ergodic random process. In this case, the average mutual information is the Shannon capacity in the usual sense. Thus, for infinite length coding intervals, it is possible to communicate at the rate of average mutual information without errors.

Practical decoding delay constraints generally prohibit using infinite blocklengths. With finite blocklengths, mutual information is a random variable and it is impossible to communicate at any non-zero rate without errors. For any finite blocklength $K > 1$, the probability of outage associated with communicating at rates up to average mutual information decreases as the channel varies more rapidly. This may lead us to the conclusion that rapidly time varying channels, if they can be measured, are beneficial.

Indeed many of the transmit diversity schemes created artificially time or frequency varying channels in order to reduce outage probabilities over finite blocklengths. Thus, if the channel is varying sufficiently rapidly, transmit diversity may not be necessary. Recall that only unconstrained signaling improved the average mutual information of the channel. The scalar diversity schemes only reduced probability of outage. If the channel is slowly varying in time, we may wish to use transmit diversity to further reduce the coherence time or coherence bandwidth of the channel. The diversity schemes may, however, be adversely affected by the real time variation in the channel. We show below that the capacity associated with all the transmit diversity schemes except time-shifting is independent of the rate of channel variation. For the time-shifting scheme, capacity decreases with time variability.
5.2 Effects of time variation on transmitter diversity

The output of the $M$ transmit antenna time-varying channel is

$$y_k = \sum_{i=1}^{M} \alpha_{i,k} x_{i,k} + v_k. \quad (5.2)$$

Since unconstrained signaling, time division, frequency shifting, and random weighting diversity schemes are memoryless, the mutual information can be computed on a symbol by symbol basis. For frequency division, the $M$-antenna channel is most easily evaluated in the frequency domain as $M$ parallel independent single transmit antenna channels, each with bandwidth $\pi/M$. For all these diversity schemes, the mutual information at each time $k$ is a function of the channel coefficient vector for that time $\alpha_k$. The average mutual information or capacity depends only on the pdf of $\alpha_k$, thus it is independent of time variation. Time variation reduces the blocklengths needed to achieve similar performance with a constant number of antennas and constant coding interval.

Time variation is only problematic for the time-shifting diversity scheme. Recall that time shifting diversity converts the vector input channel into a scalar ISI channel. The output of the time-varying time-shifting diversity channel is

$$y_k = \sum_{i=1}^{M} \alpha_{i,k} x_{k-i+1} + v_k. \quad (5.3)$$

On a time-varying channel, the time-shift diversity scheme creates time-varying ISI. We show below that time-varying ISI is worse than constant ISI since it reduces the mutual information of the channel. The frequency division scheme also causes ISI, but only the overall gain of the channel, not the ISI, changes randomly with time. The relative weights of the taps of the bandpass filter remain constant.
5.3 Time-varying ISI

The time-shifting diversity channel may be written in matrix form, \( \mathbf{Y}^N = A_N \mathbf{X}^N + \mathbf{V}^N \). For example, for \( M = 2 \) antennas and a sequence of \( N \) output symbols:

\[
\begin{bmatrix}
  Y_1 \\
  Y_2 \\
  \vdots \\
  Y_N
\end{bmatrix} = 
\begin{bmatrix}
  \alpha_{2,1} & \alpha_{1,1} \\
  \alpha_{2,2} & \alpha_{1,2} \\
  \vdots & \vdots \\
  \alpha_{2,N} & \alpha_{1,N}
\end{bmatrix} 
\begin{bmatrix}
  X_0 \\
  X_1 \\
  \vdots \\
  X_N
\end{bmatrix} + 
\begin{bmatrix}
  V_1 \\
  V_2 \\
  \vdots \\
  V_N
\end{bmatrix},
\]

(5.4)

where \( A_N \) is a matrix of random time-varying channel coefficients. As always, since the channel is known to the receiver but unknown to the transmitter, we assume i.i.d. Gaussian inputs over the two antennas with total power constrained to \( \sum_{i=1}^M E|X_{i,k}|^2 = \mathcal{E}_s \). Since the channel coefficients vary ergodically, the per-symbol mutual information between input and output sequences of length \( N \), in the limit of large blocklength, converges to the expected mutual information or equivalently the channel capacity:

\[
\lim_{N \to \infty} \frac{1}{N} I_N(\mathbf{X}^N; \mathbf{Y}^N) = E[ I(\mathbf{X}; \mathbf{Y}) ].
\]

(5.5)

The mutual information of the \( N \)-block channel is

\[
\frac{1}{N} I_N(\mathbf{X}^N; \mathbf{Y}^N) = h(\mathbf{Y}^N) - h(\mathbf{Y}^N | \mathbf{X}^N) = \frac{1}{N} \log \det(K_Y) - \log \det(K_V)
\]

(5.6)

\[
= \frac{1}{N} \log \det \left( I + A_N A_N^H \frac{\mathcal{E}_s}{MN_0} \right)
\]

(5.7)

(5.8)

where \( K_Y = A_N A_N^H \frac{\mathcal{E}_s}{MN_0} + N_0 I \) and \( K_V = N_0 I \) are the output and noise covariance matrices respectively and \( I \) is the \( N \)-dimensional identity matrix.

The matrix

\[
\mathcal{A}_N = I + A_N A_N^H \frac{\mathcal{E}_s}{MN_0}
\]

(5.9)

has non-zero entries only on its main diagonal and the \( M - 1 \) diagonals below and above the main diagonal. For example, for \( M = 2 \) this matrix is tridiagonal. We can factor \( \mathcal{A}_N = LU \)
5.3 Time-varying ISI

Figure 5.1: Mutual information as a function of the time correlation $\lambda$ between successive channel coefficients at SNR $= \sigma^2 \frac{E_r}{N_0}$ of 20 dB.

into a lower triangular matrix, $L$, with 1's on the diagonal times an upper triangular matrix $U$. Since $\det \mathbf{A}_N = \det L \det U$ and since $\det L = 1$, the determinant of $\mathbf{A}_N$ is simply equal to the product of the diagonal elements of $U$. Furthermore, the diagonal element on the $k$th row of $U$ can be written as a function of the previous $M - 1$ rows of $U$ and the $k$th set of channel coefficients $\alpha_{1,k} \alpha_{2,k} \ldots \alpha_{M,k}$. Therefore the diagonal elements of the $N \times N$ matrix $U$ can be computed recursively. This allows us to easily evaluate (5.8) for large $N$.

We use Monte Carlo simulations and the recursive determinant calculation technique to compute the mutual information of the channel for large $N$. As mentioned above above, $\frac{1}{N}$ times the per-block mutual information converges to a constant in the limit of infinite $N$. The value of $N$ needed for practical convergence is dependent upon the time correlation between the channel coefficients, $\lambda$. For $\lambda$ close to 1, the required $N$ becomes very large and it becomes impractical to compute the limiting determinant. For $M = 2$ antennas and $\lambda = 0$ (i.i.d. antenna coefficients for each time sample), the empirical variance of $\frac{1}{N} I_N$ is $1.1 \times 10^{-5}$ for $N = 10^4$. Using Monte Carlo simulations with $N = 10^6$, we find that mutual information increases with time correlation $\lambda$. In other words, time variation reduces the mutual information of an ISI channel. Mutual information is plotted as a function of $\lambda$ for $0 \leq \lambda \leq 0.7$ and assuming $\sigma^2 \frac{E_r}{N_0}$ of 20 dB in Figure 5.1. In the limit as $\lambda \to 1$, we expect the mutual information to converge to the time-invariant average mutual information of $E[\log(1 + |\alpha_1|^2 \frac{E_r}{N_0})] = 5.884$ bits/complex symbol.
5.3.1 Example: Why random time-varying ISI is worse than time-invariant ISI

The detrimental effects of time variation on an ISI channel can be explained via a simple example. Consider a 2-tap ISI channel parameterized by a single channel coefficient $\alpha_k \in \{0, 1\}$ for each time $k$. The output of the channel is

$$y_k = \alpha_k x_k + (1 - \alpha_k) x_{k-1},$$  \hspace{1cm} (5.10)

which implies that for each time $k$, either tap 1 is active and the current symbol is transmitted, or tap 2 is active and the previous symbol is transmitted. The transmitter of course does not know which tap is "good" at any time instant. We assume a binary input $x_k \in \{-1, 1\}$ where each symbol is equally likely. The evolution of the stationary random process $\{\alpha_k\}$ is described by the two-state Markov chain shown in Figure 5.2, where state 0 corresponds to $\alpha_k = 0$ and state 1 to $\alpha_k = 1$. The parameter $p$ is the probability of switching from one channel state to the other at each time. The initial state is $\alpha_k = 0$ or $\alpha_k = 1$ with probability $\frac{1}{2}$ each.

The mutual information between a sequence of $N$ input and output symbols is

$$I_N(X^N; Y^N) = H(Y^N) - H(Y^N|X^N).$$  \hspace{1cm} (5.11)

Since the channel state sequence $\alpha^N$ is assumed to be known at the receiver, $H(Y^N|X^N) = 0$. To compute $H(Y^N)$, we calculate the probability mass function

$$P(Y^N|\alpha^N) = \prod_{k=1}^{N} P(y_k|y_{k-1}, \alpha_{k-1}, \alpha_k).$$  \hspace{1cm} (5.12)
5.3 Time-varying ISI

If \( \alpha_k = 0 \) and \( \alpha_{k-1} = 1 \), then \( y_k = y_{k-1} \) with probability 1. For all other choices of \( \{\alpha_k, \alpha_{k-1}\} \), \( y_k \) is independent of \( y_{k-1} \). Let \( l \) be the number of times \( k \) such that \( \alpha_{k-1} = 1 \) and \( \alpha_k = 0 \) in the sequence of \( N \) channel states. Then,

\[
\frac{1}{N} [I_N(X^N, Y^N)] = \frac{1}{N} H(Y^N) = \frac{1}{N} \sum p(l) H(Y^N|l) = \frac{1}{N} \sum -p(l) \log_2(\frac{1}{2})^{N-l} = \frac{1}{N} \sum p(l)(N - l) = 1 - \frac{1}{N} E[l].
\] (5.13)

The probability of a transition in channel state is given by \( p \) and thus the probability of transition from state 1 to state 0 is \( \frac{p}{2} \). In the limit of infinite \( N \), the expected number of transitions from state 1 to state 0 per time sample is

\[
\lim_{N \to \infty} \frac{1}{N} E[l] = \frac{p}{2}.
\] (5.18)

From (5.17), the rate of reliable communication over this time-varying channel is \( 1 - \frac{p}{2} \). As the transition probability goes to 0, i.e., as the channel varies more slowly, the rate approaches 1. As the transition probability approaches 1, the capacity approaches \( \frac{1}{2} \) since we receive two copies of every other symbol and zero copies of the remaining symbols. This simple example suggests that in general, time-varying ISI decreases the rate of reliable communication because variations in the relative magnitudes of the ISI coefficients may cause us to receive several high SNR copies of one symbol and several low SNR copies of another symbol, rather than both high and low SNR copies of each symbol.

5.3.2 Effect of time-varying ISI as \( M \) increases to three antenna elements

We now return to our original channel model with channel coefficients modeled by the Gauss-Markov process (5.1). As mentioned above, the mutual information of the time-varying ISI channel may be computed using Monte Carlo simulations. Above, we examined the case of
Transmit diversity for time-varying channels

\[ M = 2 \] antenna elements or ISI of length 2. In this section, we provide some preliminary results on the effect of ISI as the number of antenna elements increases to 3.

In Figure 5.3 we plot the difference in mutual information of the \( M = 3 \) antenna time-varying ISI channel and the \( M = 2 \) antenna time-varying ISI channel, \( I_3 - I_2 \), as a function of SNR. We assume the channel coefficients \( \{\alpha_k\} \) are i.i.d., i.e., \( \lambda = 0 \). Interestingly, the expected mutual information for the 3-tap ISI channel is better than that of the 2-tap ISI channel above a SNR of 12 dB. Recall that in the 2-tap toy example above, we showed that time-varying ISI increased the probability of receiving two poor copies of a symbol. With three tap ISI, the probability that all copies of the symbol are poor decreases, since there are three copies rather than two. We conjecture that this may cause the effect noticed above. The multi-tap ISI problem is clearly complex and a full characterization of the ISI channel with \( M > 2 \) taps would be an interesting area for future work.

5.3.3 Upper bound for capacity when \( \lambda = 0 \)

We now consider the limiting case in which the channel coefficients are independent and identically distributed at each time \( k \), i.e., \( \lambda = 0 \). For \( M = 2 \) transmit antennas, we upperbound the per symbol mutual information in the limit as blocklength \( N \to \infty \), i.e., the channel capacity. Using this upperbound, we show that for SNR above 16.43 dB, the capacity of the time-varying ISI channel is strictly less than the average mutual information of a time-invariant ISI channel.
5.3 Time-varying ISI

Let $A_N$ be defined as in (5.9) for the case of $M = 2$ transmit antennas. Recall $A_N$ is tridiagonal in this case. Let $a_1 \, a_2 \ldots \, a_N$ denote the elements on the main diagonal of $A_N$ and let $b_1 \, b_2 \ldots \, b_{N-1}$ and $c_1 \, c_2 \ldots \, c_{N-1}$ denote the elements on the diagonals above and below the main diagonal, respectively. To simplify notation, also define $\nu = \frac{c_2}{N_0}$.

The capacity can be written as a function of the expected value of the per-symbol mutual information of a blocklength $N$ channel:

$$C_{TV\, ISI} = \lim_{N \to \infty} \frac{1}{N} \log \det A_N \quad (5.19)$$

$$= E \left[ \lim_{N \to \infty} \frac{1}{N} \log \det A_N \right] \quad (5.20)$$

$$\leq \lim_{N \to \infty} E \left[ \frac{1}{N} \log \det A_N \right]. \quad (5.21)$$

Equation (5.21) follows from Fatou’s Lemma [2]. We determine an upper bound for the expected per-symbol mutual information using Jensen’s inequality:

$$E \left[ \frac{1}{N} \log \det A_N \right] \leq \frac{1}{N} \log E[\det A_N]. \quad (5.22)$$

The determinant of the tridiagonal $N \times N$ matrix, $A_N$, can be written recursively in terms of the determinant of the $N - 1 \times N - 1$ and $N - 2 \times N - 2$ principal submatrices $A_{N-1}$ and $A_{N-2}$ [1]:

$$\det A_N = a_N \det A_{N-1} - b_{N-1} c_{N-1} \det A_{N-2}. \quad (5.23)$$

Since $A_{N-1}$ is independent of $a_N$ and $A_{N-2}$ is independent of $b_{N-1} c_{N-1}$, the expected determinant is

$$E[\det A_N] = E[a_N] E[\det A_{N-1}] - E[b_{N-1} c_{N-1}] E[\det A_{N-2}], \quad (5.24)$$

where

$$E[a_N] = E[1 + \nu \left| \alpha_{1,n} \right|^2 + \left| \alpha_{2,n} \right|^2] \quad (5.25)$$

$$= 1 + \nu \sigma_n^2 \quad (5.26)$$
Transmit diversity for time-varying channels

and

\[ E[b_{N-1}c_{N-1}] = E\left[ \frac{\nu^2}{4} |\alpha_{1,n-1}|^2 |\alpha_{2,n}|^2 \right] = \frac{\nu^2}{4} \sigma_a^4. \tag{5.28} \]

We can calculate \( E[\det A_N] \) by solving the difference equation (5.24) with initial conditions \( E[\det A_0] = 0 \) and \( E[\det A_1] = 1 + \nu \sigma_a^2 \). This yields

\[ E[\det A_N] = \gamma(t^N - s^N), \tag{5.29} \]

where

\[ \gamma = \frac{1 + \nu \sigma_a^2}{\sqrt{1 + 2 \nu \sigma_a^2}}, \tag{5.30} \]

\[ r = \frac{1 + \nu \sigma_a^2 + \sqrt{1 + 2 \nu \sigma_a^2}}{2}, \tag{5.31} \]

and

\[ s = \frac{1 + \nu \sigma_a^2 - \sqrt{1 + 2 \nu \sigma_a^2}}{2}. \tag{5.32} \]

In the limit of large \( N \),

\[ \lim_{N \to \infty} \frac{1}{N} E[\log \det A_N] \leq \lim_{N \to \infty} \frac{1}{N} \log E[\det A_N] \leq \log r \tag{5.33} \]

\[ = \log \left( \frac{1 + \frac{\xi}{N_0} \sigma_a^2 + \sqrt{1 + 2 \frac{\xi}{N_0} \sigma_a^2}}{2} \right), \tag{5.35} \]

where (5.34) results since \( r > s \).

We can compare the upperbound in (5.35) to the expected mutual information of the time shift diversity scheme applied to a time invariant channel. This corresponds to the case where \( \lambda = 1 \). In Chapter 2 the expected mutual information of a time shifting diversity channel was shown to equal the expected mutual information of the single antenna channel.
5.3 Time-varying ISI

Figure 5.4: The expected mutual information of the time-invariant ISI channel ($\lambda = 1$) is compared to the derived upperbound for the capacity of the time-varying ISI channel ($\lambda = 0$).

Figure 5.5: The capacity of the time-varying ISI channel is compared to the derived upper bound (5.35).

Explicit evaluation of the expectation in (2.32) yields

$$E[I_{TS}] = e^{\frac{1}{\sigma_A^2 E_s/N_0}} E_t \left( \frac{1}{\sigma_A^2 E_s/N_0} \right),$$

(5.36)

where $E_t(x) = \int_1^\infty \frac{e^{-xt}}{t} \, dt$. The graph in Figure 5.4 shows that the capacity of the time-varying ISI channel with $\lambda = 0$ is strictly less than the average mutual information of the time-invariant channel for values of $\sigma_A^2 E_s/N_0 > 16.43$ dB. For lower SNR values, a tighter upper bound is necessary. In Figure 5.5, we compare the upperbound (5.35) to the capacity of the time-varying ISI channel. Here, we use Monte Carlo simulations to compute the capacity of the time-varying ISI channel. (A formula for this capacity is derived in the following...
section.) It can be seen that this upper bound is not very tight. However, it is sufficient to prove that for large values of SNR, the capacity of a time varying ISI channel (for \( \lambda = 0 \)) is strictly lower than the expected mutual information of a time-invariant ISI channel.

5.3.4 Mutual information computation: \( \lambda = 0 \)

In this section, we follow an alternative approach to show that time variation on an ISI channel reduces mutual information. Recall from Section 5.3 that the mutual information of the time-varying ISI channel can be computed as the log of the product of the diagonal elements of upper triangular matrix \( U \) in the LU decomposition of \( A_N \). Furthermore, for the case of \( M = 2 \), the diagonal element associated with each row \( k \) is a function of the elements on the previous row. In other words, the diagonal elements form a first order Markov chain. We derive a stationary distribution for this Markov chain and show that this distribution is unique and represents the asymptotic distribution of the diagonal elements. From this stationary distribution, we calculate the capacity of the time-varying ISI channel for the case \( M = 2 \) and \( \lambda = 0 \).

5.3.4.1 Stationary distribution calculation

At the beginning of Section 5.3, we defined \( A_N = I + A_N A_N^H \frac{\xi}{MN} \) (5.9) and \( A_N \) in terms of the channel coefficients \( \alpha_{i,k} \) (5.4). As mentioned above, the mutual information of the time-varying ISI channel is equal to the log of the product of the diagonal elements of \( U \), where \( U \) is the upper triangular matrix \( U \) in the LU decomposition of \( A_N \). From this LU decomposition, the \( k \)th diagonal element \( d_k \) of \( U \) as a function of \( d_{k-1} \) and independent complex Gaussian random variables \( \alpha_{1,k-1}, \alpha_{1,k} \) and \( \alpha_{2,k} \) is

\[
d_k = 1 + \varphi(|\alpha_{1,k}|^2 + |\alpha_{2,k}|^2) - \frac{\varphi^2|\alpha_{1,k-1}|^2|\alpha_{2,k}|^2}{d_{k-1}} \tag{5.37}
\]

\[
= 1 + \varphi|\alpha_{1,k}|^2 + \varphi|\alpha_{2,k}|^2 \left( 1 - \varphi \frac{|\alpha_{1,k-1}|^2}{d_{k-1}} \right), \tag{5.38}
\]

- 110 -
where \( \varrho = \frac{\xi}{2N_0} \). If we define

\[
\chi_k = |\alpha_{2,k}|^2 \left( 1 - \varrho \frac{|\alpha_{1,k-1}|^2}{d_{k-1}} \right),
\]

(5.39)

then

\[
d_k = 1 + \varrho |\alpha_{1,k}|^2 + \varrho \chi_k.
\]

(5.40)

Notice that \( \alpha_{1,k} \) is independent of \( \chi_k \). By substituting (5.40) into (5.39), with some simple algebraic manipulation we can write \( \chi_k \) as a function of independent random variables \( \chi_{k-1}, |\alpha_{1,k-1}|^2, \) and \( |\alpha_{2,k}|^2 \):

\[
\chi_k = |\alpha_{2,k}|^2 \left( \frac{1}{1 + \frac{|\alpha_{1,k-1}|^2}{\varrho + \chi_{k-1}}} \right).
\]

(5.41)

Thus \( \chi_k \) is a first order discrete-time continuous state space Markov chain with transitional probability \( f_{\chi_k|\chi_{k-1}}(\xi_k|\xi_{k-1}) \):

\[
f_{\chi_k}(\xi_k) = \int_{\chi_{k-1}} f_{\chi_k|\chi_{k-1}}(\xi_k|\xi_{k-1}) f_{\chi_{k-1}}(\xi_{k-1}) \, d\xi_{k-1}.
\]

(5.42)

A stationary distribution \( f_{\chi}(\xi) \) for this Markov chain is the distribution for which \( f_{\chi_k}(\xi_k) = f_{\chi_k|\chi_{k-1}}(\xi_k|\xi_{k-1}) = f_{\chi}(\xi) \) for all \( k \) [5]. We first determine a stationary distribution for \( \chi_k \) and then using (5.40), calculate a stationary distribution for \( d_k \). Without loss of generality, we assume \( E|\alpha_{2,k}|^2 = 1 \) since \( \varrho \) can be used to account for this scaling. The pdf of \( \chi_k \) can be written in terms of the following intermediate random variables:

\[
\mu = \frac{1}{\varrho} + \chi_{k-1}
\]

(5.43)

\[
\nu = \frac{|\alpha_{1,k-1}|^2}{\mu}
\]

(5.44)

\[
\zeta = 1 + \nu
\]

(5.45)

\[
\chi_k = \frac{|\alpha_{2,k}|^2}{\zeta}.
\]

(5.46)
Transmit diversity for time-varying channels

A stationary distribution \( f_X(\xi) \) for \( \chi_k \) is a solution of the following integral equation:

\[
f_X(\xi) = \frac{d}{d\xi} \left[ \int_{\xi=1}^{\infty} \frac{d}{d\zeta} \left[ \int_{\mu=1/\varrho}^{\infty} \frac{d}{d\mu} \left[ \int_{\xi=0}^{\infty} f_X(\xi) d\xi \right] (1 - e^{-\mu(\xi - 1)}) d\mu \right] (1 - e^{-\xi \varrho}) \right]. \tag{5.47}
\]

It is straightforward to verify that

\[
f_X(\xi) = \frac{\varrho e^{-\frac{1}{\varrho} + \xi \varrho}}{E_i(\frac{1}{\varrho})(1 + \xi \varrho)}, \quad \xi \geq 0, \tag{5.48}
\]

is a stationary solution.

5.3.4.2 Ergodicity of the stationary distribution

In this section we use theorems and discussion in Section VIII.7 of Feller [5] to show that the conditional pdf \( f_{X_k|\chi_{k-1}} \) is ergodic, i.e., the stationary distribution derived above is unique and represents the asymptotic distribution of \( \chi_k \). By definition, the conditional pdf \( f_{X_k|\chi_{k-1}} \) is ergodic if there exists a pdf \( f_X \) such that \( f_{X_k} \to f_X \) independently of the initial distribution \( f_{\chi_0} \).

Theorem 2 in Feller VIII.7 states that a strictly positive regular conditional pdf \( f_{X_k|\chi_{k-1}} \) is ergodic if and only if there exists a resulting strictly positive stationary probability distribution \( f_X \). A pdf is strictly positive in \( \Omega \) if the probability of any open interval \( I \subset \Omega \) is greater than 0. A conditional pdf is strictly positive if the conditional probability of entering any interval \( I \) from each point in \( \Omega \) is greater than 0. It is clear from (5.48) that \( f_X \) is strictly positive over the set \( \Omega = \{ \chi : \chi \geq 0 \} \). The probability \( \chi_k \) is in an interval \( (a, b) \in \Omega \), for \( 0 \leq a < b < \infty \), given the previous value \( \chi_{k-1} \) is

\[
\int_a^b f_{X_k|\chi_{k-1}}(\xi) d\xi = \int_a^b \frac{e^{-\xi(\frac{1}{\varrho} + \chi_{k-1})}(1 + \xi + (\frac{1}{\varrho} + \chi_{k-1}))}{(\xi + (\frac{1}{\varrho} + \chi_{k-1}))^2} d\xi \tag{5.49}
\]

\[
= \frac{e^{-a} - e^{-b}}{1 + \frac{a}{\frac{1}{\varrho} + \chi_{k-1}}} - \frac{e^{-b} - e^{-a}}{1 + \frac{b}{\frac{1}{\varrho} + \chi_{k-1}}}. \tag{5.50}
\]

This probability is always positive, since \( e^{-a} > e^{-b} \) and since \( 1 + \frac{b}{\frac{1}{\varrho} + \chi_{k-1}} > 1 + \frac{a}{\frac{1}{\varrho} + \chi_{k-1}} \). Thus, \( f_{X_k|\chi_{k-1}} \) is strictly positive.
5.3 Time-varying ISI

Since a strictly positive stationary pdf $f_x$ exists, all that remains is to show that $f_{x_k|x_{k-1}}$ is regular. By definition, the conditional pdf $f_{x_k|x_{k-1}}$ is regular if $f_{x_0}$ uniformly continuous implies $f_{x_k}$ is uniformly continuous for all $k$. Example VIII.7(b) in Feller shows that if $f_{x_k|x_{k-1}}$ is a uniformly continuous transformation, it is also regular in the sense above. To show that $f_{x_k|x_{k-1}}$ is a uniformly continuous transformation, we show that each of the intermediate conditional pdf’s used to derive $f_{x_k|x_{k-1}}$ (Equations (5.43) – (5.46)), namely $f_{\mu|x_{k-1}}, f_{\nu|\mu}, f_{\xi|\nu}$, and $f_{x_k|\xi}$, is a uniformly continuous transformation.

A function $g(x)$ is uniformly continuous if for every $\epsilon > 0$ there exists a corresponding $\delta > 0$ such that $|x' - x''| < \delta$ implies $|g(x') - g(x'')| < \epsilon$. A conditional pdf $f_{y|x}(y_0|x_0)$ is a uniformly continuous transformation if $f_x(x_0)$ uniformly continuous implies

$$f_y(y_0) = \int_{x_0} f_{y|x}(y_0|x_0)f_x(x_0)\;dx_0$$

(5.51)

is uniformly continuous. From ((5.43) – (5.46)) it is clear that we only need to consider two types of transformations: (1) a constant shift and (2) dividing an exponential random variable by another random variable whose domain is $\frac{1}{\theta} \to \infty$ for some positive constant $\theta$.

Consider case (1): $y = x + c$ where $x$ is a random variable and $c$ is a constant. Obviously, if $f_x(x_0)$ is uniformly continuous, $f_y(y_0)$ must be uniformly continuous since the transformation simply shifts the distribution. Case (2) corresponds to $y = w/x$ where $w$ is an exponential random variable. Assume for simplicity that $E[w] = 1$. The pdf of $y$ in terms of $f_x(x_0)$ is

$$f_y(y_0) = \int_{x_0=\frac{1}{\theta}}^{\infty} x_0 e^{-y_0 x_0} f_x(x_0)\;dx_0.$$  \hspace{1cm} (5.52)

Let $y'_0 = y_0 + \delta$. Then since $e^{-y'_0} < e^{-y_0}$,

$$|f_y(y'_0) - f_y(y''_0)| = \int_{x_0=\frac{1}{\theta}}^{\infty} x_0[e^{-y'_0 x_0} - e^{-y''_0(x_0+\delta)}]f_x(x_0)\;dx_0$$

(5.53)

$$= \int_{x_0=\frac{1}{\theta}}^{\infty} x_0 e^{-x_0 \delta}[1 - e^{-\delta x_0}]f_x(x_0)\;dx_0$$

(5.54)

$$\leq [1 - e^{-\delta/\theta}] \int_{x_0=\frac{1}{\theta}}^{\infty} x_0 f_x(x_0)\;dx_0$$

(5.55)

$$= [1 - e^{-\delta/\theta}] E[x].$$

(5.56)

Since $\chi_0$ is exponentially distributed, its expected value is finite. From ((5.43) – (5.46)) it
 Transmit diversity for time-varying channels

can be seen that if \( \chi_0 \) has finite mean, the random variables at each step in each iteration have finite mean. Thus \( E[x] \) is guaranteed to be finite. To ensure \( |f_y(y'_0) - f_y(y''_0)| < \epsilon \), we choose \( \delta \) such that \( [1 - e^{-\delta}] E[x] < \epsilon \). When \( \frac{\epsilon}{E[x]} \geq 1 \), any \( \delta > 0 \) is sufficient. If \( \frac{\epsilon}{E[x]} < 1 \), we require

\[
\delta < -\frac{1}{\varrho} \ln \left[ 1 - \frac{\epsilon}{E[x]} \right].
\]

(5.57)

Obviously, switching the roles of \( y'_0 \) and \( y''_0 \) leads to the same result. Therefore the transformation is uniformly continuous.

5.3.4.3 Mutual information

In the previous sections we derived a limiting distribution for \( \chi_k \) and showed that this stationary distribution is unique. Since the diagonal elements, as defined in (5.40), are a shifted sum of the random variable \( \chi_k \) and an independent exponentially distributed random variable, a unique stationary distribution exists for \( d_k \) as well. Recall that the mutual information of the \( N \)-block channel is

\[
\frac{1}{N} I_N(X^N; Y^N) = \frac{1}{N} \log \det(A_N) = \frac{1}{N} \log \det(LU) = \frac{1}{N} \log \prod_{k=1}^{N} d_k = \frac{1}{N} \sum_{k=1}^{N} \log(d_k).
\]

(5.58) (5.59) (5.60) (5.61)

If the strong law of large numbers holds, even though \( \{d_k\} \) is not i.i.d.,

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \log(d_k) = E[\log(d_k)],
\]

(5.62)

and we can compute the per-letter capacity of the time-varying channel from the stationary distribution of \( d_k \). In this section we prove that the strong law of large numbers holds.

From Theorem 17.0.1 in Meyn and Tweedie [20], we know that the strong law of large
numbers holds for \( \log(d) \) if

i \( d_k \) is a positive Harris chain with invariant probability \( f_d \), and

ii if \( E[\log(d)] < \infty \).

The second condition is clearly satisfied, since

\[
E[\log(d_k)] \leq E[d_k] \tag{5.63}
\]

\[
= 1 + \varrho E[|\alpha_{1,k}|^2] + \varrho E[\chi_k] \tag{5.64}
\]

\[
< \infty. \tag{5.65}
\]

From the definition of Harris recurrence, it will become clear that \( \chi \) Harris recurrent implies \( d \) is Harris recurrent. Thus, we will begin by showing \( \chi \) is Harris recurrent.

The range of random variables \( \chi_k \) is \( \Omega = [0, \infty) \). Let \( \mathcal{B}(\Omega) \) be the set of Borel sets on the space \( \Omega \). Define \( L(x, A) \) as the probability of reaching a set \( A \in \mathcal{B}(\Omega) \) from a point \( x \in \Omega \) in finite time. Also, define \( Q(x, A) \) as the probability of returning to set \( A \in \mathcal{B}(\Omega) \) infinitely often from \( x \in \Omega \). From [20, Chap. 9], a Markov chain \( \chi \) is Harris recurrent if there exists a probability measure \( \psi \) on \( \mathcal{B}(\Omega) \) such that

i \( \chi \) is \( \psi \)-irreducible, and

ii for every set \( A \in \mathcal{B}(\Omega) \) with \( \psi(A) > 0 \) and for every element \( x \in A \), the probability of returning to set \( A \) infinitely often, \( Q(x, A) \), from \( x \) is equal to 1.

A Markov chain \( \chi \) is \( \psi \)-irreducible [20, Chap. 4] if

i for every set \( A \in \mathcal{B}(\Omega) \) with \( \psi(A) > 0 \), the probability of reaching \( A \) from \( x \in \Omega \) in finite time is positive, i.e., \( L(x, A) > 0 \),

ii \( \psi(A) = 0 \), then \( \psi\{y : L(y, A) > 0\} = 0 \), and

iii for any measure \( \psi' \) on \( \mathcal{B}(\Omega) \) such that \( \psi'(A) > 0 \) implies \( L(x, A) > 0 \), if \( \psi'(A) > 0 \) then \( \psi'(A) > 0 \).
Transmit diversity for time-varying channels

The probability measure \( \psi \) is an indicator function which indicates the sets which can be reached in finite time by the Markov chain. In our case, the seemingly complex definition of \( \psi \)-irreducibility is satisfied by any probability measure \( \psi \in B(\Omega) \) with support \([0, \infty)\) and no atoms.

We now prove condition (ii) for Harris recurrence. Since the Borel field on \([0, \infty)\) is generated by the intervals \([a, b]\) where \(0 \leq a < b < \infty\), we can prove condition (ii) for Harris recurrence by showing that the probability of returning to any interval \([a, b]\) infinitely often from any point in \(\Omega\) is equal to 1. We first generate a lower bound on the probability of entering \([a, b]\) from \(x\) in one step. This probability was calculated in (5.50) and is repeated for convenience. The probability \(\chi_k\) enters interval \([a, b]\) from \(\chi_{k-1} \in \Omega\) is

\[
\int_a^b f_{\chi_k|\chi_{k-1}}(\xi) \, d\xi = \frac{e^{-a}}{1 + \frac{a}{1/\theta + \chi_{k-1}}} - \frac{e^{-b}}{1 + \frac{b}{1/\theta + \chi_{k-1}}}.
\]

This quantity is guaranteed to be positive for all \(0 \leq a < b < \infty\), since \(e^{-a} > e^{-b}\) and since \(1 + \frac{b}{1/\theta + \chi_{k-1}} > 1 + \frac{a}{1/\theta + \chi_{k-1}}\). Define \(\epsilon\) to be the minimum of these strictly positive probabilities:

\[
\epsilon = \min_{\chi_{k-1}} \frac{e^{-a}}{1 + \frac{a}{1/\theta + \chi_{k-1}}} - \frac{e^{-b}}{1 + \frac{b}{1/\theta + \chi_{k-1}}}.
\]  (5.66)

Then, \(\epsilon\) is a strictly positive lower bound for the probability of arriving in set \(A\) in one step.

Let \(E_n\) denote the event of arriving in set \(A\) on step \(n\). Let \(E^c\) denote the complement of \(E\). The probability of returning infinitely often to set \(A\) is

\[
\lim_{m \to \infty} P\left( \bigcup_{n=m}^{\infty} E_n \right) = 1 - \lim_{m \to \infty} P\left( \bigcap_{n=m}^{\infty} E_n^c \right) = 1 - \lim_{m \to \infty} \lim_{m' \to \infty} P(E_{m'}^c) P(E_{m+1}^c | E_m^c) \cdots P(E_{m'}^c | E_{m'-1}^c, \ldots, E_m^c) \leq 1 - \lim_{m \to \infty} \lim_{m' \to \infty} (1 - \epsilon)^{m'-m+1} = 1.
\]  (5.67)  (5.68)  (5.69)  (5.70)

Thus, \(\chi\) and \(d\) are Harris chains with invariant probability distributions.
The invariant probability distribution \( f_d \) can be computed from \( f_x \) (5.48) and is given by

\[
f_d(d_0) = \frac{\ln(d_0)e^{-\frac{d_0}{\theta}}}{E_i\left(\frac{1}{\theta}\right)q}, \quad d_0 \geq 1.
\]  

(5.71)

From (5.62), the capacity of the time-varying channel in nats is

\[
C_{TV} = E[\ln(d)]
\]

(5.72)

\[
= 2\frac{\text{MeijerG}(4,1,\frac{1}{\theta})}{E_i\left(\frac{1}{\theta}\right)q},
\]

(5.73)

where the MeijerG function is

\[
\text{MeijerG}(m,a,z) = G_{m-1,m}^{m,0} \left( \begin{array}{c} z \\ 0 \begin{array}{cccc} 0 & \cdots & 0 \\ -1 & -1 & \cdots & a-1 & -1 \end{array} \end{array} \right)
\]

(5.74)

\[
= (2\pi)^{-1} \int_L \frac{\Gamma(-1-s)^{m-1}\Gamma(a-1-s)}{\Gamma(-s)^{m-1}} z^s ds,
\]

(5.75)

and

\[
\Gamma(x) = \int_0^\infty e^{-t}t^{x-1} dt, \quad \Re(x) > 0,
\]

(5.76)

and \( L \) is the appropriate path of integration [18].

5.3.4.4 Effect of time variation on mutual information

In Figure 5.6, we compare the capacity of the time-varying ISI channel (\( \lambda = 0 \), \( C_{TV \text{ ISI}} \)) to the expected mutual information of the non time-varying ISI channel (\( \lambda = 1 \), \( E[I_{TS}] \)). The graph clearly shows that time variation on an ISI channel reduces expected mutual information. Furthermore, our results are consistent with Monte Carlo simulations shown in Figure 5.5.

As a final remark, we note that since time-variation adversely affects the time-shifting diversity scheme, we also expect using frequency-shifting diversity on a frequency selective channel will create similar problems, since time and frequency shifting are dual schemes.
Figure 5.6: Comparison of capacity for a rapidly time-varying ISI channel, $C_{TV\,ISI}$, and expected mutual information of the non-time varying ISI channel as a function of SNR, $\sigma^2 \frac{E_x}{N_0}$. 

*Transmit diversity for time-varying channels*
Chapter 6

Conclusion

6.1 Conclusions

The focus of this research has been to analyze the benefits achievable with multiple transmit antennas at the base station. We showed that only multiple antennas using unconstrained signaling increased the capacity of the Rayleigh fading channel over the single antenna channel. Communicating at rates near capacity, however, requires coding over a large number of blocks, each of which has an associated channel coefficient that varies ergodically from block to block, or requires using transmit diversity with a large number of antennas. The utility of the transmit diversity schemes really lies in their ability to reduce the probability of outage when coding over a finite number of blocks with a finite number of antennas. Base station providers, due to cost issues, are generally willing to put a small number (2–4) of antennas on the base station. Fortunately, a significant fraction of the achievable performance is gained from a small number of antennas. For a fixed outage probability, the dB loss in signal power with respect to the AWGN channel roughly halves as the number of antennas is doubled.

The linear diversity schemes convert a slowly varying, frequency non-selective channel into a channel that varies in time or frequency. In this way, the probability of being in a deep fade contiguously is reduced. If the channel has a small coherence time or coherence bandwidth, the linear diversity schemes may be used to further reduce the coherence time or bandwidth, or they may not be necessary. As a caveat, we show that the performance of time shifting diversity, a scheme that creates periodic frequency variation, degrades as the channel varies more rapidly in time. With random time variation of the ISI coefficients, several “good” (high SNR) copies may be received at the expense of receiving only “bad” (low SNR) copies of another symbol.

Much of our research focused on the scenario in which the channel is known (measured)
Conclusion

at the receiver, but unknown at the transmitter. This model is appropriate, for example, in a broadcast scenario where the transmitter must communicate with several users simultaneously. For many point-to-point channels, information about the channel coefficients may be procured by the transmitter. Channel information can provide a factor of $M$ improvement in received SNR. We show that even a small amount of channel side information is extremely beneficial.

6.2 Future research directions

Methods and theorems that allow us to calculate the mutual information of time-invariant continuous-time channels are well developed. These methods, however, do not provide closed form solutions for the mutual information of time-varying continuous-time channels. Developing theorems to aid in the evaluation of the eigenvalues of the Karhunen-Loève expansion for time-varying channels or developing a new approach for calculating the mutual information of time-varying channels is an open research area.

In Chapter 4, we determined methods for using channel side information and evaluated their utility based on performance measures of expected SNR and expected mutual information. As mentioned above, a potentially more useful measure of performance is the probability of outage, or the probability the mutual information falls below the desired communication rate. Although this analysis appears to be difficult, probability of outage may be the best measure of performance for a system with decoding delay constraints.

Random time variation of an ISI channel was shown to diminish the rate of reliable communication. Most of the results are specific to the $M = 2$ antenna/tap ISI case. A complete characterization of the effect of time variation on multiple-tap ISI channels is an open research area.

The problem of channel identification at the receiver for a time-varying channel will become severe as the number of antennas increases. Diminishing returns will eventually set in; an incorrectly designed system might even deteriorate with additional antennas. To fully understand this issue would require a theory of communication over channels on the margin of identifiability. This theory does not presently appear to exist.

Another area of current research is the design of practical channel codes for achieving
6.2 Future research directions

near-optimal performance on the multiple transmit antenna channel. Some work along these lines is presented in an upcoming paper by Tarokh et al. [31].

In concurrent work, Fochini [6] examines the performance benefits for systems with both multiple receive antennas and multiple transmit antennas. Further analysis of multiple receive and transmit antenna systems as well as the development of coding techniques for achieving the available performance improvements are further areas of research.
Appendix A

Time shifting diversity as $M \to \infty$

In this appendix we show that for the time shifting diversity scheme, as the number of transmitters $M$ increases to infinity, mutual information converges in mean square to its expected value.

The mutual information of an $M$ transmitter channel employing time shifting diversity is given in (2.14). Without loss of generality we will assume that $E_s/N_0 = 1$. Then

$$I_{TS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 + |H_M(\omega)|^2) d\omega,$$  \hspace{1cm} (A.1)

where

$$H_M(\omega) = \sum_{i=1}^{M} \frac{\alpha_i}{\sqrt{M}} e^{-j\omega(i-1)}.$$  \hspace{1cm} (A.2)

Recall that the coefficients $\alpha_i$ are i.i.d. zero mean complex Gaussian random variables. This implies that $H_M(\omega)$ is also a complex Gaussian random variable with mean zero and variance $E[|\alpha_1|^2]$.

We will show that as the number of transmitters increases to infinity, the mutual information converges to its expected value:

$$\lim_{M \to \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 + |H_M(\omega)|^2) d\omega = E[\log(1 + |\alpha_1|^2)].$$  \hspace{1cm} (A.3)

It is sufficient to show that the variance of $I_{TS}$ goes to zero:

$$\lim_{M \to \infty} \sigma^2_{I_{TS}} = 0.$$  \hspace{1cm} (A.4)
**Time shifting diversity as \( M \to \infty \)**

To calculate the variance of \( I_{TS} \) as \( M \to \infty \), we first calculate

\[
E[I_{TS}^2] = \frac{1}{4\pi^2} E\left[\int_{-\pi}^{\pi} \log(1 + |H_M(\omega_1)|^2) d\omega_1 \int_{-\pi}^{\pi} \log(1 + |H_M(\omega_2)|^2) d\omega_2 \right] \tag{A.5}
\]

\[
= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} E\left[\log(1 + |H_M(\omega_1)|^2) \log(1 + |H_M(\omega_2)|^2) \right] d\omega_1 d\omega_2 \tag{A.6}
\]

The integral in (A.6), i.e., \( E[\log(1 + |H_M(\omega_1)|^2) \log(1 + |H_M(\omega_2)|^2)] \), is bounded. To see this, we use the Schwartz Inequality and the fact that \( \log(1 + x^2) \leq x^2 \):

\[
E \left[\log(1 + |H_M(\omega_1)|^2) \log(1 + |H_M(\omega_2)|^2) \right] \\
\leq \{ E[\log(1 + |H_M(\omega_1)|^2)] \}^{1/2} \{ E[\log(1 + |H_M(\omega_2)|^2)] \}^{1/2} \tag{A.7}
\]

\[
\leq \{ E[H_M(\omega_1)]^4 \}^{1/2} \{ E[H_M(\omega_2)]^4 \}^{1/2} \tag{A.8}
\]

\[
= 3E[|\alpha_1|^2]^2 < \infty \tag{A.9}
\]

Therefore, by the Lebesgue Dominated Convergence Theorem,

\[
\lim_{M \to \infty} E[I_{TS}^2] = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \lim_{M \to \infty} E[\log(1 + |H_M(\omega_1)|^2) \log(1 + |H_M(\omega_2)|^2)] d\omega_1 d\omega_2. \tag{A.10}
\]

It remains only to establish the following limit:

\[
\lim_{M \to \infty} E[\log(1 + |H_M(\omega_1)|^2) \log(1 + |H_M(\omega_2)|^2)] = E[\log(1 + |H_1|^2)]E[\log(1 + |H_2|^2)] \tag{A.11}
\]

where \( H_1 \) and \( H_2 \) are independent Gaussian random variables with mean zero and variance \( E[|\alpha_1|^2] \). Indeed, using (A.11) with (A.10) we obtain

\[
\lim_{M \to \infty} E[I_{TS}^2] = (E[\log(1 + |H_1|^2)]^2 = (E[I_{TS}])^2, \tag{A.12}
\]

and thus

\[
\lim_{M \to \infty} \sigma^2_{I_{TS}} = 0. \tag{A.13}
\]

To establish (A.11), we first show that in the limit as \( M \to \infty \), \( H_M(\omega_1) \) and \( H_M(\omega_2) \)
are uncorrelated for all $\omega_1 \neq \omega_2 \in [-\pi, \pi)$.

\[
E[H_M(\omega_1)H_M^*(\omega_2)] = E\left[\sum_{i=1}^{M} \frac{\alpha_i}{\sqrt{M}} e^{-j\omega_1(i-1)} \sum_{k=1}^{M} \frac{\alpha_k}{\sqrt{M}} e^{-j\omega_2(k-1)}\right]
\]
\[
= \sum_{i=1}^{M} \sum_{k=1}^{M} \frac{E[\alpha_i \alpha_k^*]}{M} e^{-j(\omega_1-i)(\omega_2-k)}
\]
\[
= \sum_{i=1}^{M} \frac{E[|\alpha_i|^2]}{M} e^{-j(\omega_1-i)(\omega_2)}
\]
\[
= \frac{e^{-j(\omega_1-\omega_2)(M+1)} - 1}{M(e^{-j(\omega_1-\omega_2)} - 1)}.
\]

Therefore

\[
\lim_{M \to \infty} E[H_M(\omega_1)H_M^*(\omega_2)] = 0
\]

for all $\omega_1, \omega_2$ such that $\omega_1 \neq \omega_2 \pmod{2\pi}$. Thus, as $M \to \infty$, $H_M(\omega_1)$ and $H_M(\omega_2)$ converge in distribution to independent Gaussian random variables.

Convergence in distribution is not enough for our needs. Let $X_M = H_M(\omega_1)$, $Y_M = H_M(\omega_2)$, $X = H_1$, and $Y = H_2$. We have shown that in the limit as $M \to \infty$, $(X_M, Y_M) \to (X, Y)$ in distribution. By Skorohad’s Theorem [2, Thm. 25.6], we can construct random variables $(\tilde{X}_M, \tilde{Y}_M)$ and $(\tilde{X}, \tilde{Y})$ such that: $(X_M, Y_M)$ and $(\tilde{X}_M, \tilde{Y}_M)$ have the same distribution for each $M$, $(X, Y)$ and $(\tilde{X}, \tilde{Y})$ have the same distribution, and $(\tilde{X}_M, \tilde{Y}_M) \to (\tilde{X}, \tilde{Y})$ with probability 1. Therefore,

\[
\lim_{M \to \infty} \log(1 + |X_M|^2) \log(1 + |Y_M|^2) = \log(1 + |X|^2) \log(1 + |Y|^2)
\]

with probability 1. If we can take expectations, (A.11) is proven. From [2, Thm. 25.12], if $\log(1 + |X_M|^2) \log(1 + |Y_M|^2)$ is uniformly integrable for $M \geq 1$, then

\[
\lim_{M \to \infty} E[\log(1 + |X_M|^2) \log(1 + |Y_M|^2)] = E[\log(1 + |X|^2)]E[\log(1 + |Y|^2)].
\]

From [2, p. 338], random variables $Z_M$ are uniformly integrable if

\[
\lim_{a \to \infty} \sup_M \int_{\{|Z_M| \geq a\}} |Z_M| \, dP = 0.
\]
Time shifting diversity as $M \to \infty$

If $\sup_M E[|Z_M|^{1+\epsilon}] < \infty$ for some positive $\epsilon$, then the $Z_M$ are uniformly integrable because

$$\int_{|Z_M| \geq \alpha} |Z_M| \, dP \leq \frac{1}{\alpha^\epsilon} E[|Z_M|^{1+\epsilon}],$$

(A.22)

which goes to zero as $\alpha \to \infty$. Let $Z_M = \log(1 + |\tilde{X}_M|^2) \log(1 + |\tilde{Y}_M|^2)$ and $\epsilon = 1$. Then using the Schwartz inequality,

$$\sup_M E\{[\log(1 + |\tilde{X}_M|^2)]^2[\log(1 + |\tilde{Y}_M|^2)]^2\} \leq \sup_M E[|\tilde{X}_M|^4|\tilde{Y}_M|^4]$$

$$\leq \sup_M (E[|\tilde{X}_M^8|])^{1/2}(E[|\tilde{Y}_M^8|])^{1/2}$$

(A.23)

$$< \infty.$$  

(A.24)

$$< \infty.$$  

(A.25)
Appendix B

Random time weighting as $M \to \infty$

In this section we show that the mutual information of the random time weighting diversity scheme converges in mean square to its expected value as the number of transmit antennas $M$ increases to infinity. The mutual information for the $M$ antenna diversity channel employing random time weighting is given in (2.22). Without loss of generality we assume that $\mathcal{E}_s/N_0 = 1$ and multiply and divide the argument of the logarithm by $M$. Then

$$I_{\text{RAN}} = \int_0^1 \log \left( 1 + \frac{||\alpha||^2}{M} M \eta \right) f_{||\alpha||^2}(\eta) \, d\eta,$$  \hfill (B.1)

where

$$f_{||\alpha||^2}(\eta) = \begin{cases} (M - 1)(1 - \eta)^{M-2} & 0 < \eta \leq 1, \\ 0 & \text{otherwise}, \end{cases} \hfill (B.2)$$

for $M \geq 2$. We show that this mutual information converges to its expected value:

$$\lim_{M \to \infty} \int_0^1 \log \left( 1 + \frac{||\alpha||^2}{M} M \eta \right) f_{||\alpha||^2}(\eta) \, d\eta = E[\log(1 + ||\alpha_1||^2)].$$  \hfill (B.3)

Recall that $|\alpha_1|^2$ is exponentially distributed since $\alpha_1$ is zero mean complex Gaussian. Defining $Y_M = M \eta$ and $X_M = \frac{||\alpha||^2}{M}$ yields

$$I_{\text{RAN}} = \int_0^M \log(1 + X_M Y_M) f_{Y_M}(y) \, dy = E_{Y_M}[\log(1 + X_M Y_M)],$$  \hfill (B.4)
Random time weighting as $M \to \infty$

where

$$f_{Y_M}(y) = \begin{cases} \frac{M-1}{M} \left( \frac{M-y}{M} \right)^{M-2} & 0 < y \leq M, \\ 0 & \text{otherwise}, \end{cases} \quad (B.6)$$

for $M \geq 2$. In the limit as $M \to \infty$, $X_M$ converges to 1 with probability 1 by the strong law of large numbers and $Y_M$ converges in distribution to an exponentially distributed random variable $Y$ with mean $E|\alpha_1|^2$. To prove (B.3) it is sufficient to show that

$$\lim_{M \to \infty} E_{Y_M}[\log(1 + X_M Y_M)] = E_Y[\log(1 + Y)]. \quad (B.7)$$

Since $Y_M$ converges to $Y$ in distribution in the limit as $M \to \infty$, by Skorohad’s Theorem [2, Thm. 25.6], we can construct random variables $\tilde{Y}_M$ and $\tilde{Y}$ such that: $Y_M$ and $\tilde{Y}_M$ have the same distribution for each $M$, $Y$ and $\tilde{Y}$ have the same distribution, and $\tilde{Y}_M \to \tilde{Y}$ with probability 1. Therefore,

$$\lim_{M \to \infty} \log(1 + X_M \tilde{Y}_M) = \log(1 + \tilde{Y}). \quad (B.8)$$

with probability 1. Since $\log(1 + X_M \tilde{Y}_M)$ and $\log(1 + \tilde{Y})$ have the same distributions as $\log(1 + X_M Y_M)$ and $\log(1 + Y)$ respectively, their expectations are equal. Thus, all that remains in proving (B.7) is to show that we can take expectations in (B.8). This requires showing that $\log(1 + X_M \tilde{Y}_M)$ is uniformly integrable. Recall from the time shift diversity proof (Appendix A) that random variables $Z_M$ are uniformly integrable if $\sup_M E[|Z_M|^{1+\epsilon}] < \infty$ for some positive $\epsilon$. Let $Z_M = \log(1 + X_M \tilde{Y}_M)$ and $\epsilon = 1$. Then

$$\sup_M E[|\log(1 + X_M \tilde{Y}_M)|^2] \leq \sup_M E[X_M^2 \tilde{Y}_M^2] \quad (B.9)$$

$$= E[X_M^2] E[\tilde{Y}_M^2] \quad (B.10)$$

$$< \infty, \quad (B.11)$$

where we have used the fact that $X_M$ and $Y_M$ and consequently $X_M$ and $\tilde{Y}_M$ are independent for each $M$. 

- 128 -
Appendix C

Justification of independent circular Gaussian codebooks for each antenna input

In this section, we further justify the use of independent circular Gaussian random variables for the antenna inputs over the unconstrained vector input channel. That is, we explain why random codebooks should be constructed with i.i.d. Gaussian symbols, each symbol a circular complex Gaussian random vector with covariance matrix equal to a scaled identity matrix. I.i.d. Gaussian signaling across antenna elements is shown to be optimal in terms of three performance criteria: (1) maximum expected mutual information, (2) minimum outage probability, and (3) in a game-theoretic sense in which nature chooses the worst possible channel coefficients and the transmitter chooses the best possible signaling scheme. The optimality of i.i.d. Gaussian signaling in maximizing expected mutual information was proven in Chapter 4. We examine performance criteria (2) and (3) below.

Recall from (2.1) that, the complex baseband output of the $M$-antenna channel is

$$Y = \sum_{i=1}^{M} \alpha_i X_i + V,$$

where the total input energy is constrained to $\sum_{i=1}^{M} E|X_i|^2 < \mathcal{E}_s$. The use of codebooks with i.i.d. Gaussian symbols follows from the assumption that the receiver knows the channel. In Section 4.2, we showed that without loss of generality, we can restrict the input to be circular Gaussian when maximizing mutual information. A similar proof can be used to show that we can restrict ourselves to consider only circular Gaussian random variables when minimizing the probability the mutual information falls below $R$. We simply modify Equation (4.16)
Justification of independent circular Gaussian codebooks for each antenna input

with the new criteria. From (4.15) we know $I(\mathbf{X}; Y) \leq \log(1 + \frac{2t}{N_0})$, therefore

$$\min_{\mathbf{X} : \text{tr}(\mathbf{X}) \leq \varepsilon_s} P[I(\mathbf{X}; Y) < R] \geq \min_{\mathbf{X} : \text{tr}(\mathbf{X}) \leq \varepsilon_s} P \left[ \log \left(1 + \frac{2t}{N_0}\right) < R \right]. \quad (C.1)$$

From the discussion in Section 4.2, $t$ is independent of the pseudo-covariance matrix of $\mathbf{X}$, thus we may restrict $\mathbf{X}$ to be circular when minimizing this lower bound. Furthermore, this lower bound (C.1) is satisfied with equality when the input is circular Gaussian. Thus we can restrict the input to be circular Gaussian without loss of generality. Since a circular complex Gaussian input vector $\mathbf{X} = [X_1, X_2, \ldots, X_M]$ is completely described by its covariance matrix $\Lambda_X = E[\mathbf{X} \mathbf{X}^H]$, choosing the optimal input corresponds to choosing the optimizing covariance matrix $\Lambda_X$. I.i.d. signaling corresponds to a scaled identity covariance matrix.

C.1 Minimizing probability of outage

Consider the Rayleigh model for the channel coefficient vector: $\alpha$ a circular complex Gaussian random vector. With an infinite number of antennas, it is possible to communicate at the rate of average mutual information without errors. For a finite number of antennas, error-free communication is not possible. Instead, we are interested in minimizing the probability of outage, which is the probability that the mutual information falls below a specified rate $R$. To maintain small probabilities of outage, communication rates $R$ less than the average mutual information are considered. We conjecture that a Gaussian input with a scaled identity covariance matrix minimizes the probability of outage for all $M$-antenna transmit diversity systems. A direct computation and minimization of the probability of outage shows this result to be true for a small number of antennas.

Minimizing the probability the mutual information falls below rate $R$,

$$\min_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} P(I(\mathbf{X}; Y) < R) = \min_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} P(\log(1 + \frac{\alpha^T \Lambda_X \alpha^*}{N_0}) < R), \quad (C.2)$$

is equivalent to minimizing the probability the SNR falls below $R' = 2^R - 1$,

$$\min_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} P(\frac{\alpha^T \Lambda_X \alpha^*}{N_0} < R'). \quad (C.3)$$
Diagonalizing $\Lambda_X = U^H D U$ and using the fact that multiplying a complex Gaussian random vector by a unitary matrix does not affect its distribution shows that the minimization corresponds to picking optimal diagonal matrix $D$,

$$
\min_{D : \text{tr}(D) \leq \varepsilon} P(\frac{\alpha^T D \alpha^*}{N_0} < R').
$$

(C.4)

Denoting the diagonal elements of $D$ by $\lambda_1, \lambda_2, \ldots, \lambda_M$ yields

$$
\min_{\{\lambda_i\} : \sum_{i=1}^M \lambda_i \leq \varepsilon} P\left(\sum_{i=1}^M \lambda_i |\alpha_i|^2 < R'\right),
$$

(C.5)

where the $|\alpha_i|^2$ are exponentially distributed random variables. The mean of $\sum_{i=1}^M \lambda_i |\alpha_i|^2 = E|\alpha_i|^2$ is independent of the choice of $\lambda_i$ and its variance is minimized by choosing $\lambda_i = \frac{\varepsilon}{M}$. Our choice of $R$ ensures that $R' < E|\alpha_i|^2$. Therefore as $M$ increases, the probability of outage goes to zero if $\lambda_i = \frac{\varepsilon}{M}$. We conjecture that this is the best signaling strategy for low probabilities of outage. For a small number of antennas, $\lambda_i = \frac{\varepsilon}{M}$ can be shown to be optimal by evaluating the first derivative of the probability of outage and showing that the second derivative is negative for all $\lambda_i \in [0, 1]$. We omit the algebra here.

If we attempt to communicate at very high rates $R' > E|\alpha_i|^2$, the probability of outage goes to 1 as the number of antennas increases with $\lambda_i = \frac{\varepsilon}{M}$. For high rates and high probabilities of outage, the optimal antennas moves toward a rank 1 matrix. In this case we want $\sum_{i=1}^M \lambda_i |\alpha_i|^2$ to have a large variance, to ensure that there is some probability of communicating at a rate larger than the mean. Practical communication systems, of course do not operate at such high probabilities of outage.

**C.2 Game theoretic approach**

Consider a scenario in which the transmitter chooses the input distribution to maximize the expected mutual information $I(X; Y, \alpha)$ (rate of reliable communication) and nature chooses the distribution of the channel coefficient vector $\alpha$, for a fixed channel gain $||\alpha||^2 = c$, to
minimize the expected mutual information. We show that

$$\max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} I(\mathbf{X}; Y) = \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} I(\mathbf{X}; Y), \quad (C.6)$$

i.e., the game has a saddle point, and the point is achieved when $\Lambda_X$ is a scaled identity matrix and $f_{\alpha}$ assigns uniform relative phases to the components of $\alpha$. Deviation from these distributions by either the transmitter or nature results in inferior performance from either player’s point of view.

Let $\mathbf{X}$ be a circular complex Gaussian input vector with independent and identically distributed components for each $X_t$, i.e., $\Lambda_X = \frac{\varepsilon_s}{M} I$, where $I$ is an $M \times M$ identity matrix. Then

$$\max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} I(\mathbf{X}; Y, \alpha) \geq \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} I(\mathbf{X}; Y, \alpha) \quad (C.7)$$

$$= \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} E \left[ \log \left( 1 + ||\alpha||^2 \frac{\varepsilon_s}{MN_0} \right) \right] \quad (C.8)$$

$$= \log \left( 1 + c \frac{\varepsilon_s}{MN_0} \right). \quad (C.9)$$

Alternatively, let $\hat{\alpha} = [\sqrt{\frac{c}{M}} e^{j\theta_1}, \ldots, \sqrt{\frac{c}{M}} e^{j\theta_M}]^T$ where the $\theta_i$ are independent and uniformly distributed between $-\pi$ and $\pi$. This yields

$$\min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} I(\mathbf{X}; Y, \alpha) \leq \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} I(\mathbf{X}; Y, \hat{\alpha}) \quad (C.10)$$

$$= \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} E \left[ \log \left( 1 + \hat{\alpha} \Lambda_X \hat{\alpha}^* \right) \right] \quad (C.11)$$

$$= \log \left( 1 + c \frac{\varepsilon_s}{MN_0} \right), \quad (C.12)$$

where (C.11) is maximized by $\Lambda_X = \frac{\varepsilon_s}{M} I$. A diagonal $\Lambda_X$ also ensures that the variance of $\hat{\alpha} \Lambda_X \hat{\alpha}^*$ is zero and there exists a non-zero rate for which we can communicate without errors. The mean of $\hat{\alpha} \Lambda_X \hat{\alpha}^*$ is maximized by transmitting equal power on each antenna. Equations (C.9) and (C.12) show that

$$\max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} I(\mathbf{X}; Y, \alpha) \geq \min_{f_{\alpha}(\alpha) : ||\alpha||^2 = c} \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \varepsilon_s} I(\mathbf{X}; Y, \alpha). \quad (C.13)$$

The inequality in the other direction is a general result for all functions of two random
\[ \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \epsilon_s} \min_{f_\alpha(\alpha) : ||\alpha||^2 = c} I(X; Y, \alpha) \leq \min_{f_\alpha(\alpha) : ||\alpha||^2 = c} \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \epsilon_s} I(X; Y, \alpha). \quad (C.14) \]

Combining the two results above shows

\[ \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \epsilon_s} \min_{f_\alpha(\alpha) : ||\alpha||^2 = c} I(X; Y, \alpha) = \min_{f_\alpha(\alpha) : ||\alpha||^2 = c} \max_{\Lambda_X : \text{tr}(\Lambda_X) \leq \epsilon_s} I(X; Y, \alpha) \quad (C.15) \]

\[ = \log \left( 1 + \frac{\epsilon_s}{MN_0} \right), \quad (C.16) \]

and, thus, the optimal input covariance matrix is a scaled identity matrix.

Suppose nature were allowed to choose the distribution of \( \alpha \) without the gain constraint, but only a power constraint of \( E|\alpha_i|^2 = \epsilon_{\alpha} \). The optimal distribution from nature's point of view is a two-point distribution, for which the probability that \( |\alpha| \) is small is very large and the probability that \( |\alpha| \) is large is small. This distribution ensures that the reliable rate of communication is extremely small. Fortunately, this channel model is not particularly realistic. The max-min problem as described above with the gain constraint shows that for any given channel gain, the best signaling strategy is a scaled identity covariance matrix. Nature's choice of uniform relative phase for the \( \alpha_i \) prevents the transmitter from coherent combining, i.e., "pointing" a beam in the optimal direction, since the optimal direction is unknown to the transmitter.
Bibliography


Bibliography


