Channel Equalization in Wireless Communication using Multiple Transmit Antennas

Mai Hong Vu

A Thesis Submitted
in Partial Fulfillment for the Degree of
Master of Engineering Science

25 August 1999

Department of Electrical and Electronic Engineering
The University of Melbourne
Declaration

The work presented in this thesis was conducted under the guidance of Professor Rob Evans in the Department of Electrical and Electronic Engineering at the University of Melbourne.

I declare that this work is the result of my original research carried out during the time that I enrolled in the Master of Engineering Science degree here, and due acknowledgments have been given for information derived from other published and unpublished work as referenced in the text. This work has not been submitted for any other degree or award at any other University or educational institution.

Mai Hong Vu
25 August 1999
Abstract

In wireless communication, it has been established that the use of multiple antennas, known as spatial diversity, can increase the channel capacity. This is the motivation for the on-going search for processing techniques that exploit spatial diversity. While receive diversity has been studied extensively, transmit diversity has only recently been investigated and is receiving increasing attention. Some transmit diversity techniques have been shown to improve system performance in a Rayleigh flat and slow fading channel as the number of antennas increases.

We study a linear structure using transmit diversity for the single user case, where the transmit signal is processed by a linear prefilter prior to being sent from each transmit antenna, and a linear equalizer is used at the receiver followed by sampling to detect the transmitted symbols. The channel is assumed to be frequency non-selective, and both cases of known and unknown channel coefficients at the receiver are considered. Using the mean square error criterion, the optimum linear equalizer is derived. The optimum set of prefilters for a Rician fading channel is then established, and the system performance is evaluated by a bound on the probability of symbol error, which shows that in both cases, the average error probability decreases to zero exponentially as the number of transmit antennas increases.
Acknowledgments

Firstly, I would like to thank my supervisor Professor Rob Evans for his guidance and constant support throughout the time I worked with him. This support has been especially invaluable during the most difficult times. I am also grateful for time he has dedicated to correct this thesis even in sickness.

Next, I must thank Dr Phillip Musumeci and Dr Barbara LaScala for their supports and the help with settling in for the second time I come back to Australia. Thanks to Phillip for all the encouragements for my study and the uplifting coffees. And I would like to thank Barbara for proof reading my thesis.

Thanks also to Dr Sergei Berezner for the helpful discussions on probability problems related to the work in this thesis.

Last but not least, I would like to thank my family and my friend Long for their supports which have been going a long way.
# Contents

1 Introduction 1

1.1 Motivation and problem description .......................... 1

1.2 Thesis outline .................................................. 4

1.3 Thesis contribution ............................................. 5

2 The Wireless Channel Characteristics, Models and Capacity 7

2.1 Characteristics of multipath fading channel ............... 9

2.1.1 Channel impulse response .................................... 9

2.1.2 Channel power spectra and coherence parameters .......... 11

2.2 Wireless channel models ....................................... 15

2.2.1 Effect of signal characteristics on the channel model .... 16

2.2.2 Statistical models for fading channels ...................... 19

2.2.3 Discrete channel models ...................................... 27

2.2.4 Channel model with multiple antennas ..................... 30

2.3 Information theoretic capacity of wireless channel with multiple antennas 31

2.3.1 Ergodic channel: Capacity as the highest achievable rate 33

2.3.2 Non-ergodic channel: Capacity versus outage probability 36

2.3.3 Capacity for different diversity techniques ................ 38

3 Transmit Antenna Diversity Techniques 48

3.1 Transmit and receive antenna diversity ....................... 49

3.1.1 Basic transmit diversity and receive diversity using optimum combining .................................................. 50
3.1.2 Orthogonal transmit diversity ........................................... 52
3.2 Time shifting ........................................................................ 57
  3.2.1 Signal model ................................................................... 57
  3.2.2 Performance bounds of different equalization techniques ...... 58
3.3 Modulation diversity ................................................................. 62
  3.3.1 Optimization criterion ...................................................... 63
  3.3.2 Minimizing the squared output Euclidean distance variance for
      single symbol inputs .......................................................... 64
  3.3.3 Impact of path propagation delay differences and noisy channel
      estimates ................................................................. 67
3.4 Phase shifting .......................................................................... 68
  3.4.1 Incorporating phase shift with a spread sequence ............... 69
  3.4.2 Combining phase sweeping transmit diversity and channel coding
      ................................................................. 72
3.5 Antenna linear precoding .......................................................... 76
  3.5.1 Linear time-invariant antenna precoding ............................ 77
  3.5.2 Linear periodically time-varying antenna precoding ............ 83
  3.5.3 Summary of linear antenna diversity techniques ............... 87
3.6 Space-time orthogonal coding ..................................................... 88

4 Equalization of A Known Wireless Channel using Multiple Transmit
   Antennas .................................................................................. 93
  4.1 Signal model .......................................................................... 95
    4.1.1 Continuous signal model ............................................... 96
    4.1.2 Discrete signal model .................................................. 97
  4.2 Optimization criterion ............................................................ 100
  4.3 Optimum equalizer ................................................................. 104
  4.4 Minimum mean square error and optimum prefilters ............... 106
  4.5 Evaluating performance by bound on probability of error .......... 110
  4.6 Summary of known channel coefficients case ........................ 114
5 Equalization of An Unknown Wireless Channel using Multiple Transmit Antennas 116

5.1 Optimum equalizer ................................. 117
5.2 Minimum mean square error and optimum prefilters ............................. 119
5.3 Evaluating performance by bound on probability of error ............................. 120
5.4 Summary of the unknown channel coefficients case ............................. 122

6 Conclusion 124

6.1 Thesis summary ........................................ 124
6.2 Further research directions ........................................ 125
# List of Figures

1.1 Linear antenna processing structure. ........................................ 3

2.1 Multipath fading channel. .................................................. 10

2.2 Relationship between delay spread and bandwidth coherence of a channel. 13

2.3 Relationship between Doppler spread and time coherence of a channel. . 14

2.4 Multipath fading channel functions and parameters. ...................... 15

2.5 Equivalent signal model in discrete time. ................................. 28

2.6 Multiple antenna receive diversity. ........................................ 39

2.7 Multiple antenna transmit diversity ........................................ 41

2.8 Time division (a) and frequency division (b) transmit diversities ....... 41

2.9 Time shifting (a) and frequency shifting (b) transmit diversities. ....... 42

2.10 Randomly time weighting transmit diversity. ............................ 43

3.1 Equivalent models for receive OC (a) and transmit (b) diversities ....... 51

3.2 Maximum-likelihood receiver for transmit diversity with orthogonal base-pulses. ................................................................. 56

3.3 Delayed transmit diversity model. .......................................... 58

3.4 Transmit antenna linear modulation diversity. ............................ 63

3.5 Transmit diversity with two branch phase sweeping technique. ........ 74

3.6 Linear time-invariant antenna precoding. .................................. 77

3.7 Linear periodical time-varying antenna precoding ........................ 83

4.1 Linear transmit diversity antenna processing. ............................. 94

4.2 Equivalent signal model in continuous time .............................. 97
4.3 Equivalent signal model in discrete time ....................... 98
4.4 General discrete signal model. ................................ 99
Mathematical Notations

Conventions

\((\cdot)\) continuous in time
\([\cdot]\) discrete in time
* convolution
\(E[\cdot]\) expectation

subscript \(R\) real part
subscript \(I\) imaginary part

superscript \(^{*}\) frequency of a sampled (discrete) signal
superscript \(\dagger\) conjugate transpose
superscript \(^{*}\) complex conjugate
Symbols

\[\begin{align*}
\alpha(t) & \quad \text{time-varying amplitude of path } k \\
\alpha_k & \quad \text{random amplitude of path } k \\
\gamma & \quad \text{instantaneous receive SNR} \\
\delta(\cdot) & \quad \text{the Dirac delta function} \\
(\Delta f)_c & \quad \text{channel coherence bandwidth} \\
(\Delta t)_c & \quad \text{channel coherence time} \\
\epsilon & \quad \text{mean square error} \\
\epsilon_o & \quad \text{optimum MSE} \\
\bar{\epsilon} & \quad \text{average MSE} \\
\hat{\omega}_0 & \quad N_0/\sigma_c^2 \\
\theta_k(t) & \quad \text{time-varying phase of path } k \\
\theta_k & \quad \text{random phase of path } k \\
\mu & \quad \text{mean of a random variable} \\
\mu_a & \quad \text{channel coefficient mean} \\
\rho & \quad \text{average receive signal-to-noise ratio} \\
\rho_0 & \quad \text{transmit SNR} \\
\sigma^2 & \quad \text{variance of a random variable} \\
\sigma^2_a & \quad \text{channel coefficient variance} \\
\sigma^2_c & \quad \text{coded information symbol variance} \\
\sigma^2_N & \quad \text{sampled noise variance} \\
\tau_k(t) & \quad \text{time-varying delay of path } k \\
\tau_k & \quad \text{random delay of path } k \\
\chi^2_\nu & \quad \text{modified chi-squared random variable with } \nu \text{ degrees of freedom} \\
\omega & \quad \text{angular frequency in Radian} \\
 a_k(t) & \quad \text{time-varying complex gain of path } k \\
 a_k & \quad \text{random complex gain of path } k \\
 a_{kl} & \quad \text{imaginary part of the complex path gain } a_k \\
 a_{kR} & \quad \text{real part of the complex path gain } a_k \\
 A & \quad \text{channel coefficient matrix} \\
 B_d & \quad \text{channel Doppler spread} \\
c(\tau; t) & \quad \text{continuous time-varying multipath fading channel impulse response} \\
c[k; l] & \quad \text{discrete multipath fading channel impulse response} \\
C(f; t) & \quad \text{time-varying frequency response of } c(\tau; t)
\end{align*}\]
$C$ channel capacity
$D$ time delay
$f$ frequency in Hz
$f_c$ carrier frequency
$g[k]$ modulation pulse
$h_i[k]$ linear prefilter prior to transmit antenna $i$
$H_i(\omega)$ frequency response of continuous $h_i(t)$
$H_i^*(\omega)$ frequency response of discrete $h_i[k]$
$H$ prefilter tap weight matrix
$I_0(x)$ modified Bessel function of the first kind and order 0
$I_n$ identity matrix of size $n$
$K$ total number of paths
$M$ number of transmit antennas
$n(t)$ continuous time additive noise
$n[k]$ discrete time additive noise
$N$ number of receive antennas
$N_0$ noise power spectrum density
$P$ total transmit power
$p_X(x)$ probability density function of a random variable $X$
$P_b$ probability of bit error
$P_e$ probability of symbol error
$P_W$ probability of word error
$r(t)$ receive signal in continuous time, excluding noise
$r$ general path amplitude
$s(t)$ transmit signal in continuous time
$t$ continuous time
$T$ signaling interval, usually equal a symbol interval
$T_c$ transmission time of the whole message
$T_m$ channel multipath spread
$u(\cdot)$ the unit step function
$u_i[n]$ output of the $i$th prefilter
$w[k]$ noise in discrete time
$W$ signal bandwidth
$x_i$ transmit signal from antenna $i$ in transmit diversity
\( \hat{x}[n] \)  \hspace{1cm} \text{estimate of symbol } x[n] \\
y(t) \hspace{1cm} \text{receive signal in continuous time, including noise} \\
y_i \hspace{1cm} \text{receive signal at antenna } i \text{ in receive diversity}
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>Asynchronous Digital Subscriber Line</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>BPSK</td>
<td>Binary Phase Shift Keying</td>
</tr>
<tr>
<td>DFE</td>
<td>Decision Feedback Equalizer</td>
</tr>
<tr>
<td>DPSK</td>
<td>Differential Phase Shift Keying</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile</td>
</tr>
<tr>
<td>FIR</td>
<td>Finite Impulse Response</td>
</tr>
<tr>
<td>ISI</td>
<td>Intersymbol Interference</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LE</td>
<td>Linear Equalizer</td>
</tr>
<tr>
<td>LTI</td>
<td>Linear Time-Invariant</td>
</tr>
<tr>
<td>LPTV</td>
<td>Linear Periodical Time-Varying</td>
</tr>
<tr>
<td>MFB</td>
<td>Matched Filter Bound</td>
</tr>
<tr>
<td>MLSE</td>
<td>Maximum Likelihood Sequence Estimation</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Square Error</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation and problem description

Personal wireless communication has been a fast growing area recently, with the booming of mobile communications and the emergence of indoor wireless communications such as wireless LAN (local area network). Since the first time that continuous communication by radio was demonstrated between ships sailing the English channel in 1897, wireless communication has matured into a technology that serves common needs, in much the same way as wired communication. After World War II where the use of wireless technology in the military provided many incentives for its growth, mobile communication systems began to find domestic use. But only recently has the boom in cordless appliances and mobile telephones taken place, and since 1994, wireless services have been experiencing customer growth rates exceeding 50% per year. Some even estimate that the number of wireless customers throughout the world will be equal to that of wireline technology in the first couple decades of the 21st century [39].

Wireless communication often faces strict constraints on transmission bandwidth and power. This together with the unpredictably changing nature of the wireless environment makes the transmission rates of wireless communication still far less than that achieved today by wired technology such as optical fiber, ADSL, etc. For example, the popular GSM mobile network used in Europe and some parts of Asia achieves a transmission rate of 270.833 kbps in a 25 MHz bandwidth, whereas the D-AMPS system used in North America offers a bit rate of 48.6 kbps in the same bandwidth, and the Japanese JDC system transmits at 42 kbps [40]. The ultimate capacity, often random, of a wireless channel is a measure of the limits in wireless communication. This wireless
channel capacity is understandably less than that of a wired channel given the same transmit power. However, recent research [10, 11] has found that by using multiple antennas at the transmitter and receiver, the capacity of a wireless channel can be substantially increased, and the increment is almost linear in the number of transmit and receive antennas. These theoretical results suggest that eventually the achievable rate will exceed gigabits per seconds in the common 25 MHz bandwidth found today [10].

A wireless channel is often characterized as having multiple fading paths. These paths are created by reflections, refraction or diffraction of the transmit signal off scattering objects between the transmitter and receiver. Due to the constantly changing nature of the wireless environment, these paths appear random to the user, with their characteristics varying as functions of time, frequency and space. Such multipath fading channels exhibit two important characteristics: frequency selectivity which represents the effects of the channel on different signal frequency components, and fading which represents how fast the channel is changing with time.

Diversity signal processing techniques exploit the existence of different effects of the channel on resolvable signal components within a domain. Results on the theoretical capacity of a wireless channel with multiple antennas strongly motivate the study of spatial diversity, which exploits the fact that different transmit and receive antenna pairs fade independently, provided the antenna spacing is larger than half of the wavelength. At very high-frequency communication in the ranges up to 10 or 20 GHz [4] corresponding to wavelengths of 3 or 1.5 cm, this spacing requirement makes it possible to employ antenna diversity on common wireless appliances such as mobile telephones, laptop computers, personal digital assistants, etc. While receive diversity has been studied extensively ([4], [10], [13] and reference therein), transmit diversity ([20], [21], [23], [22], [15], [17], [18], [12], [16], [25], [24]) and the combination of both ([26], [27], [30]) have only recently been receiving attention. These results have shown that with as few as two antennas at either the transmitter or receiver combined with appropriate signal processing, the communication error rate is significantly decreased.

In this thesis we focus our attention on single user transmit diversity for a channel that does not exhibit frequency selectivity, i.e. the channel has the same effect on all of the signal frequency components. Such channel is also called a flat fading channel and is characterized by a time-varying multiplicative complex coefficient. It is first assumed that the channel coefficients are available at the receiver but not at the transmitter.
Later this assumption is removed and we study the case where the receiver does not know the channel coefficients.

In a Rayleigh flat and relatively slow fading channel where the complex channel coefficients are modeled as Gaussian random variables with zero mean, information theoretic results show that with increasing number of transmit antennas, a random fading channel capacity approaches almost surely the capacity of a non-fading additive white Gaussian noise channel with the same average receive signal-to-noise (SNR) ratio. This suggests that the use of multiple transmit antennas is to mitigate fading in the channel. Existing transmit diversity techniques have investigated processing schemes such as linear prefiltering the transmit signals from different antennas ([23], [22], [17], [18]) which includes time shifting ([20], [21]), phase shifting ([25], [24]) as particular cases, coupled with various types of equalizers including maximum likelihood sequence estimation, decision feedback equalizer and linear equalizer, or combining antenna diversity with coding ([30], [27]) to achieve diversity gain. Amongst these, linear structures appear attractive due to their simplicity in implementation while still achieving diversity gains. Results show that with certain linear prefiltering of the data before transmitting from multiple transmit antennas, and with linear equalizer at the receiver, the channel is asymptotically transformed from a Rayleigh flat and slow fading channel into a non-fading additive white marginally Gaussian noise channel, and the asymptotic bit error rate for QPSK (quadrature phase shift keying) falls off exponentially with increasing SNR [18].

![Diagram](image)

**Figure 1.1: Linear antenna processing structure.**

These results motivate our work to further investigate this linear structure of using linear modulation and having linear prefilters prior to each transmit antenna and a linear equalizer at the receiver, which we call **linear antenna processing**, as depicted
in Figure 1.1. Whereas previous results only provide the optimum linear equalizer in the asymptotic case of an infinite number of transmit antennas, we determine the optimum linear equalizer given any finite number of transmit antennas using a mean square error (MSE) criterion. We then analyze the optimum linear prefilters and error rate for Rician flat fading channels. In a Rician flat fading channel, the multiplicative channel coefficients are modeled as complex Gaussian random variables with non-zero mean. We also extend our problem further to include both cases of known and unknown channel coefficients at the receiver.

1.2 Thesis outline

This thesis is organized into 4 main chapters following this introductory chapter as below.

Chapter 2 provides a background and overview of wireless channel multipath fading characteristics and common models that are used for a wireless channel. From these we build discrete signal models that will be used in later analysis. We then present the information theoretic results on capacity of channels with multiple antennas, together with the capacity of various receive and transmit diversity schemes.

Chapter 3 analyzes transmit diversity in Rayleigh flat and slow fading channels and compares this with receive diversity. This is then followed by a comprehensive review of transmit diversity techniques for a Rayleigh flat slow fading channel. Through the review, we point out the connections between these diversity techniques and draw an overall picture on linear transmit diversity in Rayleigh flat and slow fading channels.

Chapter 4 presents our new results on the linear antenna processing structure for the case of known channel coefficients at the receiver, assuming that linear modulation is used. We derive the optimum linear equalizer based on a mean square error optimization criterion. We then establish the optimum set of prefilters for Rician fading channels and evaluate the structure in terms of probability of symbol error for a rectangular QAM signal, which shows that the error rate approaches zero exponentially as the number of transmit antennas increases.
Chapter 5 presents our new work for the case of unknown channel coefficients at the receiver. We assume, however, that the mean and variance of these channel coefficients are available. Using the same linear structure as in Chapter 4, we study the optimum linear equalizer and optimum set of linear prefilters for Rician fading channels. Performance is also evaluated in terms of symbol error probability for a rectangular QAM signal, and a similar error bound to the known channel case is obtained. This highlights the effectiveness of transmit diversity in Rician flat fading channels even when the channel is not fully known at the receiver.

Finally we state our conclusions and future directions in Chapter 6. Mathematical proofs of some results stated in Chapters 4 and 5 can be found in the Appendix.

1.3 Thesis contribution

The contributions of this thesis can be summarized in two parts. The first part is a comprehensive review of current transmit diversity techniques for Rayleigh flat and slow fading channels in Chapter 3. The review provides an integrated framework connecting mostly independently derived results. Using an optimization criterion of minimizing the receive power variance or equivalently the normalized variance of square Euclidean distance between two output sequences corresponding to single symbol inputs, it was shown that the optimum set of linear prefilters for transmit diversity is when the tap coefficients of these filters form the rows of a unitary matrix [22]. This includes other schemes of time shifting [21, 20] and phase shifting [25] as special cases. Adopting this set of prefilters in the linear antenna processing structure, it was shown that a Rayleigh flat and slow fading channel is asymptotically transformed into a non-fading AWGN channel [18].

The second and major part of the contribution includes the derivation of the optimum equalizer for the case of a known channel at the receiver given any number of transmit antennas, followed by a novel analysis of the optimum linear prefilters and error rate in the Rician flat fading channel case for both known and unknown channels. Results for the known channel case are presented in Chapter 4. The optimum linear equalizer based on the minimum mean square error criterion given any number of transmit antennas for both Rayleigh and Rician fading channels is given in Theorem 4.2. The set of optimum linear prefilters for Rician flat fading channels is given in Theorem 4.3, which shows that the MSE converges in the mean square sense to its
minimum average lower bound as the number of transmit antennas increases to infinity, implying the convergence of symbol estimates to the symbol values in the 4th mean. An upper bound on the probability of symbol error for rectangular QAM signals is given in Theorem 4.4, which shows that with the optimum equalizer and prefilters, the average error probability for Rician fading channels decreases to zero exponentially as the number of transmit antennas increases.

For the unknown channel case, results are presented in Chapter 5. We assume that the mean and variance of the channel coefficients are available at the receiver and the analysis is applicable for Rician flat fading channels. The optimum linear equalizer based on a mean square error criterion is given by Theorem 5.1. Analysis of the set of optimum linear prefilters can be found in Sections 5.2, which results in the same optimum prefilter set as in the known channel case. It is shown that the average MSE approaches zero as the number of transmit antennas increases, implying convergence of symbol estimates to the symbol values in the mean square sense. An error bound for rectangular QAM signals is given in Theorem 5.2. With the optimum equalizer and prefilters, the error bound turns out to be equal to that of the known channel case. Thus the average probability of symbol error also decreases exponentially to zero with increasing number of transmit antennas.

The results in Chapters 4 and 5 show that transmit diversity is very effective in a Rician flat fading channel in both cases of known or unknown channels at the receiver. Even though the convergence of symbol estimates to the symbol values in the known channel case is stronger, by increasing the number of transmit antennas, the symbol error probability is driven to zero exponentially in both cases.
Chapter 2

The Wireless Channel

Characteristics, Models and Capacity

This chapter provides an overview of the wireless channel and the common mathematical models used to describe it. These form the basis for development of signal processing techniques in the later chapters. We then present information theory results on the capacity of a channel with multiple antennas at the transmitter and receiver.

Due to the environment dependent changing nature of a wireless channel, it is often characterized as a time-varying channel having multiple fading paths. Two important parameters of a time-varying multipath fading channel are the coherence bandwidth and the coherence time. The coherence bandwidth represents the maximum frequency separation at which the frequency-domain channel responses at two frequency remains strongly correlated [2]. This means that signal components differing in frequencies by more than the channel coherence bandwidth will experience different effects by the channel, which is called frequency selectivity of the channel. Similarly, the coherence time of a channel is a measure of the maximum time separation that the channel impulse responses at two time instants still remain strongly correlated. It characterizes how fast the channel is changing with time, which is the phenomenon that creates the fluctuation in the received signal and is termed fading. Thus frequency selectivity and fading are two key measures in characterizing a time-varying multipath fading channel.

From these characteristics of a wireless channel, mathematical models that describe
the channel are built. The channel response is modeled as a sum of time varying complex gains at different path delays, where the time varying complex gains represent the fading and different path delays represent the frequency selectivity of the channel. In a frequency non-selective channel, the channel response is reduced to a multiplicative complex time-varying gain. For relatively slow fading channels, a channel model often found in the literature is the quasi-stationary model where the channel is assumed to remain time-invariant for at least a signaling interval, but change randomly at the next one. Since the changes in wireless channel response parameters (i.e., the path gains, path delays and number of paths) are usually unpredictable in time, these parameters are model statistically. The common statistical distributions used to model a path gain are Rayleigh and Rician distributions, in which the complex path gain has a Rayleigh or Rician distributed magnitude and usually a uniformly distributed phase. These two models are equivalent to modeling path gains as complex Gaussian random variables with independent Gaussian real and imaginary parts having the same variance, and zero mean in the case of Rayleigh fading, non-zero mean in the case of Rician fading. This complex Gaussian model of channel gain will be used extensively in this thesis.

Due to the fading characteristic of a wireless channel, the channel capacity is less than that of an additive white Gaussian noise (AWGN) channel, and it is usually random. However, recent research [10, 11] has found that using multiple antennas at the transmitter and receiver, which is known as spatial diversity, can substantially increase a wireless channel capacity. It has been established that having spatial diversity with an equal number of antennas at both the transmitter and receiver increases the channel capacity almost linearly with the number of antennas, whereas receive diversity only will increase the capacity logarithmically, and transmit diversity only will asymptotically drive a fading channel capacity to the capacity of an AWGN channel. These results serve as the motivation for exploiting spatial diversity. While receive diversity has been study extensively, transmit diversity has only recently been investigated [20, 21, 22, 18, 25, 24], as well as the combination of both transmit and receive diversities [26, 30, 27]. The below-maximum-capacity current transmit and combined diversity techniques inspire further research into these areas.

This chapter is organized as follows: Section 2.1 analyzes the characteristics of a multipath fading channel, which generalizes the wireless channels. In Section 2.2, several common channel models are developed in both continuous and discrete time. Finally, Section 2.3 presents the information theoretic results on capacity of wireless
channels with multiple antennas, and also on the capacity of some existing spatial diversity techniques.

2.1 Characteristics of multipath fading channel

A wireless channel is often characterized as a multipath fading channel with a complex and time-varying impulse response. The multipath arises from scattering, reflection, refraction or diffraction of the radiated energy off objects in the environment. Propagation path loss and fading due to moving or random objects in the environment between the transmitter and receiver causes the time variation in the channel. This time variation also depends on the speed of the moving objects, either the transmitter or receiver, where a mobile channel is more likely to experience greater time variation than an indoor fixed wireless channel. In this section, we will study multipath fading channel impulse response, power spectra and channel parameters. These will form the basis for establishing channel models in the next section.

2.1.1 Channel impulse response

A multipath fading channel is characterized by a complex time-varying channel impulse response $c(\tau; t)$ (Figure 2.1). When an ideal impulse is transmitted over a multipath fading channel, there will be two effects on the received signal. Firstly, since the signal may follow different paths with different lengths and attenuation factors, the received signal may appear as a train of pulses with different delays and magnitudes. The second effect is due to the time varying nature of the channel, which means the nature of the multipath is varying with time. Thus the number of pulses, the delay between them and their magnitude may change from time to time as the experiment of sending the impulse is repeated. This time variation, moreover, appears to be unpredictable to the user of the channel. Therefore a time-varying multipath fading channel is often characterized statistically.

First, we will examine the effect of the channel on a modulated transmitted signal that is represented generally as

$$s(t) = Re\left[s_i(t)e^{j2\pi f_c t}\right]$$

(2.1)

where $f_c$ is the carrier frequency and $s_i(t)$ is the lowpass information carrying signal. Assuming that there are multiple propagation paths, each with a propagation delay
Figure 2.1: Multipath fading channel.

and an attenuation factor which are both time variant, the received bandpass signal without noise may be expressed as

\[
r(t) = \sum_{k=1}^{K} \alpha_k(t)s(t - \tau_k(t))
\]  

(2.2)

where \( K \) is the total number of paths, \( \alpha_k(t) \) is the \( k^{th} \) path attenuation factor and \( \tau_k(t) \) is \( k^{th} \) path propagation delay. Substituting \( s(t) \) in (2.1) yields

\[
r(t) = \text{Re}\left\{ \sum_{k=1}^{K} \alpha_k(t)e^{-j2\pi f \tau_k(t)} s_1(t - \tau_k(t)) \right\} e^{j2\pi f t}.
\]

It is apparent from this equation that the equivalent low pass received signal is

\[
r_l(t) = \sum_{k=1}^{K} \alpha_k(t)e^{-j2\pi f \tau_k(t)} s_1(t - \tau_k(t)).
\]

Since the channel bandwidth is generally much smaller than the carrier bandwidth, the system can be effectively modeled by equivalent lowpass signals through equivalent lowpass channels [1]. Since \( r_l(t) \) is the response of an equivalent lowpass channel to the equivalent lowpass signal \( s_l(t) \), it follows that the equivalent lowpass channel can be described by the time-varying impulse response

\[
c(\tau; t) = \sum_{k=1}^{K} \alpha_k(t)e^{-j2\pi f \tau_k(t) \delta(t - \tau_k(t))} = \sum_{k=1}^{K} \alpha_k(t)e^{j\theta_k(t) \delta(t - \tau_k(t))}
\]

(2.3)

where \( \theta_k(t) = -2\pi f \tau_k(t) \) is a time-varying phase sequence. This \( c(\tau; t) \) represents the response of the channel at time \( t \) due to an impulse applied at time \( t - \tau \). The channel is completely characterized by the number of multipath components \( K \) and the path variables: amplitude \( a_k(t) \), delay \( \tau_k(t) \) and phase \( \theta_k(t) \). These parameters change unpredictably with time and are often described statistically. The received signal \( r_l(t) \) therefore is also random, and when there are a large number of paths, the central limit theorem applies. This means \( r_l(t) \) may be modeled as a complex-valued Gaussian random process. Thus the channel impulse response \( c(\tau; t) \) is a complex-valued Gaussian random process in the \( t \) variable. The statistical models are described in more detail in the next section.
Large dynamic changes in the transmitting medium are required for the $\alpha_k(t)$ to change sufficiently to cause a significant change in the received signal. On the other hand, $\theta_k(t)$ will change by $2\pi$ radians whenever $\tau_k(t)$ (or in effect the path length) is changed by $1/f_c$, which is a small amount due to large carrier bandwidth. Therefore $\theta_k(t)$ can change quite rapidly with relatively small motions of the medium. This time variation of the phases $\{\theta_k(t)\}$ is the primary cause of fading phenomena in a multipath channel. The randomly time-varying phases $\{\theta_k(t)\}$ associated with the vectors $\{\alpha_k(t)e^{j\theta_k(t)}\}$ at times result in the received vectors adding constructively or destructively. This causes amplitude variations in the received signal and is termed signal fading.

The lowpass received signal $y(t) = r(t) + n(t)$ therefore becomes

$$y(t) = c(\tau; t) * s(t) + n(t) = \int_{-\infty}^{\infty} c(\tau; t) s(t - \tau)d\tau + n(t) \quad (2.4)$$

where $n(t)$ is lowpass complex-valued additive Gaussian noise.

2.1.2 Channel power spectra and coherence parameters

To see the effect of signal characteristics on a channel model, we describe a number of correlation functions and power spectral density functions that define the characteristics of a fading multipath channel. We begin with an equivalent lowpass channel impulse response $c(\tau; t)$, which is characterized as a complex-valued random process and is assumed to be wide-sense stationary, (i.e. the impulse response has a constant mean and its autocorrelation function depends only on the time difference but not the absolute time). We define the autocorrelation function of $c(\tau; t)$ as:

$$\phi_c(\tau_1, \tau_2; \Delta t) = E[c^*(\tau_1; t)c(\tau_2; t + \Delta t)] .$$

In most radio transmission media, the attenuation and phase shift of the channel associated with different paths are uncorrelated, which is usually called uncorrelated scattering [1]. With this assumption, the channel response associated with path delay $\tau_1$ and the channel response associated with path delay $\tau_2$ are uncorrelated, and the above equation becomes:

$$E[c^*(\tau_1; t)c(\tau_2; t + \Delta t)] = \phi_c(\tau_1; \Delta t)\delta(\tau_1 - \tau_2) . \quad (2.5)$$

Similarly in the frequency domain, we have the time-varying transfer function
$C(f; t)$ as the Fourier transform of $c(\tau; t)$

$$C(f; t) = \int_{-\infty}^{\infty} c(\tau; t)e^{-j2\pi f \tau} d\tau \, .$$  \hspace{1cm} (2.6)$$

Under the wide sense stationary assumption, define the autocorrelation function

$$\Phi_C(f_1, f_2; \Delta t) = E[C^*(f_1; t)C(f_2; t + \Delta t)] \, .$$

It can be shown that $\Phi_C(f_1, f_2; \Delta t)$ is related to $\phi_c(\tau_1, \tau_2; \Delta t)$ by the Fourier transform

$$\Phi_C(f_1, f_2; \Delta t) = \int_{-\infty}^{\infty} \phi_c(\tau_1; \Delta t)e^{-j2\pi f_1 \tau_1} d\tau_1 \equiv \Phi_C(\Delta f; \Delta t)$$  \hspace{1cm} (2.7)$$

where $\Delta f = f_2 - f_1$. Furthermore, the assumption of uncorrelated scattering implies that the autocorrelation function of $C(f; t)$ in frequency is a function of only the frequency difference $\Delta f = f_2 - f_1$. Therefore $\Phi_C(\Delta f; \Delta t)$ is called the spaced-time spaced-frequency correlation function of the channel.

**Delay power spectrum and coherence bandwidth**

First we will study the effect of path delays on the channel characteristics. If we let $\Delta t = 0$ and $\tau_1 = \tau_2 = \tau$ in (2.5), the resulting autocorrelation function $\phi_c(\tau; 0) \equiv \phi_c(\tau)$ is simply the average power output of the channel as a function of the time delay $\tau$. Thus the function $\phi_c(\tau)$ is termed multipath intensity profile or the delay power spectrum of the channel. The range of values of $\tau$ over which $\phi_c(\tau)$ is essentially non-zero is termed the multipath spread of the channel and is denoted by $T_m$. With $\Delta t = 0$ in (2.7), the transform relationship is simply

$$\Phi_C(\Delta f) = \int_{-\infty}^{\infty} \phi_c(\tau)e^{-j2\pi \Delta f \tau} d\tau \, .$$  \hspace{1cm} (2.8)$$

Since $\Phi_C(\Delta f)$ is an autocorrelation function in the frequency variable, it provides us with a measure of the frequency cohesion of the channel. As a result of (2.8), the reciprocal of the multipath spread is a measure of the coherence bandwidth of the channel

$$(\Delta f)_c \approx \frac{1}{T_m}$$  \hspace{1cm} (2.9)$$

where $(\Delta f)_c$ denotes the coherence bandwidth. This means that two sinusoids (single tone frequency signal) with frequency separation greater than $(\Delta f)_c$ are affected
differently by the channel. Thus when the signal bandwidth is larger than \((\Delta f)_c\), different frequency components of the signal will experience different amounts of fading and the signal is severely distorted by the channel. In this case the channel is said to be \textit{frequency selective}. On the other hand, if the signal bandwidth is small compared to \((\Delta f)_c\), the channel appears to be flat in the frequency spectrum and is termed a \textit{frequency nonselective} or \textit{flat channel}.

![Spaced-frequency correlation function](image1)

\[ |\Phi_C(\Delta f)| \]

\[ \Phi_C(\Delta f) \]

\[ \Delta f \]

\[ \left((\Delta f)_c \approx \frac{1}{T_m}\right) \]

\[ \left(\Phi_C(\Delta t) \right) \]

\[ \Phi_C(\tau) \]

\[ \tau \]

\[ 0 \]

\[ T_m \]

\[ \text{Figure 2.2: Relationship between delay spread and bandwidth coherence of a channel.} \]

\textbf{Doppler power spectrum and coherence time}

We will focus now on the time variation of the channel as measured by the parameter \(\Delta t\) in \(\Phi_C(\Delta f; \Delta t)\). Define the Fourier transform of \(\Phi_C(\Delta f; \Delta t)\) with respect to variable \(\Delta t\) to be

\[ S_C(\Delta f; \lambda) = \int_{-\infty}^{\infty} \Phi_C(\Delta f; \Delta t) e^{-j2\pi \lambda \Delta t} d\Delta t \ . \quad (2.10) \]

With \(\Delta f = 0\) and denoting \(S_C(0; \lambda) \equiv S_C(\lambda)\), the above relation becomes

\[ S_C(\lambda) = \int_{-\infty}^{\infty} \Phi_C(\Delta t) e^{-j2\pi \lambda \Delta t} d\Delta t \ . \quad (2.11) \]

The range of values of \(\lambda\) over which \(S_C(\lambda)\) is essentially non-zero is called the \textit{Doppler spread} \(B_d\) of the channel. It represents the spectral broadening of the transmitted signal due to the time variation in the channel. From (2.11), we observe that if the channel is time-invariant, that is \(\Phi_C(\Delta t) = 1\), then \(S_C(\lambda) \equiv \delta(\lambda)\), which means there is no spectral broadening observed in transmitting a pure tone. The function
$S_C(\lambda)$ is a power spectrum that gives the signal intensity as a function of the Doppler frequency $\lambda$ and hence it is called the *Doppler power spectrum* of the channel.

Since $S_C(\lambda)$ is related to $\Phi_C(\Delta t)$ by the Fourier transform, the reciprocal of $B_d$ is a measure of the coherence time of the channel. That is

$$\langle \Delta t \rangle_c \approx \frac{1}{B_d} \quad (2.12)$$

where $\langle \Delta t \rangle_c$ denotes the **coherence time** of the channel. Coherence time represents the time separation during which the channel impulse response at two time instances are still strongly correlated, and therefore it measures how fast a channel changes in time. A slowly varying channel will have a large coherence time, or equivalently, a small Doppler spread, and vice versa.

![Diagram](image)

**Figure 2.3:** Relationship between Doppler spread and time coherence of a channel.

**The scattering function**

We have shown a Fourier transform relationship between $\phi_c(\tau; \Delta t)$, $\Phi_C(\Delta f; \Delta t)$ and a Fourier transform relationship between $\Phi_C(\Delta f; \Delta t)$ and $S_C(\Delta f; \lambda)$. Now, define a new function $S(\tau; \lambda)$ as

$$S(\tau; \lambda) = \int_{-\infty}^{\infty} \phi_c(\tau; \Delta t) e^{-j2\pi \lambda \Delta t} d\Delta t .$$

Then from (2.7) and (2.10), the following Fourier transform relation is apparent

$$S(\tau; \lambda) = \int_{-\infty}^{\infty} S_C(\Delta f; \lambda) e^{j2\pi \lambda \Delta f} d\Delta f .$$
Furthermore, \( S(\tau; \lambda) \) and \( \Phi_C(\Delta f; \Delta t) \) are related by the double Fourier transform

\[
S(\tau; \lambda) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi_C(\Delta f; \Delta t) e^{-j2\pi \lambda \Delta t} e^{-j2\pi \lambda \Delta \Delta f} d\Delta f d\Delta t.
\]

The function \( S(\tau; \lambda) \) is called the scattering function of the channel. It provides a measurement of the average power of the channel as a function of time delay \( \tau \) and the Doppler frequency \( \lambda \).

The relationship between the channel correlation and power spectrum functions and the coherence parameters can be summarized in figure 2.4.

![Diagram showing relationships between channel correlation and power spectrum functions and coherence parameters](image)

Figure 2.4: Multipath fading channel functions and parameters.

### 2.2 Wireless channel models

Having established the power spectrum functions and coherence parameters in the previous section, we now proceed to building the mathematical models for a wireless channel that can be used to study and analyze the performance of a particular system. First we will study the effect that the transmitted signal has on establishing a channel model. Since the channel is often characterized by its statistical properties, we will then present the common distributions used to model fading path amplitude, phase and delay time with some insights into the theoretical explanation and empirical justification of
these models. Finally, we develop discrete channel models for channel with multiple antennas, and these will be the models used in the remainder of this thesis.

2.2.1 Effect of signal characteristics on the channel model

Here we consider the effect of signal characteristics on the appropriate channel model. As before, let \( s_i(t) \) be the equivalent lowpass signal transmitted over the channel and \( S_i(f) \) be its frequency content. Then the equivalent lowpass received signal, exclusive of additive noise, can be written as

\[
r_i(t) = c(\tau; t) * s_i(t) = \int_{-\infty}^{\infty} c(\tau; t) s(t - \tau) d\tau
\]

or in terms of frequency spectrum

\[
r_i(t) = \int_{-\infty}^{\infty} C(f; t) S_i(f) e^{j2\pi ft} df.
\]

The channel has two distortion effects on the transmitted signal. The first effect is caused by multipath spread, or equivalently, the coherence bandwidth of the channel. If the bandwidth \( W \) of the signal is larger than the channel coherence bandwidth \( (\Delta f)_c \), then \( S_i(f) \) is subjected to different gains and phase shift across the band, and the channel is termed frequency selective. The other effect is due to the time variation in the channel, which is grossly characterized by the time coherence \( (\Delta t)_c \), or equivalently, the Doppler spread \( B_d \). This type of distortion results in a variation of received signal strength and is termed fading. Thus frequency selectivity and fading are two different types of distortion in a multipath time-varying channel.

The effect of the channel on the transmitted signal depends on our choice of signal bandwidth and signal duration. The rapidity of fading of the channel is characterized by the parameter time coherence \( (\Delta t)_c \) or Doppler spread \( B_d \). If the signal duration satisfies \( T \ll (\Delta t)_c \) or equivalently, \( W \gg B_d \), the channel is called slow fading, which means that the channel attenuation and phase shift are essentially fixed for the duration of at least one signaling interval. Otherwise the channel is called fast fading where each use of the channel can result in a different attenuation and phase shift of the transmitted signal. Reference [2] has diagrams of channel frequency responses for two different mobile systems which exhibit almost opposite characteristics with regard to frequency selectivity and time variation in the channel.
Suppose that it is possible to select the signal bandwidth $W \ll (\Delta f)_c \approx 1/T_m$ and the signaling interval $T$ such that $T \ll (\Delta t)_c \approx 1/B_d$, i.e., the channel is frequency nonselective and slow fading. Then when the condition $W \approx 1/T$ holds, this implies that $T_m$ and $B_d$ must satisfy $T_m B_d < 1$. The product $T_m B_d$ is called the spread factor of the channel. If $T_m B_d < 1$, the channel is said to be underspread, otherwise it is overspread.

**Frequency Nonselective Channel**

This is the case where the signal duration $T \gg T_m$, the delay spread of the channel. The time delays of different paths are much less than the signal duration, so that the signal on these paths cannot be resolved as distinct pulses. These unresolvable paths add vectorially and are treated as a single path in analysis. All of the frequency components in $S_i(f)$ undergo approximately the same attenuation and phase shift during transmission, and the channel introduces a negligible amount of intersymbol interference. This also implies that the time-variant transfer function $C(f; t)$ of the channel is a complex-valued constant in the frequency variable within the bandwidth of $S_i(f)$. Since $S_i(f)$ has its frequency components concentrated around the vicinity of $f = 0$, $C(f; t) = C(0; t)$. Consequently, in a frequency nonselective channel, the equivalent lowpass received signal (2.13) becomes

$$r_i(t) = C(0; t) \int_{-\infty}^{\infty} S_i(f)e^{j2\pi ft}df = C(0; t)s_i(t)$$

(2.14)

i.e. the received signal is simply the transmitted signal multiplied by a complex-valued random process, which represents the time variation of the channel. Let $C(0; t) = a(t) = \alpha(t)e^{j\theta(t)}$ then, in this case, the envelope of the channel impulse function is simply the magnitude of the channel multiplicative gain $\alpha(t)$. Both $\alpha(t)$ and $\theta(t)$ are random processes.

**Frequency Selective Channel**

The frequency selective channel is modeled as a tapped delay line. In this case, the signal bandwidth $W$ is greater than the channel coherence bandwidth $(\Delta f)_c$, thus the frequency components of the transmitted signal $S(f)$ with separation greater than $(\Delta f)_c$ will experience different gains and phase shifts. When $W \gg (\Delta f)_c$, the multipath components in the channel response that are separated in delay by at least $1/W$ are
resolvable. The Sampling Theorem may be applied to represent the resolvable received signal components. Assuming that there exists a resolvable path in every contiguous frequency interval $W$, the time-varying channel impulse response can be written as [1]

$$c(\tau; t) = \sum_{k=1}^{K} a_k(t) \delta(\tau - k/W)$$

and the corresponding transfer function is

$$C(f; t) = \sum_{k=1}^{K} a_k(t)e^{-j2\pi fk/W}$$

where $a_k(t)$ is the complex-valued channel gain of the $k^{th}$ multipath component and $K$ is the number of resolvable multipath components. Since the multipath spread is $T_m$, it follows that the number of practically resolvable paths is

$$K = \lfloor T_m W \rfloor + 1$$

The tap gains $\{a_k(t)\}$ are usually modeled as wide-sense stationary, mutually uncorrelated random processes due to uncorrelated scattering.

**Quasi-stationary channel model**

Recall from (2.3) the channel impulse response is of the form

$$c(\tau; t) = \sum_{k=1}^{K} \alpha_k(t)e^{j\theta_k(t)} \delta(t - \tau_k(t))$$

where $K$ is the total number of paths and $\alpha_k(t)$, $\theta_k(t)$ and $\tau_k(t)$ are the time-varying amplitude, phase and time delay of each path $k$ respectively. The time invariant version of this model, first suggested by Turin and was analyzed by Hashemi [3], has been used successfully in wireless communication. The channel reduces to

$$c(\tau; t) = \sum_{k=1}^{K} \alpha_k e^{j\theta_k} \delta(t - \tau_k) = \sum_{k=1}^{K} a_k \delta(t - \tau_k)$$

(2.15)

where $a_k = \alpha_k e^{j\theta_k}$ is the complex-valued response of path $k$. This model is reasonably mathematically tractable and is quite suitable for a quasi-stationary (slow fading) channel, i.e. a channel which is slowly time-varying where the response remains relatively time-invariant for at least a signaling interval. Thus the channel can be measured using some training sequence or blind estimation algorithm and the measurement result can be used to transmit the signal until the next signaling interval, when the channel is remeasured and channel parameter values are updated. Although the impact of noisy
channel estimation is not analyzed in this thesis, existing analytical results [18] and
simulation [22] have shown that most of the diversity techniques presented do not dis-
play sensitivity to small inaccuracies in channel measurement as exhibited by some
other coherence combining (e.g. beamforming) strategies.

This model will be used extensively throughout this thesis. The complex path gains
\( \alpha_k \) with amplitude \( \alpha_k \) and phase \( \theta_k \), and path time delay \( \tau_k \) are random variables. In
the next section, we will establish the common statistical distributions used to model
these path parameters.

### 2.2.2 Statistical models for fading channels

Since there are usually a large number of scatterers that contribute to the signal at
the receiver in a wireless channel, the Central Limit Theorem can be applied to model
the channel response as a complex-valued Gaussian random process. Although this is
still an approximation, the Gaussian model is successful in predicting the measured
statistics of the signal to good accuracy in most cases over the ranges of interest for
the variables involved, and thus its use is justified [4].

#### Distribution of path amplitudes

As previously pointed out, the paths that have a delay spread smaller than the inverse
of the channel coherence bandwidth are not resolvable and are therefore combined as a
single path and the envelope of the combined path is observed. Thus if the path delay
difference satisfies \( |\tau_{ij} - \tau_{ik}| < 1/(\Delta f)_c \) for \( i, j = 1, 2, \ldots, p \) then the single resolvable
path

\[
a_k = \alpha_k e^{i\theta_k} = \sum_{i=1}^{p} \alpha_k e^{i\theta_{ki}}
\]

will be used to represent all of these \( p \) paths. The envelope \( \alpha_k \) is a random variable.
To simplify notation, denote \( r = \alpha_k \) for any resolvable equivalent path \( k \). A number
of statistical distributions have been used in modeling this amplitude \( r \) of a fading
channel.

**Rayleigh distribution**

In an environment where there are many scattering objects and there is no direct
path between the transmitter and receiver, the channel impulse response is usually
modeled as a zero mean Gaussian random process. The envelope \( r \) of the channel
response at any time has a Rayleigh probability density function

\[ p_R(r) = \frac{2r}{\delta^2} e^{-r^2/\delta^2} \cdot u(r) \]  
(2.17)

where \( \delta^2 = E(r^2) \) and \( u(r) \) is the unit step function (i.e., \( u(r) = 1 \) if \( r > 0 \) and \( u(r) = 0 \) if \( r < 0 \)). The mean and variance of \( r \) are \( \sqrt{\pi/2} \sigma \) and \((2 - \pi/2)\sigma^2 \) respectively.

The Rayleigh model is often used because of its theoretical tractability and empirical justification. The theoretical basis follows from a model by Clarke, which is described in Hashemi [3]. Assume that a transmitted signal arrives at the receiver by \( p \) paths, where the \( i^{th} \) path signal has complex strength \( r_i e^{j\theta_i} \). All the received paths are then added vectorially at the receiver. In this model, over a small area and in the absence of a line-of-sight path, the \( r_i \) are assumed to be approximately equal, therefore the received signal is

\[ r e^{j\theta} = r' \sum_{i=1}^{p} e^{j\theta_i} \]

where \( r' \) is the common received signal amplitude of the different paths. At high frequency, the path phase \( \theta_i \) is very sensitive to change in the path length since the phase changes by \( 2\pi \) when the path length changes by a wavelength. Thus the phases \( \theta_i \) are modeled as independent uniformly distributed random variables over \([0, 2\pi)\). The problem then becomes that of finding the distribution of the envelope and phase of the sum of many sinusoids with constant amplitude \( r' \) and uniformly distributed random phases \( \theta_i \). The real and imaginary component (I and Q) of the received signal

\[ I = r' \sum_i \cos \theta_i \quad , \quad Q = r' \sum_i \sin \theta_i \]

can be shown to be uncorrelated with zero mean and the same variance, and by the Central Limit Theorem, they are Gaussian distributed. Therefore the distributions of the amplitude \( r = \sqrt{T^2 + Q^2} \) and the combined phase \( \theta = \arctan(I/Q) \) are Rayleigh and uniform respectively, based on well established results which can be found in [6].

Obviously the assumption that the attenuation of all paths are equal is unrealistic. However, it has been shown that even when the magnitudes of the paths are not equal but there is no single one of them which contributes a major fraction of the received power (i.e., \( r_i^2 \ll \sum_i r_i^2 , \forall i = 1 \ldots p \)) then the Rayleigh distribution can still be used to describe the channel amplitude. The Rayleigh model has been shown empirically to be a good fit in some environments, especially where there are many scatterers and no direct path between the transmitter and receiver, and for a fast fading component.
This model establishes that when a channel is Rayleigh fading, the channel response can be expressed equivalently in either polar form as in (2.16) or quadrature form

\[ a_k = a_{kR} + j a_{kI} . \]  

(2.18)

The polar form is formulated as having Rayleigh distributed amplitude and uniformly distributed phase, while in the quadrature form, the real and imaginary parts are identically independently Gaussian distributed with zero mean. Thus \( a_{kR} \) and \( a_{kI} \) both have Gaussian probability density

\[ p_A(a) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{a^2}{2\sigma^2} \right) . \]

In the sequel, we will make use of the quadrature model extensively.

**Rician distribution**

When there exists a dominant path together with low level scattered multipath components in the environment, the channel response can no longer be modeled as having zero mean. In this case, the envelope of the channel response is modeled according to a Rician distribution. The dominant path can be a line-of-sight or a path that experiences much less attenuation than others. The theoretical explanation for this model is that, when a such path exists, the received signal can be considered as a sum of two vectors

\[ re^{j\theta} = u e^{j\beta} + v e^{j\gamma} \]

where \( u e^{j\beta} \) is the random scattering component with \( u \) being Rayleigh distributed and \( \beta \) being uniformly distributed, and \( v e^{j\gamma} \) is the strong path component. Fixing the amplitude and phase of the strong path component, Rice (see [3]) has shown that the distribution of \( r \) and \( \theta \) becomes

\[ p_{R,\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} \exp \left( -\frac{r^2 + v^2 - 2rv \cos(\theta - \gamma)}{2\sigma^2} \right) \cdot u(r) \]

where \(-\pi \leq (\theta - \gamma) \leq \pi\). Furthermore, the amplitude and phase of the strong path is usually changing, and \( \gamma \) is a random variable uniformly distributed over \([0, 2\pi]\). Randomizing \( \gamma \) causes \( r \) and \( \theta \) to become independent, with \( \theta \) having a uniform distribution and \( r \) a Rician distribution given by

\[ p_R(r) = \frac{r}{\sigma^2} \exp \left( -\frac{r^2 + v^2}{2\sigma^2} \right) I_0 \left( \frac{rv}{\sigma^2} \right) \cdot u(r) \]  

(2.19)
where $I_0(x)$ is the modified Bessel function of the first kind and order 0 which has the series expansion

$$I_0(x) = \frac{1}{\pi} \int_0^\pi e^{x \cos \phi} d\phi = \sum_{k=0}^{\infty} \frac{x^{2k}}{4^k k! \Gamma(k + 1)}$$

where $\Gamma(k)$ is the Gamma function defined as $\Gamma(x) = \int_0^\infty \xi^{x-1} e^{-\xi} d\xi$ with $\Re(x) > 0$. The variable $v^2$ represents the power of the strong path.

Obviously when $v$ goes to zero in the above PDF, i.e. the strong path diminishes, the amplitude distribution becomes Rayleigh. Thus Rayleigh fading is a special case of Rician fading. On the other hand, if the power in the strong path is considerably higher than the Rayleigh component, then $r$ and $\theta$ are both approximately Gaussian, $r$ having mean that equals $v$ and $\theta$ having zero mean, and the Rician distribution is approximated around its mode by a Gaussian distribution.

Various analyses of collected data from different environments have also shown that a Rician distribution is a good fit in many cases, especially when there exists a line-of-sight (LOS) path between the transmitter and receiver. Some empirical results have indicated that in extensive temporal fading, even in the absence of a LOS path, the Rician distribution shows much better fit to the data than the Rayleigh distribution.

Again, the Rician model can be expressed in quadrature form (2.18) where the real and imaginary parts of the channel response are independent Gaussian random variables with non-zero means and the same variance. In other word, the $a_{kR}$ and $a_{kI}$ are both Gaussian random variables having independent distributions

$$p_A(a) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(- \frac{(a - \mu_i)^2}{2\sigma^2}\right)$$  \hspace{1cm} (2.20)$$

where $\mu_i$, $i = 1, 2$ are the mean values of $a_{kR}$ and $a_{kI}$ respectively, and $\sigma^2$ is their variance. Then the amplitude $r = \sqrt{a_{kR}^2 + a_{kI}^2}$ has Rician distribution with parameter $v^2 = \mu_1^2 + \mu_2^2$. This model is briefly mentioned in [8]. The phase in this model will have a distribution deviating from uniform, which is expected when the channel response is sampled at symbol rate, as explained below. This model again contains the quadrature Rayleigh model as a special case. Since we will make use of this model extensively, we define a special complex Gaussian random variable as follows.

**Definition 2.1.** A special complex Gaussian random variable is a complex-valued random variable that has independent real and imaginary parts which are Gaussian distributed with the same variance.
With this definition, the channel gain $a_k$ in both Rayleigh and Rician fading are special complex Gaussian random variables. Let $\mu_k$ and $\sigma_k^2$ be the mean and variance of $a_k$, then the real and imaginary parts of $a_k$ are independent Gaussian random variables with mean $\mu_1$ and $\mu_2$ respectively, such that $\mu_1^2 + \mu_2^2 = \mu_k^2$, and variance equal to $\sigma_k^2/2$. In case of Rayleigh fading, all the means are zero ($\mu_1 = \mu_2 = \mu_k = 0$). Some properties of these special complex Gaussian random variables are given in Lemma 1 in the Appendix.

Other distributions

The above two distributions, Rayleigh and Rice, are the most common distributions used in the literature to model the amplitude of a multipath fading channel. There are other well known distributions used in modeling the envelope of the channel response, including the Nakagami-$m$, log normal, Weibull and Suzuki distributions [3].

The Nakagami-$m$ distribution is a more realistic model proposed by Nakagami. Where the Rayleigh model assumes the lengths and therefore attenuation factors of different paths are the same and their phases are random, the Nakagami model permits these path lengths to also be random. The distribution of the channel amplitude is given by

$$p_R(r) = \frac{2m^m r^{2m-1}}{\Gamma(m) \Omega^m} \exp \left( -\frac{mr^2}{\Omega} \right) \cdot u(r)$$

where $\Omega = E[r^2]$ and $m = (E[r^2])^2/\text{var}[r^2]$, with the constraint $m \geq 1/2$. The Nakagami distribution contains many others as special cases. It reduces to Rayleigh fading for $m = 1$ and to a one-sided Gaussian distribution for $m = 1/2$. It also approximates, with high accuracy, the Rician distribution, and approaches a lognormal distribution under certain conditions. Nakagami distributions have been shown to be the best fit for data signals received in urban radio multipath channels. However, application of this model in the literature has generally been neglected.

The lognormal distribution often has been used to explain large scale variation of the signal amplitudes in a multipath fading environment. The channel amplitude $r$ has the distribution

$$p_R(r) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left( -\frac{(\ln r - \mu)^2}{2\sigma^2} \right) \cdot u(r)$$

so that $\ln r$ has a normal (Gaussian) distribution. The theoretical explanation for this model is that, due to multiple reflections in the multipath environment, the fading phenomenon can be characterized as a multiplicative process of the path amplitudes,
which results in lognormal distribution in the same manner as addition gives rise to normal distribution by the Central Limit Theorem. There is significant empirical justification for this model in urban and ionospheric propagation. However, the distinction between Rayleigh and Rician distributions and lognormal distribution is that the previous two models assume statistics of the channel do not change over the area under consideration, whereas when the channel exhibits random effects on the parameters of the distributions, a lognormal distribution is more suitable to model the channel amplitude. Thus, Rayleigh and Rician fading are usually used to model channels which have spatial homogeneity such as local areas, and the lognormal distribution is used in global areas.

Other distributions used for channel amplitude include the Suzuki and Weibull distributions. The Suzuki distribution is a mixture of Rayleigh and lognormal distributions. It has an elegant theoretical explanation whereby the main wave, which has a lognormal distribution, is broken into many paths in the local area of the receiver due to scattering by local objects, and thus creates many subpaths with equal amplitudes and random phases. Therefore the signal envelope has a Rayleigh distribution with the deviation \( \sigma \) being lognormal distributed, which gives rise to the Suzuki distribution. Suzuki distributions have been shown to be a better fit to large area data than the lognormal distribution, although it is not commonly used due to its complicated PDF which is in integral form. The other distribution is Weibull, which also provides a good fit for some mobile radio channels, although there is no theoretical explanation for encountering this distribution.

**Distribution of path phases**

As noted before, the path phases are very sensitive to changes in the environment and are usually modeled as having a uniform distribution over \([0, 2\pi)\). Even though wireless system performance can be very sensitive to the statistical properties of the path phase sequence \(\{\theta_k\}\), where the subscript \(k\) denotes path \(k\), no empirical model for the phase sequence has been reported in the literature [3]. This may be due to the difficulties in measuring the phases of individual multipath components. Considering the geometry of a path, since the phase changes by \(2\pi\) as the path length changes by only a wavelength (e.g. 30cm at 1GHz), for an assembly of points, it is reasonable to model the phase by a uniform distribution. However, at a small sampling distance, the sampled phase sequence of a path can be strongly correlated, especially when the channel response is
sampled at the symbol rate, and large deviations from the uniform distribution can be expected. It has been found that for a fixed delay path the phase values of different samples are correlated, whereas the adjacent detectable multipath components of the same sampled profile have independent phases (uncorrelated scattering phenomenon).

It is therefore accurate to say that the absolute phase value of a multipath component at a fixed point in the space is not important, but more emphasis should be placed on the changes in signal phase as the channel changes. Two models have been used to describe the phase increment of a fixed delay path: a random phase increment model using a Gaussian distribution and a deterministic phase increment model, both with initial phase generated according to an uniform distribution on \([0, 2\pi]\). There is no theoretical bases for choosing a Gaussian distribution for the random phase increment model, and the deterministic model is based on the over-simplistic assumption that the arrival angle of each multipath component caused by a single scatterer remains the same for small spatial separation. All the models have been applied in simulation, but not derived from empirical data. Therefore the accuracy of these models remains a question.

**Distribution of path delay sequence**

The distribution of the path arrival time sequence \(\{\tau_k\}\) has received insufficient attention in the literature. These arrival times form a random point process on the positive time axis, except when there is a LOS, then its arrival time \(\tau_0\) should be excluded since it is not random. Therefore it is more appropriate to look at the distribution of \(\{\tau_k - \tau_0\}\). Several candidate point process models exist. First is the well known Poisson distributed arrival times with exponentially distributed inter-arrival times, which assumes that the obstacles which cause multipath fading are located with complete randomness in space. Although this models provides quite a good fit for insensitive receivers (i.e. for high threshold values), it has been found inadequate when the threshold is lowered. This is probably due to the fact that scatterers are not located with total randomness as assumed. Modified Poisson models are proposed, including the \(\Delta - K\) model, Weibull (i.e. non-exponential) inter-arrival model, and double Poisson model [3]. Again, these models fit better in some cases but are unsuccessful in others. Other models include Gilbert's burst noise model and the pseudo-Markov model which are concerned with correlated events.
Forward and reverse channels

Having established the statistical models for path amplitude, phase and delay time, it should be pointed out that while these statistical properties hold for both forward and reverse channels, the actual values of these parameters are most often uncorrelated between those two links. This issue is called the principle of reciprocity. This principle implies that the channel is identical on the forward and reverse links as long as the channel is measured at the same frequency and at the same time instant. However, this is only true for time-division duplexing (TDD) systems when the time separating a receive from a transmit frame is very small compared to the channel coherence time. In frequency-division duplexing (FDD), which most wireless systems are, the separation between the forward and the reverse links frequencies is about 5% of the mean frequency, and therefore the principle of reciprocity cannot be applied (since forward and reverse link parameters are not measured at the same frequency). In fact, even though the paths used by both links in a multipath channel can be assumed to be identical, that is the number of paths, the path delays and the path angles are the same for both links, the path amplitudes and phases are not; moreover, they will be largely uncorrelated. Also within any one link, different paths fade independently. Given that the fading appears as a multiplicative gain, the channel between the forward and reverse links therefore appears to be uncorrelated [2].

There is another aspect of multipath fading channels that has not been addressed here: the angle spread or space selective fading. This is the effect of the spread of angles of departure or arrival at the antenna due to its spatial location. The distribution of this parameter is dependent on the environment and antenna height, and in some cases, this angle spread is statistically related to the path delay. The reason that the angle spread has not been analyzed is due to its lack of application in antenna diversity, which uses stochastic models of the channel with the assumption of a large number of multipath reflections, as opposed to a directive antenna array or beamforming, which is based on the assumption that there exists some dominant direction of arrival for the signal. In a directive array, the antennas are assumed to know the arrival angle almost perfectly, and the array pattern is adjusted accordingly for the transmitter to focus its energy in the direction of the receiver, or for the receiver to maximize antenna gain in the direction of transmitter. In diversity, on the other hand, the array does not have full knowledge of the channel: the phases are generally unknown and there is not a meaningful notion of direction of arrival (the angle of arrival is assumed to be
uniformly distributed over $[0, 2\pi)$ due to the signal being received after bouncing from surrounding scatterers in all directions), thus the array pattern is adjusted accordingly to the statistical measurements of the channel. Although the distinction between diversity and beamforming can be blurred in many applications, we note that the work and results presented in this thesis are related to diversity.

2.2.3 Discrete channel models

Generally, we see discrete channel models in the literature since such models are usually easier to analyze than continuous models, and they greatly simplify the simulation process. They also lead naturally to transmitter and receiver algorithms having efficient implementations on digital signal processor-based architectures.

Discrete time channel models are based on the continuous time models that we have developed above. It is also understood that the lowpass equivalent channels are being modeled instead of bandpass channels and so the subscript $l$ is dropped for convenience.

Suppose the signal bandwidth is $W$ and the received waveform is sampled at the symbol rate $T = 1/W$, followed by discrete time processing. Let $s[k]$ be the representation of the transmitted signal $s(t)$ in terms of its Nyquist T-spaced samples, i.e. $s[k] = s(kT)$. The received signal (2.4) can be transformed to

$$y(t) = \sum_{k} c(t; k)[s[k] + n(t)]$$

where

$$c(t; k) = \int c(t; t - \tau) \frac{\sin(\pi(\tau - kT)/T)}{\pi(\tau - kT)/T} \, d\tau.$$  

Let $y[n]$ be the sequence arising from filtering the received signal $y(t)$ by an ideal lowpass filter with cutoff frequency $\pm W/2$ and then sampling at rate $T$. Since the low pass filter is a linear operation on $y(t)$, we have

$$y[n] = c[n; k] * s[n] + w[n] = \sum_{k} c[n; k]s[n - k] + w[n]$$

where $y[n] = y(nT)$, $w[n] = n(nT)$ is the noise sample, and $c[n; k] = c(nT; k)$ is the kernel of the discrete channel response. The noise is often modeled as a zero mean, complex-valued, circularly symmetric, stationary Gaussian random sequence.

\footnote{We use the notation $x[n]$ to denote discrete values and $x(\cdot)$ to denote a continuous signal. The double notation $c(t; k)$ indicates it is continuous in time $t$ and discrete in delay $k$.}
with variance \( \sigma_N^2 = E[|w[n]|^2] = N_0W = N_0/T \) (where \( N_0 \) is the noise spectrum density). The kernel \( c[n;k] \) represents the response of the channel at time \( n \) to a unit sample input at time \( n-k \). The discrete channel has a time-varying frequency response \( C(\omega;n) \) defined as

\[
C(\omega;n) = \sum_k c[n;k] e^{-j\omega k T}
\]

(2.23)

which is related to the spectrum \( C(\omega;t) \) of the continuous channel (2.6) by

\[
C(\omega;n) = \sum_k C(\omega - \frac{2\pi k}{T}, nT).
\]

Assuming linear modulation of the transmit signal \( s(t) \), the general equation that represents linear modulation is

\[
s(t) = \sum_n c_n \cdot g(t - nT)
\]

where \( c_n \) are the coded information symbols and \( g(t) \) is the modulating pulse, which is usually assumed to have a frequency spectrum as the square root of a Nyquist pulse. The representation of the modulated signal \( s(t) \) by its Nyquist samples suggests the discrete representation for the linear modulation scheme

\[
s[n] = \sum_k c_n \cdot g[n - k] = c_n * g[n]
\]

(2.24)

where \( g[n] = g(nT) \). Thus the discrete received signal can be written as

\[
y[n] = c_n * g[n] * c[n;k] + w[n].
\]

(2.25)

Consider now the kernel \( c[n;k] \) of the discrete channel. This kernel can be reflected in the quasi-stationary model (2.15). For each value of the delay \( k \), \( c[n;k] \) is a wide-sense stationary complex-valued random sequence in time \( n \), which corresponds to the time realization of the path gain \( a_k \) in (2.15) (assuming the \( a_k \) are resolvable). Using either a Rayleigh or Rician fading model, this sequence is a special complex Gaussian random
sequence. Furthermore, due to uncorrelated scattering, the sequences corresponding to distinct values of delay $k$ (i.e. distinct resolvable paths) are statistically independent.

Except when there is feedback from receiver to transmitter, the transmitter generally does not know the value of the channel kernel $c[n; k]$. On the other hand, the receiver is usually able to infer information about the fading channel kernel from the received waveform, through either a training sequence sent by the transmitter or by applying a “blind” channel estimation algorithm. Therefore it is generally assumed that the receiver has perfect knowledge of the channel coefficients, an assumption which allows us to develop useful bounds on the performance attainable in practice. In later sections, we will also study the case where only the mean and variance of the kernel $c[n; k]$ are available at the receiver for each path delay $k$.

When $C(\omega; n)$ does not vary with $n$, the channel is time nonselective (or slow fading). The input output relation involves simple convolution, i.e.

$$ y[n] = \sum_k c[k] s[n-k] + w[n] = c[n] * s[n] + w[n] \quad (2.26) $$

where the unit sample response $c[k]$ is given by $c[k] \equiv c[0; k]$ and has the associated frequency response $C(w) \equiv C(w; 0)$. In such a channel, temporal diversity cannot be exploited.

When $C(\omega; n)$ does not vary with $\omega$, i.e. $C(\omega; n) \equiv C(0; n)$, the channel is frequency nonselective and we have $c[n; k] = c[n] \delta[k]$, where $\delta[k]$ is the Dirac delta function, i.e.

$$ \delta[k] = \begin{cases} 
1 & k = 0 \\
0 & otherwise
\end{cases} $$

In this case, the input output relation becomes a simple multiplicative relationship

$$ y[n] = c[n] \cdot s[n] + w[n] \quad . \quad (2.27) $$

On such channels, spectral diversity cannot be exploited.

Finally when the channel is both frequency nonselective (flat fading) and time nonselective (slow fading), the input output equation becomes

$$ y[n] = a \cdot s[n] + w[n] \quad (2.28) $$

where $a$ is a complex-valued random variable. On such channels, neither temporal nor spectral diversity can be exploited.
2.2.4 Channel model with multiple antennas

Spatial diversity implies using multiple antennas at either the transmitter, receiver or both, and can often be applied in all above cases. It exploits the fact that different gains $c[n; k]$ are encountered by different elements of an antenna array. Using the discrete model from above, we will build a model for multiple antennas, frequency non-selective and relatively slow fading channels.

Focusing on the single user case, assume that the number of antennas at the transmitter is $M$ and at the receiver is $N$. The transmitted signal has total power constrained to $P$ regardless of the number of transmitting antennas. Antennas are physically spaced far enough apart to create independent fading between each transmit-receive antenna pair. This antenna separation needs to be only half a wavelength for independent fading [4], thus the assumption is quite acceptable for high frequency transmission.

As previously analyzed, frequency non-selective and slow fading channels are multiplicative. The input-output relation can be written as

$$ y = Ax + n $$

(2.29)

where $x \in C^M$ is transmit vector, $y \in C^N$ is receive vector, $A$ is a $N \times M$ random complex matrix of channel coefficients and $n \in C^N$ is the noise vector. We employ a Rician fading model with Rayleigh fading as a special case, where the entries $a_{i,j}$ of the channel matrix $A$ are special complex Gaussian random variables. Since the transmit antennas are placed relatively closely to each other, it is reasonable to assume that the power gains and variances of the coefficients between different transmit antennas and the receive antenna are the same, i.e., all $a_{i,j}$ have mean $\mu$ and variance $\sigma_a^2$ (when $\mu = 0$, the channel is Rayleigh fading). Due to independent fading between each pair of transmitting-receiving antennas, these channel coefficients are independent. The noise components of $n$ are zero-mean special complex Gaussian random variables with variance $\sigma_N^2 = N_0W = N_0/T$. The noises at different receiving antennas are also independent.

Furthermore, the transmitter is subject to a power constraint by \footnote{The symbol $^\dagger$ denotes the complex conjugate transpose matrix.}

$$ E[x^\dagger x] \leq P. $$

(2.30)

Let $\rho_0$ be the average transmit SNR, i.e. $\rho_0 = P/(N_0W) = P/\sigma_N^2$. Since the received
signal at antenna $j$ can be written as

$$y_j = \sum_{i=1}^{M} a_{ji}x_i + n_j, \quad j = 1 \ldots N,$$

the average signal power at receive antenna $j$ is

$$P_j = E\left[ \left( \sum_i a_{ji}x_i \right)^2 \right] = \sum_i E[a_{ji}x_i^2] = \sigma_n^2 \cdot \sum_i E[x_i^2] = P\sigma_n^2.$$

Thus the average receive SNR at any receive antenna will be the same, independent of $M$, and equal to

$$\rho = \frac{P \cdot \sigma_n^2}{\sigma_N^2} = \frac{P \cdot \sigma_n^2}{N_0W} = \rho_0 \cdot \sigma_n^2 \quad (2.31)$$

We will use this average SNR extensively in the following section and subsequent chapters.

Spatial diversity has recently been recognized as a means to substantially improve the capacity of wireless channels ([10], [11]). Various techniques have been studied in the past ([20], [21], [25], [23], [22], [15], [24], [30]), and this is now becoming an active research topic ([26], [17], [18], [12], [27], [28]). In the following section, we will present the information theory findings on the capacity of a wireless system using multiple antennas. In Chapter Two, we will discuss in detail some spatial diversity techniques found in the literature.

### 2.3 Information theoretic capacity of wireless channel with multiple antennas

The information capacity of an additive white Gaussian noise (AWGN) communication system has been established by the well known Shannon formula, which gives the highest rate at which information can theoretically be transmitted without error. This capacity acts as the upper bound for the on-going search for practical coding and signal processing techniques to achieve a rate close to the bound and with a negligible error rate. In the case of a fading channel, due to the fluctuation in the channel amplitude, the capacity is understandably less than for a non-fading additive Gaussian channel.

However, recent research has found that using multiple antennas at both ends of a wireless system, the ultimate capacity of a band-limited wireless communication system is increased substantially. As Foschini and Gans state in [10], compared to the baseline of using a single antenna at each end, which by Shannon’s classical formula, scales
as one more bit/cycle for every 3dB increase in signal-to-noise ratio (SNR), using \( n \) multiple antennas at both ends, the scaling is almost like \( n \) more bits/cycle for each 3dB increase in SNR. This result is important since it brings closer to reality the possibility of having high speed data over wireless communication channels. With the emerging trend of wireless communication together with the existing wired networks, the idea of having a mobile computing device, integrating the functions of a laptop computer, cellular phone, personal digital assistance, digital camera, video game, calculator and many other remote devices; which is capable of browsing the Internet, transmitting and receiving high speed data such as multimedia, is becoming nearer to practice.

The specific implication of the channel coherence time \((\Delta t)_c\) on a particular system affects the notion of the system’s capacity greatly. Of particular relevance are the signal bandwidth \( W \) and the transmission duration of the whole message (codeword) \( T_c \). It should be emphasized here that we distinguish between channel symbol duration \( T \) (of order \( W^{-1} \)) which characterizes fast fading or slow fading, and the transmission duration of the whole message \( T_c \). This latter parameter characterizes the channel as ergodic \((T_c \gg (\Delta t)_c)\) or non-ergodic \((T_c \ll (\Delta t)_c)\) according to the variability of the fading process in terms of the whole codeword transmission duration \([13]\). Even if \( T \ll (\Delta t)_c \) which implies slow fading, the total transmission time \( T_c \) may be so large that \( WT_c \gg 1 \) and the channel can be viewed as ergodic. This in turn has a major impact on the notion of channel capacity. In an ergodic channel, \( T_c \gg (\Delta t)_c \) means that the total transmission duration is long enough to reveal the long-term ergodic properties of the fading channel, giving rise to ergodic or average capacity. This is the standard capacity in the Shannon sense which implies that if the transmission rate is lower than capacity, the error probability will decay exponentially with the transmission length for a good (random) code. For the non-ergodic case, there is no capacity in the strict sense that allows arbitrary small error communication with a rate below the capacity. When the channel is non-ergodic, for any given channel realization, there is always a non-zero probability, independent of the code length, that the chosen transmission rate, however small, is still greater than the instantaneous capacity of the channel. In other word, the channel produces a random capacity. This gives rise to the notion of capacity versus outage probability \([14]\), where the latter is defined as the probability that the transmitted rate is greater than the capacity of the channel. The capacity versus outage probability is measured by the probability that the channel can support a given rate, for example, capacity at 99\% is the capacity that the channel can
provide with 99% probability, which amounts for 1% outage.

We will focus on the information theoretic capacity of a Rayleigh fading channel in this section, although some results presented below are general and can be applied to any fading channel distribution. As commonly found in the literature, the receiver is usually assumed to know the channel state information (i.e., the realizations of channel coefficients at each time instant) perfectly, whereas transmitter may or may not have such knowledge. Though, it was shown that in the asymptotic case (number of transmit antennas is very large) where perfect channel state information is available to both transmitter and receiver so that the optimal “water-filling” [5] can be applied to maximize capacity, only for low signal-to-noise ratio (SNR) there is a substantial fourfold increase in capacity, while as SNR approaches infinity, the advantage of having the channel state information at the transmitter disappears ([13] and reference therein). Also it should be noted that having channel state information available at the transmitter may require feedback, which cannot generally be exploited in applications such as broadcasting. All the information theory capacity results presented here assume that channel state information is available to the receiver only.

### 2.3.1 Ergodic channel: Capacity as the highest achievable rate

For the case where the channel is ergodic, it has been established by Telatar [11] that the capacity of a Rayleigh fading channel is given by

\[
C = E \left[ \log_2 \det \left( \mathbf{I}_N + \frac{\rho}{M} \mathbf{A} \mathbf{A}^\dagger \right) \right] \quad \text{bps/Hz}
\]

where $\mathbf{I}_M$ is the unit matrix of size $M$. This capacity is achieved when the transmitted vector $\mathbf{x}$ is a random vector with i.i.d (independently identically distributed) zero-mean, special complex Gaussian entries. The covariance of $\mathbf{x}$ therefore satisfies $E[\mathbf{x}\mathbf{x}^\dagger] = (P/M)\mathbf{I}_M$, which implies equal power transmitted from each transmit antenna.

Let $\hat{\mathbf{A}} = \frac{1}{\sigma_0} \mathbf{A}$ be the normalized channel coefficient matrix, i.e., the entries of $\hat{\mathbf{A}}$ are special complex Gaussian random variables with zero mean and unit variance, then the capacity equation can be rewritten as

\[
C = E \left[ \log_2 \det \left( \mathbf{I}_N + \frac{\rho}{M} \hat{\mathbf{A}} \hat{\mathbf{A}}^\dagger \right) \right] \quad \text{bps/Hz}.
\]

(2.32)

Note that for a fixed number of receive antennas $N$, since the entries of matrix $\hat{\mathbf{A}}$ are i.i.d Gaussian random variables, by the Strong Law of Large Numbers, $\frac{1}{M} \hat{\mathbf{A}} \hat{\mathbf{A}}^\dagger \rightarrow \mathbf{I}_N$ almost surely as $M \rightarrow \infty$. Thus the asymptotic capacity in the limit of the number of
transmitting antennas is

$$C = N \log_2(1 + \rho),$$

which means the capacity increases almost linearly with the number of receive antennas when the number of transmit antennas is very large.

**Statistical analysis of channel capacity in special cases**

Here we will adopt the modified chi-squared notation as used in [10] to represent the amplitude distribution of a complex Gaussian random variable. Since each entry $\tilde{a}_{ij}$ of the $\mathbf{A}$ matrix is an independent special complex Gaussian random variable with zero mean and unit variance, which means the real and imaginary parts of $\tilde{a}_{ij}$ are zero mean independent Gaussian random variables with variance equal to $1/2$, $|\tilde{a}_{ij}|^2$ is a modified central chi-squared random variable with two degrees of freedom $\chi^2_2$ (normally the chi-squared distribution applies for sum of square of independent Gaussian random variable with unit variance). The sum of any $n$ squared different channel coefficient magnitudes is also a chi-squared distributed random variable with degree $2n$ denoted by $\chi^2_{2n}$. The distribution of a modified central chi-squared random variable with $\nu$ degrees of freedom $X = \chi^2_{\nu}$ is given by

$$p_{\chi}(x) = \frac{1}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} \cdot u(x).$$

Consider the following special cases. Using the above notation, the capacity of the baseline single transmit and single receive antenna system ($M = N = 1$) is given by

$$C_{1,1} = E \left[ \log_2 (1 + \rho \chi^2) \right] = \int_0^\infty \log_2 (1 + \rho x) e^{-x} dx. \quad (2.33)$$

Since log is a concave function, applying Jensen’s inequality, we have

$$E \left[ \log_2 (1 + \rho \chi^2) \right] \leq \log_2 \left( 1 + \rho \cdot E \left[ \chi^2 \right] \right) = \log_2 (1 + \rho)$$

where the expression on the right hand side is the capacity of non-fading additive Gaussian noise channel with the same SNR, which we shall call the equivalent non-fading channel. Thus the capacity of a fading channel is always less than the capacity of an equivalent non-fading channel.

When only receive diversity is present ($M = 1$), using the determinant identity $\text{det}(\mathbf{I} + \mathbf{A} \mathbf{B}) = \text{det}(\mathbf{I} + \mathbf{B} \mathbf{A})$ where $\mathbf{I}$ is the identity matrix of appropriate size, the
capacity can be written as

\[
C_{1,N} = E \left[ \log_2 \left( 1 + \rho \chi_{2N}^2 \right) \right] = \frac{1}{\Gamma(N)} \int_0^\infty \log_2 \left( 1 + \rho x \right)x^{N-1}e^{-x}dx
\]  

(2.34)

Whereas when only transmit diversity is present \((N = 1)\), the capacity is

\[
C_{M,1} = E \left[ \log_2 \left( 1 + \frac{\rho}{M} \chi_{2M}^2 \right) \right] = \frac{1}{\Gamma(M)} \int_0^\infty \log_2 \left( 1 + \frac{\rho}{M} x \right)x^{M-1}e^{-x}dx
\]  

(2.35)

In effect, we can see that both transmit diversity only and receive diversity only cases improve the fading channel's capacity over the single antenna case. However, given the same number of antennas in both cases, the capacity provided by receive diversity is obviously greater than that of transmit diversity. The capacity of receive diversity only increases approximately as a logarithmic function of the number of receive antennas \(N\) and, as \(N \to \infty\), the capacity goes asymptotically to \(\log(1 + \rho N)\). On the other hand, transmit diversity affects the capacity such that, when the number of transmit antennas is very large, by the Strong Law of Large Numbers, \(\frac{1}{M} \chi_{2M}^2 \to 1\) almost surely as \(M \to \infty\). Thus, when the number of transmit antennas is large

\[
C_{M,1} = E \left[ \log_2 \left( 1 + \frac{\rho}{M} \chi_{2M}^2 \right) \right] \xrightarrow{a.s.} \log_2(1 + \rho) \text{ as } M \to \infty
\]

i.e., the capacity of the fading channel approaches that of the equivalent non-fading channel. Therefore, the impact of transmit diversity can be taken as combating the fading effect of the channel. Note that this transmit diversity bound, unlike the receive diversity case, does not depend on the number of transmit antennas \(M\), so as \(M\) increases beyond a certain value, the capacity does not increase significantly as a result of increasing \(M\). This is the diminishing return effect of multiple transmit antennas.

**Evaluation of channel capacity**

Using the distribution of eigenvalues of a random matrix, Telatar [11] has computed the capacity of a single user channel with \(M\) transmit antennas and \(N\) receive antennas to be

\[
C_{M,N} = \int_0^\infty \log_2 \left( 1 + \frac{\rho}{M} \chi_{2M}^2 \right) \sum_{k=0}^{m-1} \frac{k!}{(k + n - m)!} \left[ I_{k}^{n-m}(\lambda) \right]^2 \lambda^{n-m}e^{-\lambda}d\lambda
\]  

(2.36)

where \(m = \min\{M, N\}\) and \(n = \max\{M, N\}\) and

\[
I_k^{n-m}(\lambda) = \frac{1}{k!} \int e^{\lambda \gamma - n} d\lambda \left( e^{-\lambda} x^{n-m+k} \right)
\]
is the associated Laguerre polynomial of order $k$.

**Equal number of transmit and receive antennas**

In case that the number of transmit and receive antennas is the same, i.e. $M = N$, application of (2.36) yields

$$C_{N,N} = \int_0^\infty \log_2 \left( 1 + \frac{\rho \lambda}{N} \right) \sum_{k=0}^{N-1} L_k(\lambda)^2 e^{-\lambda} d\lambda$$

(2.37)

where $L_k(\lambda) = L_k^0(\lambda) = \frac{1}{\pi} e^{\lambda} \frac{d^k}{d\lambda^k} e^{-\lambda} x^k$ is the Laguerre polynomial of order $k$. Telatar [11] observed that the capacity in this case can be well approximated by a linear function of the number of antennas $N$

$$C_{N,N} \approx N \int_0^4 \log_2 (1 + \rho \nu) F(\nu) d\nu$$

where the function $F(\nu)$ is given by

$$F(\nu) = \begin{cases} 
\pi^{-1} \sqrt{\nu - 1/4} & 0 < \nu \leq 4 \\
0 & \nu > 4 
\end{cases}$$

Thus when the number of transmit and receive antennas is equal, the channel capacity increases almost linearly with the number of antennas $N$. This is like having $N$ parallel non-interfering deterministic channels transmitting at the same time. The finding is a truly ground-breaking result.

**2.3.2 Non-ergodic channel: Capacity versus outage probability**

In a non-ergodic channel, the channel time coherence is assumed to be very large compared to the duration of transmitting a message (i.e. a burst duration when the communication mode is bursty), therefore the average capacity cannot be realized. Using the quasi-static model where the random channel selected at each burst is essentially not changing, the capacity at each transmission is a random variable and determination of the probability distribution function of capacity will be the main objective.

The transmitted signal vector is assumed to be composed of $M$ independent special complex Gaussian components. Similar to the result in the previous case (2.32), the generalized, but in this case, random capacity can be expressed by

$$C = \log_2 \det \left( I_N + \frac{\rho}{M} \tilde{A} \tilde{A}^\dagger \right) \text{bps/Hz}.$$  

(2.38)

This capacity expression does not require a specific distribution of the channel matrix $A$. However, in this case, it has not been rigorously proven that the maximum capacity
is achieved when the transmit vector \( \mathbf{x} \) consists of independent special complex Gaussian components. Choosing a different joint distribution of the transmit components will increase the capacity for some values of \( \mathbf{A} \) and decrease it for others. Thus the chosen distribution of \( \mathbf{x} \) may not be optimal in some global sense, but this may be the saddle point solution to a max-min problem in which nature chooses a distribution of \( \mathbf{A} \) to maximize outage probability [12].

From here we will use the Rayleigh model for the channel matrix \( \mathbf{A} \). Again capacity of different combinations of transmit and receive antennas are considered. Similar expressions to the ergodic channel capacity (2.33, 2.34, 2.35) are established, but without using expected values. The single transmit and single receive antenna channel capacity is given by \( C_{1,1} = \log_2(1 + \rho \chi_2^2) \). The capacity for receive diversity only is

\[
C_{1,N} = \log_2(1 + \rho \chi_2^N) \tag{2.39}
\]

and the capacity for transmit diversity only is

\[
C_{M,1} = \log_2\left(1 + \frac{\rho}{M} \chi_2^M\right). \tag{2.40}
\]

These capacities are now random variables and cannot be compared directly. However, it can be shown that the fading combating effect of transmit diversity still applies, as \( C_{M,1} \rightarrow \log_2(1 + \rho) \) almost surely as \( M \rightarrow \infty \) by the Law of Large Numbers. Thus even in the case of a non-ergodic channel, having many transmit antennas will stabilize the fading channel and make the random capacity of the channel approach the deterministic value of AWGN channel capacity. The outage probability therefore will be driven to zero with an increasing number of transmit antennas when the transmission rate is less than \( \log_2(1 + \rho) \text{bps/Hz} \).

Upper bound and lower bound on capacity

When full diversity at the transmitter and receiver are present, the upper bound for the generalized capacity (2.38) is given by

\[
C_{M,N} \leq \sum_{i=1}^{M} \log_2\left(1 + \frac{\rho}{M} \chi_2^{N(i)}\right) \tag{2.41}
\]

where the \((i)\) indexes at the chi-squared variable indicate independent chi-squared variables with \(2N\) degree of freedom. This upper bound can be viewed as the capacity of the artificial case when each of the \(M\) signals transmitted by the transmit antennas is received perfectly by a set of \(N\) receive antennas without any interference from others,
i.e. the channel is viewed as composed of \( M \) parallel uncoupled channels, with each channel having \( N \) receive antennas.

Also using results on random matrices, Foschini and Gans in [10] have derived the lower bound for capacity in case that the number of transmitting antennas and receive antennas are equal, \( M = N \). The lower bound, with probability one, is given by

\[
C_{N,N} > \mathcal{L}(N) = \sum_{k=1}^{N} \log_2 \left( 1 + \frac{\rho}{N} \chi^2_{2k} \right)
\]

The bound is quite tight for a large number of antennas and with high receive SNR \( \rho \).

It can be shown that

\[
\frac{C_{N,N}}{N} > \frac{\mathcal{L}(N)}{N} \rightarrow (1 + \frac{1}{\rho}) \log_2(1 + \rho) - \log_2 e - \epsilon_N + o(N^{-1})
\]

as \( N \to \infty \), where the random variable \( \epsilon_N \) has a Gaussian distribution with mean

\[
\epsilon_N = \frac{1}{N} \log_2 (1 + \rho)^{-1/2}
\]

and variance

\[
\text{var}(\epsilon_N) = \left( \frac{1}{N \cdot \ln 2} \right)^2 \left[ \ln(1 + \rho) - \frac{\rho}{1 + \rho} \right]
\]

The variance of \( \epsilon_N \) vanishes when \( N \) is large. Thus with large \( N \), the dominant term of \( C_{N,N} \) scales at least linearly with increasing \( N \). So even in the non-ergodic case, the capacity of a channel with equal numbers of antennas at both the transmitter and receiver also scales approximately linearly with the number of antennas.

This capacity with large \( N \) may seem to be unreasonable in term of \( \text{bps}/\text{Hz} \), but it was pointed out in [10] that the capacity in term of \( \text{bps}/\text{Hz} \) per dimension asymptotically approaches some constant for each fixed value of SNR. For example, at 0 dB SNR, the capacity is about 0.8 \( \text{bps}/\text{Hz}/\text{dimension} \) for \( N = 32 \), which is close to the capacity of 1 \( \text{bps}/\text{Hz} \) of a single antenna non-fading channel at 0 dB SNR. It is conjectured that the capacity per dimension asymptotically approaches the capacity of the equivalent single antenna non-fading channel.

### 2.3.3 Capacity for different diversity techniques

Although the ultimate capacity with multiple antennas at both ends is very large, there is not yet a signal processing technique that can achieve this capacity. Various diversity techniques have been studied in the literature, each offering some fraction of the available capacity. We will give examples of the capacities of some common diversity techniques below.
Receive diversity

The use of multiple antennas at the receiver has been studied extensively. Techniques such as selection diversity and maximal ratio combining (optimum combining) have been derived ([4], [10], [13] and references therein). In this case, the transmission is from a single transmit antenna to $N$ receive antennas (Figure 2.6). The received signal at each antenna is

$$y_i[n] = a_i \cdot x[n] + w_i[n] \quad i = 1 \ldots N$$

![Diagram](image)

Figure 2.6: Multiple antenna receive diversity.

Selection diversity is the technique where only the strongest signal among the receive antennas is selected to be processed. Its capacity is given by

$$C = \max_i \log_2 \left( 1 + \rho |a_i|^2 \right) = \log_2 \left( 1 + \rho \cdot \max_i |a_i|^2 \right)$$

where the maximization is over \{i | 1 \leq i \leq N\} and $a_i$ is the channel coefficient between the transmit antenna and the receive antenna $i$.

For maximal ratio combining, which is optimal for receiver diversity, a sufficient statistic $\{y[n]\}$ for the maximum likelihood decision on the transmit data is obtained as

$$y[n] = \sum_{i=1}^{N} y_i[n] \cdot a_i^* = x[n] \sum_{i=1}^{N} |a_i|^2 + \sum_{i=1}^{N} w_i[n] \cdot a_i^* = r'[n] + w'[n]$$

where $w'[n]$ is the equivalent noise, which is a linear combination of the Gaussian noises at each branch and therefore is also Gaussian with zero mean. It can be shown that the resulting combined SNR $\gamma$ is maximized and equal to the sum of the SNR at each
branch

$$\gamma = \frac{E |v'[n]|^2}{E |w'[n]|^2} = \rho_0 \sum_{i=1}^{N} |a_i|^2.$$  \hspace{1cm} (2.43)

The combined channel therefore is equivalent to a single Gaussian fading channel with fade coefficient equal to \((\sum_{i=1}^{N} |a_i|^2)^{1/2}\). The capacity in this case is given by

$$C = \log_2 \left( 1 + \rho_0 \sum_{i=1}^{N} |a_i|^2 \right) = \log_2 \left( 1 + \rho \cdot \chi^2_N \right)$$

This capacity is the same as the maximum achievable capacity using receive diversity only (2.39). We see that the capacity of optimum combining is larger than that of selection diversity, although selection diversity has the advantage of simpler implementation.

With this capacity for optimum combining receive diversity, the formula for the lower bound of capacity for full transmit and receive diversity (2.42) suggests that in some sense, one might be able to embed \(N\) optimum combining systems with degree \(k = 1, 2, \ldots, N\) in an \((N, N)\) transmit-receive antenna system, where each transmit antenna transmits with power \(P/N\). This leads to a signal processing method called the layered space-time architecture \([26]\) described briefly at the end of this section.

**Transmit diversity**

Recently transmit diversity has received considerable attentions ([20], [21], [25], [23], [22], [17], [18], [12]). Here we will examine the capacity of some transmit diversity techniques, which can be categorized into six schemes: unconstrained signaling, time division, frequency division, time-shifting, frequency-shifting and random weighting techniques \([12]\). Unconstrained signaling means the system accepts a vector input from multiple transmit antennas, whereas the other five schemes use linear processing to convert the vector-input channel to a scalar-input channel.

It is found that the performance of all the transmit diversity schemes mentioned depends only on the distribution of the magnitude of the channel coefficients \(|a_i|\), not their relative phases, and for some schemes, the performance depends only on the distribution of total antenna gain \(||a|| = \sum_{i=1}^{M} |a_i|^2\), where \(a\) is the vector of channel coefficients \([a_1, a_2, \ldots, a_M]^T\). Again we assume that for scalar input, the input signal \(\{x[n]\}\) is a sequence of independent zero-mean special complex Gaussian random variables with variance \(E|x[n]|^2 = P\). For a vector input, the input sequences \(\{x_i[n]\}\) of different transmit antennas are also independent and have the same variance \(E|x_i[n]|^2 = P/M\).
Unconstrained signaling

\[
y[n] = \sum_{i=1}^{M} a_i \cdot x_i[n] + w[n]
\]

With the transmit signals \( x_i[n] \) at each time \( n \) being independent zero-mean special complex Gaussian random variables with the same variance \( E[x_i[n]^2] = P/M \), the output \( y[n] \) is a zero-mean Gaussian random variable with variance \( (P/M) \sum_{i=1}^{M} |a_i|^2 + \sigma_N^2 \), and the capacity is

\[
C_{opt} = \log_2 \left( 1 + \frac{P}{M} \sum_{i=1}^{M} |a_i|^2 \right) = \log_2 \left( 1 + \frac{P}{M} \cdot \chi_{2M}^2 \right). \tag{2.44}
\]

This is the same capacity as the maximum achievable transmit diversity capacity in (2.40). In this case the capacity only depends on the total antenna gain \( ||a||^2 \).

Time division and frequency division:

Figure 2.7: Multiple antenna transmit diversity

Figure 2.8: Time division (a) and frequency division (b) transmit diversities
Time division and frequency division (Figure 2.8) are found to have the same mutual information, since the channel in these cases can be viewed as a set of independent parallel channels. The generated signals are orthogonal and will remain orthogonal at the receiver due to the lack of intersymbol-interference because of the quasi-stationary and frequency nonselective channel model. Thus we will present the result for the time division case only. Time division is implemented by spatially cycling through each transmit antenna one at a time. That is we cycle through all $M$ transmit antennas periodically with period $M$. The output of this channel at time $n$ is

$$y[n] = a_k \cdot x[n] + w[n], \quad k \equiv n (\text{mod} M + 1).$$

Beside the ease of implementation, this method reduces the variance of the received signal and of the random capacity as the number of transmit antennas increases, even more rapidly than when there is receive diversity only. The intuitive reason is that the cycling of the transmit antennas helps to avoid being stuck with a bad channel for a long time. The capacity for spatial cycling is

$$C_{td} = C_{fd} = \frac{1}{M} \sum_{i=1}^{M} \log_2 \left( 1 + \rho_0 |a_i|^2 \right) = \frac{1}{M} \sum_{i=1}^{M} \log_2 \left( 1 + \rho \cdot \chi^2_{2(i)} \right). \quad (2.45)$$

Compared with the unconstrained signaling case, applying Jensen’s inequality $\text{Average}(\log t) \leq \log(\text{Average} t)$ gives $C_{td} = C_{fd} \leq C_{opt}$ with equality if and only if all the antenna gains $|a_i|$ are the same.

**Time shifting and frequency shifting**

![Diagram](image)

Figure 2.9: Time shifting (a) and frequency shifting (b) transmit diversities.

Similarly, time shifting and frequency shifting (Figure 2.9) are duals, and we will describe only the time shifting method here. In time shifting diversity, the transmitter sends delayed versions of the same input signal via all the transmit antennas, i.e. the
ith antenna transmits the input signal delayed by \( i - 1 \) time steps (the step is usually at least a symbol interval for uncorrelated receive signals). The output of the time shifting channel is

\[
y[n] = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i \cdot x[n - i + 1] + w[n].
\]

Taking Fourier transforms of both sides yields

\[
Y(\omega) = \frac{1}{M} \sum_{i=1}^{M} a_i \cdot X(\omega) e^{-j\omega(i-1)} + N(\omega)
\]

where \( N(\omega) \) is the spectrum of the discrete noise \( w[n] \). Therefore the frequency response of the time shifting channel can be written as

\[
A(\omega) = \sum_{i=1}^{M} a_i e^{-j\omega(i-1)}.
\] (2.46)

Thus the channel capacity is given by

\[
C_{ts} = C_{fs} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log_2 \left( 1 + \frac{\rho_0}{M} \sum_{i=1}^{M} e^{-j\omega(i-1)} a_i \right)^2 d\omega.
\] (2.47)

Again applying Jensen’s inequality, we have \( C_{ts} = C_{fs} \leq C_{opt} \) with equality if and only if a single component of \( a \) is non-zero.

**Randomly time weighting**

![Randomly time weighting](image)

Figure 2.10: Randomly time weighting transmit diversity.

Finally for the randomly time weighted diversity scheme (Figure 2.10), again a scalar input input \( x \) is mapped linearly into a vector output \( [x_1 \ x_2 \ldots \ x_M]^T \), but the mapping here is random. The motivation for this random time weighting is that the performance of the systems depends only on \( ||a|| \) and not the individual coefficient magnitudes. The vector input is generated by multiplying the scalar input by a complex
unit-magnitude random vector \( \mathbf{b}_n = [b_1[n] \ldots b_M[n]]^T \) chosen randomly and uniformly over the surface of the unit sphere, i.e.

\[
x_n = x[n] \cdot \mathbf{b}_n .
\]

The output of the channel is then

\[
y[n] = \mathbf{a}^T \mathbf{b}_n \cdot x[n] + w[n] .
\]

The weighting \( \mathbf{b}_n \) is assumed to be known at the receiver so that the receiver can extract the original information sent. This can be achieved by selecting \( \mathbf{b}_n \) pseudo-randomly according to some prearranged scheme. Using a circular unitary ensemble matrix decomposition to write \( \mathbf{b}_n = \mathbf{U}_n [1 \ 0 ; \ldots \ 0]^T \) where \( \mathbf{U}_n \) is a random unitary matrix, and similarly with \( \mathbf{a} \), the channel output may be written as

\[
y[n] = ||\mathbf{a}|| \cdot \mu_n x[n] + w[n]
\]

where \( \mu_n \) denotes the single upper left entry of an \( M \times M \) matrix drawn from the circular unitary ensemble. The capacity of the channel therefore is

\[
C_{ran} = \int_0^1 \log_2 (1 + \mu ||\mathbf{a}||^2 \eta) \cdot f_{||\mu||^2}(\eta) \, d\eta
\]

(2.48)

where \( f_{||\mu||^2}(\cdot) \) is the probability density function of the squared magnitude of any entry \( \mu \) of a random matrix drawn from the circular unitary ensemble. It can be shown [12] that \( f_{||\mu||^2}(\eta) \) is given by

\[
f_{||\mu||^2}(\eta) = \begin{cases} 
(M - 1)(1 - \eta)^{M-2} & 0 < \eta \leq 1 \\
0 & \text{otherwise}
\end{cases} \quad \text{with} \quad M \geq 2
\]

Applying Jensen’s inequality, again we have \( C_{ran} \leq C_{opt} \) where the equality happens only when \( ||\mathbf{a}||^2 \) = 0.

**Significance of transmit diversity**

For all of the above transmit diversity schemes, unconstrained signaling offers the largest capacity, which is the maximum achievable capacity for transmit diversity. However, in unconstrained signaling the input signal is a vector, hence this can complicate the detection process at the receiver, whereas the input signal in other cases is scalar and therefore can be more easily reconstructed at the receiver. This is the reason we study different transmit diversity techniques. In fact, the scalar input schemes effectively transform antenna diversity into other forms of diversity: frequency diversity in
case of frequency division and time shifting, and time diversity in case of frequency shifting, time division and random weighting technique.

To see the effect of multiple transmit antennas, we examine the capacities as the number of transmit antennas goes to infinity. For unconstrained signaling, as previously analyzed, $C_{opt} (2.44)$ converges almost surely to the capacity $\log_2 (1 + \rho \sigma^2_0)$ of an equivalent additive white Gaussian noise channel as $M \to \infty$. Similarly, the capacity of other diversity schemes also converge to its expected value as $M \to \infty$, although the limit in this case is equal to the expected capacity of a single antenna fading channel. For time and frequency division (2.45), applying the Strong Law of Large Numbers, we have

$$C_{td} = C_{fd} \xrightarrow{a.s.} E[\log_2 (1 + \rho \cdot \chi^2_2)].$$

For time and frequency shifting (2.47), it is proven [12] that the capacity converges in a mean-squared sense (i.e. the variance approaches zero as $M \to \infty$) to its expected value

$$C_{ts} = C_{fs} \xrightarrow{m.s.} E[\log_2 (1 + \rho \cdot \chi^2_2)].$$

The result has not been proven for the random weighting case but it is conjectured that convergence will apply. Thus, as the number of transmit antennas become large, the variability in the channel effectively disappears. This means that the outage probability is effectively driven to zero by using many transmit antennas, giving way for a reliable communication rate. Again, this suggests that the effect of multiple transmit antennas is to combat fading in the channel, therefore asymptotically transforming a random fading channel into a deterministic one.

**Combining receive and transmit diversities**

Here we give examples of some diversity methods combining multiple transmit and receive antennas. The first method is spatial cycling using one transmitter at a time with optimum combining receive diversity. The cycling ensures nontrivial dwelling of the $M$ transmit antennas and therefore helps to reduce the probability of being stuck with a bad channel for long time. This is the same principle as time-division transmit diversity. The capacity in this case is given by

$$C = \frac{1}{M} \sum_{i=1}^{M} \log_2 (1 + \rho \cdot \chi^2_{2N}).$$
The variance of this random capacity goes to zero as $M \to \infty$, even more rapidly than the receive diversity only case (2.39). In fact, of all the diversity methods mentioned here, this method exhibits the smallest variance.

When the number of transmit and receive antennas are the same ($M = N$), the spatial nulling-out method can be applied, where the transmit antennas send independent signals with the same power, and receiver $k$ only tries to detect message from transmit antenna $k$ and treats signals from other antennas as noise by nulling out all other signals. This approximates having $N$ parallel channels. When there is no interference between these parallel channels, i.e. $A = I_N$, the capacity is given by

$$
C = N \log_2 \left(1 + \frac{\rho}{N}\right) .
$$

The capacity of the real channel is understandably less due to interference between the transmit antennas. Moreover, the nulling process can enhance the noise significantly in some cases. An improvement to this method is to cycle the separately encoded data streams over all the transmit antennas instead of sending each of them to a fixed transmit antenna. The cycling helps to average out the capacity of all of the sub-streams and thus the total capacity is equal to the sum of the capacity of all sub-streams, whereas without cycling, we are stuck with the capacity equal to the number of antennas times the capacity of the worst sub-channel.

Another relatively new diversity technique [26] uses the same number of transmit and receive antennas, where the transmit antennas also sends independent sub-streams with equal power. The receive signals are processed in the following way: the signal from the first transmit antenna is detected using optimum combining, where signals from all other antennas are treated as noise, then the detected signal is extracted from the received signal and the process repeats until the signal from the last transmit antenna is detected. The technique is termed layered spaced-time architecture. Again transmit cycling can be employed to average out the fading effect in each sub-channel. The capacity equation of an $(N, N)$ transmit and receive antennas system in this case is given by [26]

$$
C = \sum_{k=1}^{N} \log_2 \left(1 + \frac{\rho}{N} \frac{\sigma^2}{N_{2k}}\right) \text{ bps/Hz} .
$$

This capacity is the same as the lower bound capacity of full antenna diversity in (2.42).

The probability distribution of the capacity of different diversity schemes were contrasted by numerical examples [10]. At a low outage probability of about 1%, the
systems capacities increase in the order: no spatial diversity, transmit diversity only, selection receive diversity, optimum combining receive diversity, transmit spatial cycling with optimum combining receiver, and layered spaced-time method, given the same number of antennas in each diversity. The spatial nulling-out methods only give large capacity at low probability of about 80 – 90%, and their capacities reduce considerably to less than all other diversity methods at 1% outage probability.

Receive diversity has been studied quite extensively. More recently, transmit diversity is receiving increasing attention ([20], [21], [23], [22], [15], [17], [18], [12], [16], [25]). The effect of transmit diversity only is to combat fading in the channel. In the following chapter, we will provide a review of the transmit diversity techniques available in the literature.
Chapter 3

Transmit Antenna Diversity Techniques

In this chapter we will analyze the performance of transmit diversity and receive diversity, and provide a literature review of current transmit antenna diversity techniques. Again we focus on the quasi-stationary channel model, where the channel is assumed to be frequency non-selective and slow fading so that it can be represented by multiplicative independent special complex Gaussian random coefficients which reflects Rayleigh or Rician fading.

Transmit diversity only, as shown in the previous chapter, provides less capacity than receive diversity only. Receive diversity has been studied extensively and the optimum method of maximal ratio combining has been derived. Transmit diversity techniques are thus often compared to receive diversity schemes having the same number of antennas. In most cases, the receiver is assumed to know the channel coefficients perfectly whereas the transmitter does not possess such knowledge. The Rayleigh fading channel model is employed in all the diversity techniques discussed below.

Section 3.1 analyzes two transmit diversity structures: a basic transmit diversity where the same signal is sent from all transmit antennas, and an orthogonal transmit diversity structure where signals transmitted from different antennas are orthogonal. These structures are compared with receive diversity using optimum combining. Analysis shows that in a Rayleigh flat fading channel, when all the transmit antennas send the same signal, the system performance is exactly the same as a single antenna channel given the same total transmit power. The orthogonal structure, on the other hand, achieves a diversity gain equal to that of receive diversity using optimum combining.
However, it does this at the cost of a bandwidth expansion factor equal to the number of transmit antennas. Thus a useful transmit diversity technique must process the transmit signal and send different signals from different antennas so that a diversity gain in performance is possible, while introducing no bandwidth expansion or only a negligible amount.

Various transmit diversity techniques have been studied in the literature. These include sending delayed versions of the same signal from different transmit antennas [21, 20], using modulation diversity at the transmit antennas [23, 22], introducing phase shifts at different antennas [24, 25], imposing linear antenna pre-filters [17], and coding [20] using a space-time block code [27, 28]. In all cases, it is shown that the system performance improves as the number of transmit antennas increases. The antenna pre-filter is a generalization of the other linear (non-coding) methods, in which the data signal is processed by a separate linear filter prior to transmitting from each antenna. The linear prefilters are chosen so that the tap weights of these filters form the rows of a unitary matrix, which is similar to modulation diversity [22] and includes time shifting diversity [21, 20] and phase shifting diversity [25] as specific cases. However, unlike most of the other techniques where maximum likelihood sequence estimation (MLSE) is used at the receiver, which has complexity growing exponentially with the sequence length, the linear antenna pre-filter scheme employs a simple linear equalizer to detect the transmit symbols. This linear prefilter scheme asymptotically transforms a flat and slow Rayleigh fading channel into a non-fading AWGN channel. In the following Sections from 3.2 to 3.6, we will attempt to describe these techniques and draw an overall picture of the available transmit diversity techniques.

3.1 Transmit and receive antenna diversity

We study two structures of transmit diversity in this sections. The first one is a basic scheme where all transmit antennas send the same signal, and we compare this with receive diversity using optimum combining. It is found that no diversity gain can be obtained from this basic transmit diversity scheme for a Rayleigh flat fading channel, which suggests that signals transmitted from separate antennas must be different in order to obtain a diversity gain. For the second part, we analyze an orthogonal transmit diversity scheme where the signals transmitted from different antennas are orthogonal to each other. This scheme is shown to achieve a diversity gain equivalent to that of re-
ceive diversity using optimum combining, but at the cost of large bandwidth expansions which are not always feasible in wireless communication. An effective transmit diversity scheme therefore must not expand the bandwidth while still achieving diversity gain.

3.1.1 Basic transmit diversity and receive diversity using optimum combining

In this section we analyze basic transmit diversity where all transmit antennas send the same signal and compare this with receive diversity using optimum combining. It will be shown that for Rayleigh flat fading, sending the same signals from all transmit antennas does not improve the system’s performance over the single antenna fading channel.

Consider the basic $M$ transmit diversity case where the same signal $x[n]$ is sent from different transmit antennas with equal power $P/M$. The receive signal can be written as

$$y[n] = x[n] \sum_{i=1}^{M} a_i + w[n]$$

where $w[n]$ is white Gaussian noise with variance $\sigma_N^2$. This is equivalent to a single antenna Gaussian fading channel with fade coefficient

$$a_i = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i$$

(3.1)

given the same transmit signal $x[n]$ with total power $P$.

On the other hand, receive diversity with $M$ receive antennas using optimum combining (OC), as suggested by (2.43), is equivalent to a single antenna Gaussian fading channel with fade coefficient

$$a_r = \left( \sum_{i=1}^{M} |a_i|^2 \right)^{1/2}.$$  

(3.2)

The equivalent receive signal for receive diversity with optimum combining can be written as

$$y[n] = x[n] \left( \sum_{i=1}^{N} |a_i|^2 \right)^{1/2} + w[n].$$

These equivalent fading channels for transmit diversity and receive diversity using optimum combining are depicted in figure 3.1.
Figure 3.1: Equivalent models for receive OC (a) and transmit (b) diversities

Since the channel is flat, there is no intersymbol interference and a symbol-by-symbol detector is optimum. As the additive noise has constant power, the instantaneous symbol error probability for a given symbol alphabet is determined by the power of the desired receive signal (that is the receive signal excluding noise) or equivalently the receive SNR. From (3.1), the average received SNR for transmit diversity is

$$ \rho_r = \frac{P}{\sigma_N^2} \cdot \frac{1}{M} E \left[ \sum_{i=1}^{M} a_i^2 \right] = \frac{P}{\sigma_N^2} \cdot \sigma_a^2 = \rho $$

which equals the average SNR for a single antenna fading channel given the same transmit power, whereas the average SNR for receive diversity with optimum combining is

$$ \rho_r = \frac{P}{\sigma_N^2} \cdot E \left[ \sum_{i=1}^{M} |a_i|^2 \right] = M \rho . $$

Due to the exponential decay of the error function, the average symbol error probability is dominated by the probability of deep fades in the instantaneous signal power. For receive antenna diversity, the probability of deep fades is reduced since the signal components from each path always add up constructively due to processing at the receiver. For transmit diversity however, the receiver cannot resolve the individual paths from the different transmit antennas, thus signals from different antennas can add up constructively or destructively. This difference between transmit and receive diversity is reflected in the normalized variance of the instantaneous receive signal power. In Rayleigh fading where the $a_i$ are zero-mean special complex Gaussian random variables, the equivalent fade coefficient $a_i$ for transmit diversity (3.1) is a linear combination of these Gaussian random variables and therefore is also a special complex Gaussian random variable, with zero mean and variance $\sigma_a^2$. Thus the receive power variance normalized by the average receive power for transmit diversity is

$$ \frac{\sigma_{\text{var}}^2}{\left( E[|a_i|^2] \right)^2} = \frac{E \left[ \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i \right]^4}{\left( E \left[ \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i \right]^2 \right)^2} - 1 = \frac{2\sigma_a^4}{\sigma_a^2} - 1 = 1 $$

(3.3)
since for zero-mean special complex Gaussian random variables, $E|a_i|^4 = E|a_i|^4 = 2\sigma_a^4$ (see Appendix, Lemma 1). For receive diversity, the normalized receive power variance is

$$
\frac{\sigma_{a_r}^2}{(E|a_r|^2)^2} = \frac{E[\sum_{i=1}^{M} |a_i|^2]^2}{(E[\sum_{i=1}^{M} |a_i|^2]^2)^2} - 1 = \frac{\sum_i E|a_i|^4 - M \sigma_a^4}{M^2 \sigma_a^4} = \frac{1}{M}.
$$

(3.4)

Thus for receive antenna diversity, the normalized variance is inversely proportional to the number of antennas whereas for transmit antenna diversity, this variance is independent of the number of antennas. As a result, the probability density function of the instantaneous receive signal power does not depend on the number of transmit antennas for a given average power.

The analysis above shows that in a Rayleigh flat fading channel, when the signals transmitted from all antennas are the same, multiple transmit antennas do not perform any better than a single transmit antenna since the mean and the variance of the instantaneous receive power are the same in both cases. Receive diversity with optimum combining on the other hand increases the average receive power and reduces the power variance by the number of antennas. The reason is that in Rayleigh fading, the transmit diversity equivalent fading coefficient $a_t$ (3.1) has exactly the same distribution as individual fading coefficients $a_i$. This suggests that a diversity gain can only be obtained using multiple transmit antennas if the signals sent from each antenna are different.

### 3.1.2 Orthogonal transmit diversity

This section will describe a method of sending the same digital information from each transmit antenna but with different modulation parameters so that the signals transmitted from different antennas are orthogonal. It is somewhat similar to the classical theory of $M$-ary orthogonal signaling in additive white Gaussian noise channels, which can achieve arbitrary small error rate as $M$ increases to infinity, provided the transmit SNR is larger than a certain value. This orthogonal signaling method, however, expands the bandwidth requirement by a factor equal to the number of transmit antennas. First we will describe the classical theory of $M$-ary orthogonal signaling in an AWGN channel, then we will apply this to orthogonal transmit diversity.

**$M$-ary orthogonal signaling in an additive white Gaussian noise channel**

First we will study the $M$-ary orthogonal signaling in an additive white Gaussian noise channel. $M$-ary orthogonal signaling consists of $M$ equal-energy orthogonal signal
waveforms which are a special case of multidimensional signals. These signals can be represented by $M$-dimensional vectors

$$
\mathbf{s}_1 = [\sqrt{\mathcal{E}} \ 0 \ 0 \ \ldots \ 0 \ 0 ] \\
\mathbf{s}_2 = [0 \ \sqrt{\mathcal{E}} \ 0 \ \ldots \ 0 \ 0 ] \\
\vdots \quad \vdots \\
\mathbf{s}_M = [0 \ 0 \ 0 \ \ldots \ 0 \ \sqrt{\mathcal{E}}]
$$

where $\mathcal{E}$ is the symbol energy. One way to construct these signals is by using frequency-shift keying modulation with lowpass base pulses defined as

$$
s_{im}(t) = \sqrt{\frac{2\mathcal{E}}{T}} \, e^{j2\pi m \Delta f t} \quad m = 1, 2, \ldots, M \quad 0 \leq t \leq T
$$

where $T$ is a signaling interval, i.e. a pulse duration. The value of $\Delta f$ is chosen to be $\Delta f = \frac{1}{2T}$ to ensure orthogonality [1].

In an AWGN channel, the optimum detector for these signals selects the signal resulting in the largest cross-correlation between the received vector $\mathbf{y}$ and each of the $M$ possible transmitted vectors $\{\mathbf{s}_i\}$, where the cross-correlation is defined as

$$
R(\mathbf{y}, \mathbf{s}_i) = \mathbf{y} \cdot \mathbf{s}_i = \sum_{k=1}^{M} y_k s_{ik}, \quad i = 1, 2, \ldots, M.
$$

For example, suppose $\mathbf{s}_1$ is the transmitted signal, then the received vector is

$$
\mathbf{y} = [\sqrt{\mathcal{E}} + n_1 \ n_2 \ \ldots \ n_M]
$$

where $n_i, i = 1 \ldots M$ are zero-mean, mutually independent Gaussian random variables with equal variance $\sigma_N^2$. A bank of $M$ decorrelators is used at the receiver to produce the outputs

$$
R(\mathbf{y}, \mathbf{s}_1) = \sqrt{\mathcal{E}} (\sqrt{\mathcal{E}} + n_1) \\
R(\mathbf{y}, \mathbf{s}_i) = \sqrt{\mathcal{E}} \cdot n_i, \quad i \neq 1.
$$

In the case of binary orthogonal signals ($M = 2$), it is shown [1] that the average probability of symbol error is given by

$$
P_2 = Q\left(\frac{\mathcal{E}}{2\sigma_N^2}\right). \quad (3.5)
$$

An upper bound on the probability of symbol error for the general $M$ orthogonal signal case can be derived from this binary case. If we view the detector for $M$ orthogonal signals as one that makes $M-1$ binary decisions between the correlator output $R(\mathbf{y}, \mathbf{s}_1)$
that contains the signal and the other $M - 1$ correlator outputs $R(y, s_i)$, $i \neq 1$, the
probability of error is upper-bounded by the union bound of the $M - 1$ events. That
is, if $E_i$ represents the event that $R(y, s_i) > R(y, s_1)$ for $i \neq 1$, then the symbol error
probability $P_M$ of the $M$-ary orthogonal signaling is bounded by

$$P_M = P\left(\bigcup_{i \neq 1} E_i\right) \leq \sum_{i \neq 1} P(E_i)$$

and thus

$$P_M \leq (M - 1)P_2 = (M - 1)Q\left(\sqrt{\frac{E_s}{2\sigma_N^2}}\right) < M \cdot Q\left(\sqrt{\frac{E_s}{2\sigma_N^2}}\right)$$

where $P_2$ is the probability of a symbol error in binary orthogonal signals (3.5). This
bound can be further simplified by upper-bounding $Q\left(\sqrt{\frac{E_s}{2\sigma_N^2}}\right) < \exp\left(-\frac{E_s}{4\sigma_N^2}\right)$, thus

$$P_M < M \exp\left(-\frac{E}{4\sigma_N^2}\right) = 2^k \exp\left(-k\frac{E_b}{4\sigma_N^2}\right) = \exp\left[-k\left(\frac{E_b}{2\sigma_N^2} - 2\ln 2\right)\right]$$

where $k$ is the number of bits in a symbol ($k = \log_2 M$) and $E_b$ is the bit energy.
As $k \to \infty$, or equivalently the number of symbols $M \to \infty$, the probability of error
approaches zero exponentially, provided that the bit SNR satisfies $E_b/2\sigma_N^2 > 2\ln 2 \approx
1.42 dB$. However, this bound is quite loose, especially at sufficiently low SNR, due to
the bound on the $Q$ function being loose. A tighter upper bound can be derived for
$E_b/2\sigma_N^2 < 2\ln 2$ as [1]

$$P_M < 2 \exp\left[-k\left(\frac{E_b}{2\sigma_N^2} - \sqrt{\ln 2}\right)^2\right].$$

This bound is also tighter than the first one for large values of $M$. Consequently,
$P_M \to 0$ as $k \to \infty$ provided that $E_b/2\sigma_N^2 > \ln 2 \approx -1.6 dB$. This value $-1.6 dB$ is the
minimum required SNR per bit to achieve an arbitrary small probability of error in
the limit as $M \to \infty$, and it is called the Shannon limit of an additive white Gaussian
noise channel.

Although orthogonal signal modulation can provide arbitrarily low probability of
error as the signal dimension increases, it does so at the cost of bandwidth expansion.
Let $T$ be the baseband symbol duration. If the $M = 2^k$ orthogonal signals are con-
structed by means of orthogonal carriers with minimum frequency separation of $1/2T$
for orthogonality, then the bandwidth requirement for transmission of $k = \log_2 M$
information bits at rate $R = k/T \text{ bps}$ is

$$W = \frac{M}{2T} = \frac{M}{2(k/R)} = \frac{M}{2\log_2 M} R.$$
Thus the bandwidth increases as \( M \) increases. In terms of normalized data rate \( R/W \) (bps/Hz) versus the SNR per bit \( (\mathcal{E}_b/2\sigma_N^2) \) required to achieve a given error probability, the \( M \)-ary orthogonal signals yield \( R/W \ll 1 \) and this ratio decreases as \( M \) increases due to increasing bandwidth. However, the SNR per bit required to achieve a given error probability decreases as \( M \) increases [1]. As a result, \( M \)-ary orthogonal signals are useful for power-limited channels that have sufficiently large bandwidth to accommodate a large number of signals. These properties are opposite to other multiphase signaling schemes such as PSK (phase shift keying), QAM and PAM (pulse amplitude modulation). Since wireless communication usually has a limited power budget, this suggests a possible use of the orthogonal modulation technique here.

**Orthogonal transmit diversity**

The idea of orthogonal signals is applied to transmit antenna diversity by having orthogonal modulation base-pulses at different antennas. Consider QAM-type modulation, which covers a large class of linear modulation. The same sequence of information symbols \( I_n \) is modulated using different orthogonal pulses at different transmit antennas. The transmit signal from antenna \( i \) is

\[
x_i(t) = \sum_n I_n \cdot g_i(t - nT).
\]  

(3.6)

Where the base-pulses \( \{g_i(t)\} \) at different transmit antennas are orthogonal, i.e.

\[
\int_{-\infty}^{\infty} g_i(t) \cdot g_j(t - nT) = \delta_{ij} \cdot \delta_{n0}
\]

with the \( \delta_{nm} \) function defined as

\[
\delta_{nm} = \begin{cases} 
1 & n = m \\
0 & \text{otherwise}
\end{cases}
\]

Without loss of generality, it is assumed that \( \{g_i(t - nT); -\infty \leq n \leq \infty\} \) is a set of orthonormal functions, i.e.

\[
\int_{-\infty}^{\infty} g_i(t - n_1 T)g_j(t - n_2 T) = \delta_{n_1,n_2}.
\]

The received signal is

\[
y(t) = \sum_i \sum_n a_i \cdot I_n \cdot g_i(t - nT) + n(t).
\]
The maximum-likelihood receiver which consists of a bank of matched filters matched to the base-pulses \( \{g_i(-t)\} \), followed by sampling at rate \( T \), can be used in this case to detect the transmitted symbols (Figure 3.2). The sampled signal after the matched filter \( i \) is

\[
 r_i[n] = y(t) * g_i(-t) \big|_{t=nT} \\
 = \int_{-\infty}^{\infty} \sum_j \sum_k \alpha_j \cdot I_k \cdot g_j(t - \tau + kT)g_i(-\tau) d\tau \big|_{t=nT} + \int_{-\infty}^{\infty} n(t - \tau)g_i(-\tau) d\tau \big|_{t=nT}.
\]

Due to the orthogonality of the base-pulses, the above equation becomes

\[ r_i[n] = a_i \cdot I_n + w[n] \quad (3.7) \]

where \( w[n] \) are the T-spaced samples of \( n(t) * g_i(-t) \) and are thus also white Gaussian noise with the same variance as the original additive white noise of the channel because \( \{g_i(t)\} \) is an orthonormal set. But (3.7) is also the equation for the received signal at antenna \( i \) in receive diversity. Thus sampled outputs of the matched filter bank are the same as the received signals at different antennas in receive antenna diversity, and optimum combining can be used to detect the transmitted symbols. Therefore, by employing orthogonal signals at different transmit antennas, one can obtain performance equivalent to receive diversity using optimum combining. However, as discussed above, this method suffers a severe bandwidth expansion, which is not always feasible in communication channels.

![Figure 3.2: Maximum-likelihood receiver for transmit diversity with orthogonal base-pulses.](image)

In short, orthogonal transmit diversity, although achieving a diversity gain equal to that of receive diversity using optimum combining, is not a desirable technique since it requires a large transmission bandwidth. In the following sections, we will describe
transmit diversity techniques that result in no bandwidth expansion or a negligible amount. These includes time shifting diversity in Section 3.2 as a special case of modulation diversity in Section 3.3, phase shifting methods in Section 3.4 and a generalized linear antenna processing in Section 3.5. A non-linear transmit diversity technique involving coding is described in Section 3.6.

3.2 Time shifting

Time shifting transmit diversity is a technique where multiple transmit antennas are used to transmit delayed versions of a signal, therefore transforming a flat fading channel into frequency-selective fading [20, 21]. When maximum likelihood sequence estimation (MLSE) is used at the receiver to equalize the signal, results show [21] that transmit diversity with \( M \) transmit antennas provides a diversity gain within 0.1 dB of that with \( M \) receive antennas.

3.2.1 Signal model

In this scheme, the digital signal \( s(t) \) is transmitted by each antenna with a D second delay between each antenna (Figure 3.3). The total transmit power is constrained by \( P \) and each transmit antenna transmits with equal power \( P/M \). The delay between antennas is chosen so that the signals transmitted by each antenna are uncorrelated, i.e.,

\[
E[s(t)s(t + D)] = 0
\]

For uncorrelated receive signals from each transmit antenna in the flat fading case, the delay \( D \) has to be at least one symbol period \( T \) (i.e. \( D \geq T \)). To simplify the equalizer at the receiver, a delay of \( D = T \) is considered (in the case of frequency selective fading, i.e., when delay spread is significant, a longer delay is needed for uncorrelated receive signals, for example with delay spread of \( \pm T \) then \( D \geq 2T \) is needed). The output of the channel in discrete form with T-spaced samples (2.22) is therefore

\[
y[n] = \sum_{i=1}^{M} a_i \cdot s[n - i + 1] + w[n].
\]

When \( s[n] \) is a linearly modulated signal as in (2.24), the above expression can be re-written as

\[
y[n] = \sum_{i=1}^{M} a_i \sum_{k} I_k \cdot g[n - i + 1 - k] = \sum_{k} I_k \sum_{i=1}^{M} a_i \cdot g[n - i + 1 - k].
\]
Let $g_a[n]$ be the equivalent baseband channel up to the receiver front end then

$$g_a[n] = \sum_{i=1}^{M} a_i \cdot g[n - i + 1]$$

and the received signal can be simplified to $y[n] = I_n \ast g_a[n]$.

![Diagram](image)

**Figure 3.3: Delayed transmit diversity model.**

### 3.2.2 Performance bounds of different equalization techniques

Here we present results on performance bounds and comparisons for three different equalizers: maximum likelihood sequence estimation (MLSE), linear equalizer (LE) and minimum-mean square error decision feedback equalizer (MMSE-DFE).

**MLSE performance bound and comparison to receive diversity**

The MLSE is the optimum equalizer which will give the lowest error, but also is the most complicated and computational intensive equalization method based on Viterbi algorithm [1]. Consider the ideal MLSE with infinite length. The optimum criterion is designed to detect a sequence of symbols $\hat{\mathbf{I}} = (\ldots \hat{I}_0, \ldots, \hat{I}_l, \ldots)$ such that

$$\hat{\mathbf{I}} = \arg \min_{\mathbf{I}} \sum_{n} \left| y[n] - \sum_{k} \hat{I}_k \cdot g_a[n - k] \right|^2.$$  

With MLSE, the probability of bit error rate (BER) for a given channel can be well approximated given the probability of minimum distance error event [20], (accurately for low BER), by the relation

$$P_b \approx Q \left( \sqrt{\frac{\sigma_m^2}{\sigma_N^2}} \right)$$

where the function $Q(\cdot)$ is given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-t^2/2} dt$$
and $d_{min}^2$ is the minimum distance over all possible error events, defined as

$$d_{min}^2 = T \min_e |e \ast g_a|^2 = T \min_{e_n} \left| \sum_n e_n \cdot g_a[k - n] \right|^2$$

with $e_n = \hat{I}_n - I_n$, $e = \ldots e_0 \ldots e_l \ldots$ is the error sequence and $g_a$ is the impulse response sequence of the discrete channel.

The error performance of MLSE can be lower bounded by the matched filter bound, based on known results. The matched filter bound assumes that all the past and future data symbols have been decoded correctly and that a decision is to be made about a transmitted symbol $I_0$ [20]. The probability of BER for the matched filter bound is also given by (3.8), where the minimum distance is the distance squared between two data sequences which are identical in all positions except one, and thus proportional to the overall equivalent channel energy

$$d_{min}^2_{|MFB} = T |e_0|^2 \left| \sum_k g_a[k] \right|^2 = T \mathcal{E}_g^s \cdot |e_0|^2 \sum_{i=1}^{M} |a_i|^2$$

where $\mathcal{E}_g^s$ is the energy of the discrete pulse $g[n]$. Assuming that $g(t)$ has a frequency spectrum that is the square root of a Nyquist pulse spectrum, then $T \mathcal{E}_g^s = 1$ and the minimum distance for the matched filter bound is

$$d_{min}^2_{|MFB} = |e_0|^2 \sum_{i=1}^{M} |a_i|^2. \quad (3.9)$$

But this is also in the same form as the output signal power of maximum ratio combining receiver diversity (2.43), therefore the performance of the matched filter bound is the same as that of receive diversity, except for a possible proportionality in gain due to the symbol error energy $|e_0|^2$. When the transmit data is a 4-level complex quadrature amplitude modulation (QAM) where the information symbols are of the form $I_n = \alpha_n + j \beta_n$ with $\alpha_n, \beta_n = \pm 1$, the average BER (3.8), averaging over the fading channel statistic is [20]

$$P_{\bar{b}|MFB} = \left( \frac{1 - \mu}{2} \right) \sum_{k=0}^{M-1} \binom{M-1-k}{k} \left( \frac{1 + \mu}{2} \right)^k \quad (3.10)$$

where $\mu = \sqrt{\frac{2}{1 + \rho}}$ and $\rho$ is the SNR in (2.31). This BER decreases exponentially with the number of antennas $M$, which shows that receive diversity using optimum combining achieves an $M$-fold diversity gain.

Thus, the comparison between MLSE and the matched filter with an $M$-symbol-spaced impulse response is the same as the comparison between time shifting transmit
diversity using ideal MLSE and receive diversity using optimum combining. Previous results have shown that the MLSE receiver can achieve the matched filter bound for any flat fading channel with \( M = 2 \). For \( M > 2 \), there is some degradation in performance of the MLSE compared to the matched filter bound. Define the degradation as

\[
\text{Degradation} = \frac{d^2_{\text{min}}}{d^2_{\text{min}}|\text{MLSE}}.
\]

This degradation and \( d^2_{\text{min}} \) are random variables since the channel response is a random variable. Using simulation, the probability distribution of the degradation and the effect of the degradation on the average BER and the distribution of BER are studied in [21] for the case the input data is binary phase shift keying (BPSK). In this model, the input is an independent binary random sequence \( I = \{I_n\} \) with outcomes \( \pm 1 \) equally likely.

Results on the probability distribution of the degradation for \( M = 3, 4, 6, 10, 20 \) and 30 show that the probability that MLSE cannot achieve the matched filter bound on a given channel decreases with increasing number of transmit antennas \( M \). For \( M = 3 \), this probability is less than 9% whereas for \( M = 30 \), the MLSE achieved the matched filter bound for all but one channel out of 10,000 simulated. Of interest is also the worst case degradation. Previous results have found the worst degradation for a real channel in the cases \( M = 3, 4, 6 \) to be 2.3, 4.2, and 7.0 dB respectively. For a complex fading channel, it is expected that the worse degradation will be greater. However, since the channel is random, these worst case degradations and the channels for which MLSE cannot achieve the matched filter bound, occur with some probability. Simulation results found these worst degradations at probability greater than \( 10^{-4} \) quite significantly smaller than those known for real channels: for \( M = 3, 4, 6 \), the simulated worst degradations are 2.2, 3.6 and 5.2 dB respectively. As \( M \) increases, the worst degradation increases and the probability of worst case degradation decreases (at least for probabilities greater than \( 10^{-4} \)).

Next, the effect of the degradation on the average BER and distribution of BER were studied. Since the degradation occurs with low probability, the effect of the degradation on average BER can be negligible. Thus in a rapidly fading environment where the average BER is of prime interest, transmit diversity can achieve the full diversity gain as receive diversity. However, in a slow fading environment, the effect of the degradation on the distribution of the BER must be considered. This effect depends on the value of \( d^2_{\text{min}}|\text{MFB} \) for each channel where the degradation occurs. If a large degradation occurs
only for large values of \(d_{min}^2|_{MBF}\) then it does not affect the distribution of the BER significantly, whereas if it happens for small values of \(d_{min}^2|_{MBF}\), the BER distribution can be severely affected. Again by simulation, it is found that probability distributions of \(d_{min}^2|_{MLSE}\) and \(d_{min}^2|_{MBF}\) differ by less than 0.1 dB in normalized \(d_{min}^2\) (normalized to its mean value). Results show that the channels for which MLSE cannot achieve the matched filter bound are generally not the channel with low \(d_{min}^2|_{MBF}\). Thus the degradation with MLSE does not significantly affect the BER distribution. It was concluded that transmit diversity with MLSE achieves within 0.1 dB of the diversity gain of the receive diversity even in a slow fading environment [21].

In [21], the performance of MLSE time shifting transmit diversity was only compared with optimum receive diversity in terms of degradation but not BER. It is not clear how the 0.1 dB difference in probability distribution of \(d_{min}^2\) in the two cases affect BER. But this shows that the average BER of time shifting transmit diversity, taken as the expected BER value over the distribution of \(d_{min}^2\), will be close to that of optimum receive diversity.

**Linear equalization and decision feedback equalization**

As alternatives to MLSE, linear equalizer (LE) or minimum mean square error with decision feedback equalizer (MMSE-DFE) can be used to tradeoff between complexity and performance.

For simplicity, the error bound for a zero-forcing optimum combiner (where the coefficients of all symbols other than the one being detected are forced to be zero after equalization) was considered. The only interference in the zero-forcing equalizer is the output Gaussian noise. Using quadrature phase shift keying (QPSK) with coherence detection (the information symbols are \(I_n = \{00,01,11,10\}\) with equal probability), a symbol error will result when the noise after equalization is greater than \(1/\sqrt{2}\) which is half the distance between two adjacent symbols. Using the Chernoff bound, it can be shown that the bit error rate is exponentially tightly upper bounded by [20]

\[
P_b \leq \exp\left(-\frac{1}{2 \cdot MSE}\right)
\]

where \(MSE\) is the mean square error. In the case of LE, the \(MSE\) for a zero-forcing solution is given by [34]

\[
MSE = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{N_0}{|A(\omega)|^2} \, d\omega
\]
and for the MMSE-DFE it is given by

\[ MSE = \exp \left[ -\frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \ln \left( \frac{N_0}{|A(\omega)|^2} \right) d\omega \right] \]

where \( N_0 \) is the noise power spectrum density and the time shifting channel \( A(\omega) \) is given in (2.46).

Using simulation, the performance of LE and MMSE-DFE were studied for QPSK signals in [20] for the cases when the number of transmit antennas \( M = 2 \) and 4. For both equalizers, the system performance improves with the number of transmit antennas, and the MMSE-DFE shows superior performance compared to the LE given the same number of antennas, although both are worse than the matched filter case. With \( M = 2 \), DFE requires 4 dB more SNR than the matched filter and LE requires 6 dB more to obtain the same BER of at least \( 10^{-4} \). The system performance in the asymptotic cases where the number of antennas goes to infinity were also compared, and results show that \( M = 4 \) improves the performance over \( M = 2 \) by about half (in terms of extra SNR required in dB) of the maximum performance improvement possible as \( M \to \infty \) for both equalizations.

### 3.3 Modulation diversity

In this section, we will describe a modulation diversity technique that does not expand the bandwidth and still provides diversity gain with multiple transmit antennas, in some cases equals to that of receive diversity gain [23, 22]. However, the method increases the receiver complexity compared to the orthogonal transmit diversity described in Section 3.1.2, since it requires adaptive equalization to achieve a diversity gain.

The motivation behind this modulation diversity is to minimize the normalized variance of the instantaneous signal power, which is equivalent to minimizing the normalized variance of the squared Euclidean distance in a channel with intersymbol interference. In effect, this will minimize the average error probability as analyzed in Section 3.1.1.

The proposed scheme is depicted in Figure 3.4. Each branch is QAM modulation of the same base-pulse \( g(t) \), therefore there is no bandwidth expansion. Modulation diversity is performed by the symbol spaced finite impulse response (FIR) pre-filters \( \tilde{h}_i \). These filters perform a “precoding” of the input symbols, so that the transmitted
signal from antenna $i$ is

$$u_i[n] = \sum_k x[k] \cdot h_i[n - k] .$$

(3.11)

### 3.3.1 Optimization criterion

The aim is to optimize the linear filters $\tilde{h}_i$ such that the normalized variance of the squared Euclidean distance is minimized. This is a logical extension of the measure “normalized variance of the instantaneous receive signal power” in a channel with intersymbol interference which is introduced by the pre-filters.

Let $L$ be the length of the pre-filter $\{\tilde{h}_i\}$. The average received signal power should be the same for different sets of prefilter weights to be comparable, thus the tap weights must satisfy

$$\sum_{i=1}^{M} \sum_{n=1}^{L} |h_i[n]|^2 = 1 .$$

(3.12)

The Euclidean distance between two received signal sequences $r^{(1)}$ and $r^{(2)}$, which correspond to the input symbol sequences $\tilde{x}^{(1)}$ and $\tilde{x}^{(2)}$, is

$$d_E^2 = \sum_n \left| r^{(1)}[n] - r^{(2)}[n] \right|^2 = \sum_n \left( \sum_{i=1}^{M} \sum_{k} \gamma[k] \cdot h_i[n - k] \right)^2$$

where $\gamma = \tilde{x}^{(1)} - \tilde{x}^{(2)}$ denotes the sequence of symbol differences. Let $R_{hh}$ denote the correlation of the pulse responses of the pre-filters, i.e

$$R_{hh}[n_1, n_2, i_1, i_2] = \sum_k h_{i_1}[k - n_1] \cdot h_{i_2}^*[k - n_2]$$
then the Euclidean distance becomes

\[ d_E^2 = \sum_{i_1}^{M} \sum_{i_2}^{M} \sum_{n_1}^{n_2} \sum_{n_2}^{n_2} a_{i_1} \cdot a_{i_2}^* \cdot \gamma[n_1] \cdot \gamma^*[n_2] \cdot R_{hh}[n_1, n_2, i_1, i_2]. \]  

(3.13)  

The average Euclidean distance is given by

\[ E_{a_1, a_2, \ldots, a_M}[d_E^2] = \sigma_a^2 \sum_{i=1}^{M} \sum_{n_1}^{n_1} \sum_{n_2}^{n_2} \gamma[n_1] \cdot \gamma^*[n_2] \cdot R_{hh}[n_1, n_2, i, i]. \]  

(3.14)  

The average Euclidean distance is the sum of the average squared Euclidean distance obtained in each branch, and it does not depend on the cross-correlation of the pulse responses of individual pre-filters. This is a result of the statistical independence of the path coefficients.

From (3.13) and (3.14) above, the variance of the squared Euclidean distance between two received sequences can be written as

\[ \sigma_{d_E}^2 = (E|a|^4 - \sigma_a^4) \sum_{i_1}^{M} \sum_{i_2}^{M} \sum_{n_1}^{n_1} \sum_{n_2}^{n_2} \gamma[n_1] \gamma^*[n_2] \gamma[m_1] \gamma^*[m_2] R_{hh}[n_1, n_2, i_1, i_2] R_{hh}[m_1, m_2, i, i] 
+ \sigma_a^4 \sum_{i_1, i_2 \neq i_1}^{M} \sum_{n_1}^{n_1} \sum_{n_2}^{n_2} \sum_{m_1}^{m_1} \sum_{m_2}^{m_2} \gamma[n_1] \gamma^*[n_2] \gamma[m_1] \gamma^*[m_2] R_{hh}[n_1, n_2, i_1, i_2] R_{hh}[m_1, m_2, i_2, i_1]. \]  

(3.15)

The first term is the sum of the variance in a single transmit branch (i.e., the path between a transmit antenna and the receiver), whereas the second term represents the covariance between different transmit diversity branches.

The optimum pre-filters tap weights follow from the optimum value for the pulse correlation \( R_{hh} \). Equation (3.15) shows that the optimal \( R_{hh} \) that minimizes the Euclidean variance depends on the sequence of symbol differences \( \gamma \) being considered. Since the sequence of symbol differences in turn depends on the input symbol sequences \( \bar{x}^{(1)} \) and \( \bar{x}^{(2)} \), there is no general expression for the Euclidean distance when input symbol sequences are arbitrary. Though in the case when the input sequences are single symbols, the Euclidean distance is obtained for a sequence of symbol differences with only one non-zero entry. In the following, we provide the analysis for this case, i.e., minimizing the variance of squared Euclidean distance between the output sequences corresponding to single symbol inputs.

### 3.3.2 Minimizing the squared output Euclidean distance variance for single symbol inputs

When the input sequences \( \bar{x}^{(1)} \) and \( \bar{x}^{(2)} \) are single symbols, the sequence of symbol differences has only a single non-zero entry \( \gamma_0 \). Thus the normalized variance of the
squared Euclidean distance follows from (3.14) and (3.15) as

\[
\frac{\sigma^2_d}{E[d_E^2]} = (E|a|^4 - \sigma_a^4) \cdot \sum_i |R_{hh}(0, 0, i, i)|^2 + \sigma_a^4 \cdot \sum_{i_1} \sum_{i_2 \neq i_1} |R_{hh}(0, 0, i_1, i_2)|^2. \tag{3.16}
\]

Note that the constraint (3.12) implies

\[
\sum_{i=1}^M R_{hh}[0, 0, i, i] = 1 \tag{3.17}
\]

thus the mean squared Euclidean distance in this case is

\[
E[d_E^2] = |\gamma_0|^2.
\]

The first summation in (3.16) depends only on the autocorrelation function of the pulse response of the pre-filters. It describes the impact each branch has on the normalized variance. The same variance would be obtained for receive diversity with optimum ratio combining. Employing the inequality \(\sum \alpha_i^2 \leq \sum |\alpha_i|^2\), with the condition (3.17), this sum achieves its minimum value \(1/M\) when the autocorrelation of all pre-filters are equal

\[
R_{hh}[0, 0, i, i] = \frac{1}{M} \forall i = 1 \ldots M. \tag{3.18}
\]

The above condition implies that transmit power from all antennas should be equal, i.e. each antenna thus transmits with power \(P/M\). This constraint is called equal power constraint.

The second summation in (3.16) describes the impact of the cross-correlation of the pre-filter pulse responses on the variance. The result is always non-negative, thus the optimum ratio combining receive diversity sets a lower bound on the variance of the Euclidean distance obtained in transmit diversity. This second summation can be minimized by choosing the pre-filter coefficients such that the resulting impulse responses are orthogonal for a zero time shift, i.e.

\[
R_{hh}[0, 0, i_1, i_2] = \sum_n h_{i_1}[n] \cdot h_{i_2}^*[n] = 0 \quad \forall i_1 \neq i_2. \tag{3.19}
\]

This condition is called the orthogonality condition. It is convenient here to introduce a matrix notation of the FIR pre-filters tap weights. Let \(\mathbf{H}\) be a \(M \times L\) matrix in which the row \(i\) contains the tap weights for FIR filter \(i\), i.e.

\[
H[i, n] = h_i[n]
\]
then the orthogonality condition and the equal power constraints become

\[ \mathbf{H} \cdot \mathbf{H}^T = \frac{1}{M} \cdot \mathbf{I}_M \]  \hspace{1cm} (3.20)

where \( \mathbf{I}_M \) is the \( M \times M \) identity matrix.

Equation (3.20) does not have a unique solution, as described in [22]. For the case \( L > M \), i.e. the number of pre-filter taps is greater than the number of transmit antennas, several solutions are available which all satisfy the optimum criterion. However, the intersymbol interference introduced by the pre-filters means a Viterbi equalizer is needed at the receiver, which increases the complexity exponentially with the number of taps. For this reason, \( L > M \) is usually not desirable. Only for cases where the propagation delay from different transmit antennas are different (for example, due to transmit antennas being placed very far apart) then \( L > M \) can be beneficial.

For the case \( L = M \), equation (3.20) has a number of solutions. For example, one solution is

\[ \mathbf{H} = \frac{1}{\sqrt{M}} \cdot \mathbf{I}_M \]

i.e., each pre-filter has only one non-zero tap and implements a delay. This scheme then becomes exactly the same as the time shifting scheme described in the previous section, where the \( i \)th antenna transmits the signal with a delay \( \tau_i = (i - 1) \cdot T \), where \( T \) is a symbol period. This solution however is not unique. All possible solutions in this case are equivalent not only with respect to the optimization criterion of minimizing the variance of the squared Euclidean distance, but also with respect to the system performance (averaged over the fading coefficients). It can be seen that the \( M \) fading branches with FIR pre-filters \( \mathbf{h}_i \) and fading coefficients \( a_i \) can be merged into one branch with an FIR filter \( \mathbf{h} \) with tap weights defined as

\[ h[n] = \sum_{i=1}^{M} a_i \cdot h_i[n] \] \hspace{1cm} (3.21)

Since each channel coefficient \( a_i \) is a special complex Gaussian random variable, the equivalent filter is completely characterized by the cross correlation matrix of its tap weights. Let \( \mathbf{a} = (a_1, a_2, \ldots, a_M)^T \) denote the fading vector, we have

\[ \mathbf{h} = \mathbf{H}^T \cdot \mathbf{a} \]

and the cross-correlation matrix follows as

\[ \Lambda_{hh} = E[\mathbf{h} \cdot \mathbf{h}^\dagger] = E[\mathbf{H}^T \cdot \mathbf{a} \cdot \mathbf{a}^\dagger \cdot \mathbf{H}] = \sigma_a^2 \cdot \mathbf{H}^T \mathbf{H} \]
since the fading coefficients are statistically independent special complex Gaussian with zero mean. With condition (3.20), the cross-correlation matrix becomes

\[ \Lambda_{hh} = \frac{\sigma^2_a}{M} \cdot I_M . \]  

(3.22)

Thus for \( L = M \), the system’s performance averaged over the fading vector \( \tilde{a} \) is the same for all sets of pre-filters \( \tilde{h}_i \) satisfying (3.20). Equation (3.22) shows that the optimum equivalent filter \( \tilde{h} \) has statistically independent identically distributed (i.i.d.) taps. This means that by adding pre-filters satisfying (3.20), the channel is effectively transformed from a transmit antenna diversity system into a time diversity system where the fading coefficients in adjacent “time slots” are independent.

For the case \( L < M \), we can map all branches into a stochastic filter \( \tilde{h} \) as described in the previous case (3.21), and the best we can obtain are \( L \) statistically independent taps. This requires no more than \( L \) transmit antennas, therefore all cases with more antennas than filter taps have an \( L \) antenna equivalent.

Employing MLSE at the receiver, simulation results in [23] and [22] show an improvement in BER at a given SNR for a system with two transmit antennas compared to a single antenna system, and also provide a comparison between transmit diversity and maximum ratio combining receive diversity. Compared to a system with no diversity, two transmit antenna diversity requires 4.5 dB less total transmit power at a BER of \( P_b = 10^{-2} \) and 11 dB less power at a BER of \( P_b = 10^{-3} \). Compared with receive diversity, the simulations show that transmit diversity with two transmit antennas using MLSE achieves the same performance as equivalent receive diversity with optimum combining. For the case of three transmit antennas however, a loss in performance compared to receive diversity optimum combining is observed at low SNR. These observations agree with the simulation results for the time shifting transmit diversity technique with MLSE described in the previous section, which was also derived independently in [20].

3.3.3 Impact of path propagation delay differences and noisy channel estimates

The impact of propagation delay differences between different transmit branches and of noisy channel estimation on system performance were also examined in [23, 22]. All transmit diversity methods, besides application in a multiple antenna device, can also have application in simulcast situations where several adjacent base-stations simulta-
neously transmit the same message [23]. From the start, it has always been assumed that the delay differences between different transmit antennas is negligible, but when the transmit antennas are placed very far apart as in simulcast situations, this may no longer be the case. With different path delays, additional intersymbol interference (ISI) arises apart from the ISI caused by the pre-filters, and this can have a severe effect on narrowband base pulses $g(t)$. In this case, not all the solutions of (3.20) are equivalent with respect to the optimization criterion of minimizing the variance of squared Euclidean distance. For example, some delay differences may result in cancelation of the received signal after going through the pre-filters. An example was given in [22] for the case of two transmit antennas with three different versions of the pre-filter coefficient matrix, showing system performances at varying delay differences between the two transmit branches.

The effect of noisy channel estimation, i.e., the channel coefficients $\{a_i\}$ are not known perfectly at the receiver, was studied by simulation in [22] for the case of 2-branch diversity, assuming no delay difference. A training pilot sequence $(0, 1, 0)$, chosen for simplicity, is inserted periodically into the transmitting sequence to estimate the channel. The repetition rate of the pilot sequence determines the fraction of the transmit energy devoted to channel measurement. In slow fading, this scheme does not affect the bit rate whereas it may reduce the bit rate significantly for fast fading. Simulation for two different repetition rates show that the relative gain obtained by the diversity scheme is fairly insensitive to the quality of the channel estimation.

### 3.4 Phase shifting

Another diversity method is to introduce phase shifts at different transmit antennas, thereby creating fast fading to increase diversity. Different phase shifting transmit diversity techniques have been studied in [25] and [24], where the first technique incorporates phase shift with a spreading sequence, and the second combines phase shift with interleave coding to obtain diversity gain. The first technique [25] however increases the transmission bandwidth by a factor of at least the number of transmit antennas due to the effect of the spreading sequence, whereas the second technique [24] introduces negligible bandwidth expansion, but the performance is evaluated in terms of coding rather than gain by transmit antenna diversity. The effect of phase shift alone on wireless system performance appears to have not been studied in the literature. In
the following sections, we will describe these two phase shifting techniques combined with a spreading sequence and interleave coding.

### 3.4.1 Incorporating phase shift with a spread sequence

This technique first spreads the transmit symbol into a faster bit rate using a spreading sequence, then introduces different and periodical phase shifts to the signal transmitted from different antennas [25], thus transforming the channel from flat fading into frequency selective fading, where a RAKE receiver [4] can be employed to detect the signal components from resolvable paths.

The signal model is as follows. Assuming binary signaling, the information symbol sequence \( \{I_n\} \) \((I_n = \pm 1)\) with symbol interval \( T \) is multiplied by a binary spreading sequence \( \{c_n\} \) with bit interval \( T_c \ll T \), which is also known as the chip sampling rate. Each data symbol then contains \( N = T/T_c \) chips, and a symbol interval is denoted as \( [0, N - 1] \), representing \( N \) chip intervals. Define \( p_N[n] = \begin{cases} 1 & n \in [0, N - 1] \\ 0 & \text{otherwise} \end{cases} \), then the information signal can be written as

\[
s[n] = \sum_{k=-\infty}^{\infty} I_k \cdot p_N[n - kN]
\]

whereas the spreading sequence \( c[n] \) can be written as

\[
c[n] = \sum_{k=-\infty}^{\infty} c_k \delta[n - k] .
\]

Then the signal after spreading becomes

\[
x[n] = s[n] \cdot c[n] .
\]

Assume that all antennas transmit with equal power. To create diversity using multiple transmit antennas, a symbol interval of \( N \) chips is divided into a number of sub-intervals equal to the number of transmit antennas \( M \), during each of which, a periodical time-varying phase offset \( \theta_i[n] \) is applied to the signal at antennas \( i \). Let \( \nu = N/M \), and assume that \( \nu \geq 1 \), i.e. the spreading gain is at least the same as the number of antennas. Denote \( a[n] \) as the time-varying composite fade coefficient seen at the receiver due to all \( M \) transmit antennas, then it can be expressed as

\[
a[n] = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i e^{j\theta_i[n]} \quad (3.23)
\]
where $1/\sqrt{M}$ is a powernormalizing factor to ensure the same total transmit power irrespective of the number of antennas. The time-varying phase offset $\{\theta_i[n]\}$ in the $k^{th}$ symbol interval $[kN, kN + N - 1]$ is chosen as

$$\theta_i[n] = \frac{2\pi(i - 1)m}{M}$$

with $m = 1, 2, \ldots, M$ satisfying

$$kN + (m - 1)\nu \leq n \leq kN + m\nu - 1.$$  

In other words, during a symbol interval $T$ of length $N$ chips, the signal sent from each antenna assumes $M$ phase values in $\{2\pi\frac{m}{M} \mid m = 0, \ldots, M - 1\}$ in the $M$ sub-intervals $[kN + m\nu, kN + (m + 1)\nu]$, i.e.,

$$R_{aa} = E[a[kN + m_1\nu] \cdot a[kN + m_2\nu]] = \sigma_a^2 \cdot \delta_{m_1, m_2}, \quad m_1, m_2 \in [0, M - 1].$$

The proof of this is straightforward, since

$$R_{aa} = \frac{1}{M} \cdot E\left[\left(\sum_{i=1}^{M} a_i \cdot e^{j2\pi(i-1)m_1/M}\right)\left(\sum_{i=1}^{M} a_i^* \cdot e^{-j2\pi(i-1)m_2/M}\right)\right] = \sigma_a^2 \cdot \sum_{i=1}^{M} \exp\left(j2\pi(m_1 - m_2)(i - 1)\right)$$

$$= \begin{cases} \frac{\sigma_a^2}{M} \cdot \frac{1 - \exp\left(j2\pi(m_1 - m_2)/M\right)}{1 - \exp\left(j2\pi(m_1 - m_2)/M\right)} = 0 & m_1 \neq m_2 \\ \sigma_a^2 & m_1 = m_2 \end{cases}.$$  

Similarly, it can be shown that the composite spreading sequence $c[n]e^{j\theta_i[n]}$ at each antenna is an orthogonal sequence during a symbol interval $[kN, kN + N - 1]$, i.e.,

$$\sum_{n=kN}^{n=kN+N-1} (c[n]e^{j\theta_i[n]}) \cdot (c[n]e^{-j\theta_k[n]}) = \sum_{m=1}^{M} \exp\left(j2\pi(i - 1)m/M\right) \cdot \exp\left(-j2\pi(k - 1)m/M\right) = \delta_{ik}.$$  

Thus this technique is equivalent to sending $M$ orthogonal signals from different transmit antennas, which is the $M$-ary orthogonal signaling described in the previous section. Therefore the signal bandwidth is expanded by a factor of $N$, which is the spreading factor, otherwise the transmission rate is reduced by a factor of $N$. Again, an $M$ matched filter bank can be employed at the receiver to process the received signal.

Two other receiver techniques are proposed in [25] for two modulation schemes BPSK (binary phase shift keying) with coherence detection and DPSK (differential
binary phase shift keying), where the receive signal is detected and processed independently in each of the sub-intervals \([kN + (m - 1)\nu, kN + m\nu - 1]\) and then summed up over all \(M\) sub-intervals in each symbol period. With BPSK modulation using coherence detection, fading coefficient estimation is required at the receiver, whereas in the case of DPSK, the signal is differentially detected and fading coefficient estimation is not required. We will describe these receivers briefly below.

Using the composite fade coefficient (3.23), the receive signal can be written as

\[
y[n] = a[n] \cdot x[n] + w[n]
\]

where \(w[n]\) is the additive Gaussian noise. Also denote \(a_m^k\) as the composite fade coefficient seen at the receiver during the \(m\)th sub-interval of the symbol interval \([kN, kN + N - 1]\), i.e.

\[
a_m^k = a[kN + m\nu] = \sum_{i=1}^{M} a_i \cdot e^{j2\pi(i-1)m / M}
\]

**BPSK system with coherence detection**

In this scheme, it is assumed that the receiver has perfect knowledge of the values of these complex coefficients \(a_m^k\). At the receiver, each receive symbol is divided into \(M\) sub-intervals, then the signal in each sub-interval is despread by multiplying with the same spreading sequence \(c[n]\). The despread signal of the \(m\)th sub-interval during symbol interval \([kN, kN + N - 1]\) is therefore

\[
y_m^k = \sum_{n=kN + m\nu}^{kN + (m+1)\nu-1} y[n]c[n] = \nu a_m^k I_k + w_m^k
\]

where \(w_m^k\) denotes the noise component after processing the \(m\)th sub-interval of the \(k\)th symbol. These sub-interval signals \(y_m^k\) are weighted independently by the conjugate transpose of \(a_m^k\), and then summed up to produce the decision output for the \(k\)th symbol as

\[
z^k = \sum_{m=1}^{M} y_m^k \cdot a_m^k = I_k \nu \sum_{m=1}^{M} |a_m^k|^2 + \sum_{m=1}^{M} w_m^k .
\]

The added noise component after processing is a linear combination of the original noise components and therefore is complex white Gaussian noise. It can be shown that the bit error probability, averaged over fading statistics, is given by the same expression as (3.10), which gives rise to \(M\) order diversity.
This technique is also applicable when there are multiple resolvable paths, i.e., the spreading transforms the channel from flat frequency into frequency selective fading. The same receive processes can be used, but with multiple processing branches for the multiple resolvable paths, which resembles the RAKE receiver. More diversity can be gained through these multipaths as the result of frequency diversity. Similar error expressions are obtained with \( ML \) order of diversity where \( L \) is the number of resolvable multipaths.

**Binary DPSK receiver**

This scheme employs differential modulation, thus the demodulation of DPSK does not require an estimate of the carrier phase, which may be attractive in channels whose characteristics vary very fast compared to the symbol rate. The information symbol sequence is first differentially encoded. This differential symbol sequence then goes through spreading and phase shifting in the same way as described above. Similarly, the receive signal in each symbol interval is divided up into \( M \) sub-intervals and despread by multiplying with the spread sequence to produce \( y_m^k \). Now instead of multiplying these \( y_m^k \) with the conjugate of the composite fade coefficient, the decision output is detected differentially as

\[
z^k = \sum_{m=1}^{M} y_m^k \cdot (y_m^{k-1})^*
\]

where it is assumed that the channel does not vary very much in successive symbol intervals, i.e., \( a_m^k = a_m^{k-1} \). It is found that the average bit error rate is given by [25]

\[
\bar{P}_b = \frac{1}{2^{2M-1}(M-1)!(1+\rho)^M} \sum_{k=0}^{M-1} \sum_{i=0}^{M-k-1} \binom{2M-1}{i} \frac{(M-1+k)!}{k!} \left( \frac{\rho}{1+\rho} \right)^k.
\]

Thus a diversity of order \( M \) is obtained. Again, when the channel is frequency selective with \( L \) resolvable paths, multi-branch processing at the receiver can be employed, which gives rise to diversity of order \( ML \).

**3.4.2 Combining phase sweeping transmit diversity and channel coding**

In multipath fading channels, channel coding is a well-known technique for combating the distortion effect produced by multipath. Since the wireless channel usually produce errors in bursts, when channel coding is applied, the received signal envelopes associated
with each bit among a codeword must be subjected to independent fading to reduce BER. One way to achieve this is to employ bit interleaving, especially in slow fading channels, in order to randomize the bursty errors. However, excessively long interleaving is not desirable since it introduces delays. Another BER reducing technique is to employ frequency hopping in which the transmit signal is hopped cyclically or randomly among a number of frequencies, each separated by at least the coherence bandwidth of the channel. Multiple transmit antennas can be used to reduce the bit interleaving length or the number of hop frequencies by a phase sweeping transmit diversity scheme as proposed in [24]. The two transmit antenna case was studied, where the carrier for one antenna is phase modulated independently of the original signal modulation by a time-varying sweeping function (Figure 3.5). Two different phase sweeping functions are considered, and the system performance is measured in terms of the word error rate (WER) of an \((n, k)\) block error correction code. The effect of the fading rate on the WER of system with and without phase sweeping transmit diversity and on the required depth of interleaving code for a given WER was studied. Comparison with receive diversity was also provided [24].

**Performance criterion**

The WER for an \((n, k)\) block error correction code with \(t\)-bit error correction capability, is given by

\[
p_W(\gamma) = \sum_{i=t+1}^{n} \binom{n}{i} p_b^i (1 - p_b)^{n-i}
\]  

(3.24)

where \(p_b\) is the BER and \(\gamma\) is the SNR. For binary DPSK, \(p_b = \frac{1}{2} e^{-\gamma}\), it is found that for very slow Rayleigh fading where the signal envelope is almost constant over several codewords, the average performance (averaged over the distribution of \(\gamma\)) cannot be significantly improved by increasing the error correction capability \(t\), even if bit interleaving is employed. On the contrary, with very fast fading, the signal level does not fade over the entire codeword, and random bit errors are produced. Increasing the error correction capability \(t\) in fast fading therefore significantly improves the WER performance. In fact the error correction \(t\) in fast fading is equivalent to receive diversity with \(t\) antennas using selection combining [24].

Bit interleaving or frequency hopping, which make use of the decorrelation property of fading in time and frequency domains respectively, can be employed to create fast fading and therefore improve the WER. However, these schemes are effective only
when the fading is not too slow. Since the two schemes have similar effects, we will concentrate on bit interleaving here. If block bit interleaving of \( m \)-bit depth is used, the equivalent fading rate (or the maximum Doppler spread) is increased by \( m \) times, while a decoding time delay of \( mnT \) is incurred, where \( T \) is the bit duration. Therefore, where long delays are not allowed, performance cannot be improved significantly by the use of bit interleaving only. This can be solved by employing multiple transmit antennas to produce forced fading. An existing scheme is antenna switching, where each of the transmit antennas are switched periodically so that only one antenna is active at any one time, but this scheme produces periodic abrupt phase changes in the receive signal, thereby causing periodic errors. In order to avoid this problem, a phase sweeping scheme is introduced in [24] where the carrier phase of each transmit antenna is phase modulated independently by a continuous time-varying sweeping function. Analysis is given for case of two transmit antennas.

![Figure 3.5: Transmit diversity with two branch phase sweeping technique.](image)

Using binary DPSK for simplicity, the receive signal is the sum of two independent fading signals which are transmitted from two transmit antennas

\[
\begin{align*}
y_1(t) &= \alpha_1(t)\cos[\omega_c t + \theta_1(t) + \Phi_x(t) + \eta(t)] \\
y_2(t) &= \alpha_2(t)\cos[\omega_c t + \theta_2(t) + \Phi_x(t)]
\end{align*}
\]

(3.25)

where \( \alpha_i(t) \) and \( \theta_i(t) \) (\( i = 1, 2 \)) are the random envelope and the random phase due to fading, \( \Phi_x(t) \) is the information bearing phase and \( \eta(t) \) is the hopping phase. The instantaneous receive SNR of each signal is \( \gamma_i(t) = \alpha_i(t)^2 / \sigma_N^2 \) where \( \sigma_N^2 \) is the noise power. Due to independent fading between different antennas, \( \gamma_1 \) and \( \gamma_2 \) are independent. When \( \eta(t) \) varies linearly and is uniformly distributed over \( [0, 2\pi] \), the BER for binary DPSK is

\[
p_b = \frac{1}{2} e^{-\gamma_1 - \gamma_2} I_0 \left( 2\sqrt{\gamma_1 \gamma_2} \right)
\]

74
and the average WER is found [24] to be

\[
P_W = \int_0^\infty \int_0^\infty p_W(\gamma_1, \gamma_2) \frac{1}{(\tilde{\gamma}/2)^2} \cdot \exp \left[ -\frac{\gamma_1}{(\tilde{\gamma}/2)} - \frac{\gamma_2}{(\tilde{\gamma}/2)} \right] \cdot d\gamma_1 d\gamma_2
\]

where \(\tilde{\gamma}\) is the average total SNR and \(p_W(\gamma_1, \gamma_2)\) is given in (3.24) with the above BER \(p_b\). Numerical results in [24] showed that two antenna phase sweep diversity outperforms the single transmitter case, and with perfect bit interleaving, i.e. \(m \to \infty\), then phase sweeping even outperforms selection combining receive diversity at WER of less than \(10^{-2}\) and maximum ratio combining at WER less than \(10^{-3}\).

Effect of phase sweeping diversity on interleaving depth

Two sweeping functions were studied: linear \(\eta(t) = 2\pi f_H t\) and sinusoidal \(\eta(t) = \Delta \theta \sin(2\pi f_H t)\), where \(f_H\) is the sweeping frequency which determines the rate of forced fading. This sweeping frequency \(f_H\) can be considered equivalent to the maximum Doppler spread, which determines the fading rate and also the frequency expansion of the scheme. Assuming block bit interleaving with depth \(m\), the phase \(\eta_i\), \((i = 1, 2, \ldots, n)\) of the \(i\)th bit of a \(n\)-bit codeword is

\[
\eta_i = \eta(m \cdot T \cdot i) = \begin{cases} 2\pi m f_H \cdot T \cdot i & \text{for } \eta(t) = 2\pi f_H t \\ \Delta \theta \sin(2\pi m f_H \cdot T \cdot i) & \text{for } \eta(t) = \Delta \theta \sin(2\pi f_H t) \end{cases}
\]

To obtain a uniform distribution of the sweeping phase over \([0, 2\pi)\) in a single codeword, in the case of linear sweeping function, it is required that

\[
mn f_H T \geq 1. \quad (3.26)
\]

For the sinusoidal phase sweeping function, clearly to have the sweeping phase cover a \(2\pi\) circle, we must have \(\Delta \theta \geq \pi\). Let \(\Delta \theta = \pi\), if we choose \(mn f_H T = 1\) as in the linear function case, the phase \(\eta_i\) will vary two times as fast as in the linear function case, as it will vary \(0 \to \pi \to 0 \to -\pi \to 0\) within a single codeword whereas the linear function case \(\eta_i\) varies \(0 \to \pi \to 2\pi\) within a single codeword. Therefore the instantaneous SNR of the combined signal in the sinusoidal case will vary twice as fast as in the linear case. Since the variation rate of the SNR corresponds to the maximum Doppler frequency, which in turns affects the interleaving depth, this suggests that the sweeping frequency \(f_H\) for the sinusoidal case should be chosen as half of that in the linear case. As a consequence, for the sinusoidal function case, the following conditions are required

\[
\Delta \theta > \pi, \quad mn f_H T > \frac{1}{2}. \quad (3.27)
\]
The effect of phase sweeping diversity on interleaving depth can be illustrated by the following example. Consider a $BCH(23,12)$ code with 3 bit error correction capability at $32K$ pbs transmission rate at carrier frequency $1.45$ GHz. Assuming that the Doppler spread is $f_D = 1$ Hz without transmit diversity, then the required interleaving depth for a certain error rate is $m \approx 1390$ bits, resulting in a very large interleaving size (approximately 32000 bits). Using phase sweeping techniques with sweeping frequency $f_H = 140$ Hz in the linear case, and $f_H = 70$ Hz in the sinusoidal case, the required interleaving depth is reduced to $m = 10$ for the same error rate. Thus, phase sweeping significantly reduces the required interleaving depth while introducing a negligible bandwidth expansion (of $70$ Hz or $140$ Hz for $32K$ bps transmission rate in this example). Extensive experimental results [24] also confirm these analysis results.

### 3.5 Antenna linear precoding

Results in information theory suggest that the capacity of a channel, in principle, can be approached arbitrarily closely through the use of a suitable coding scheme [12]. However, the computational complexity of such a scheme prohibits its practical application. This section will develop a framework for a class of efficient linear signal processing algorithms to exploit transmit diversity in a flat frequency and slow fading channel, without incurring bandwidth expansion. The scheme, in a sense, is a generalization of the previous diversity schemes: time shifting, modulation diversity, phase shifting. Emphasis is placed on linearly prefiltering the signal before transmission from each antenna, and linear equalization at the receiver, which has been shown to asymptotically transform the channel from flat and slow Rayleigh fading into a non-fading additive white marginal Gaussian noise channel as the number of transmit antennas increases to infinity [18, 17].

The setup is that the transmit signal is processed by a linear time-invariant FIR prefilter of length $M$ at each of the $M$ transmit antennas before transmitting, which is referred to as “linear antenna precoding”. At the receiver, a linear equalizer is used to detect the transmitted symbols. It has been shown that when the prefilter taps make up a unitary matrix, the Rayleigh fading channel is asymptotically transformed into a non-fading AWGN channel, based on which the optimum equalizer is derived [18]. Analysis in this reference shows some performance improvement with the number of transmit antennas, compared to single antenna case. However, the optimum linear
equalizer in this case has infinite length and an unrealizable response. A solution has been proposed by having periodically time-varying prefilters instead of time-invariant ones. This structure achieves the same performance as the time-invariant prefilters, but introduces a finite delay equal to twice the number of antennas [18]. In the followings, we will provide the analyses into these two diversity techniques.

3.5.1 Linear time-invariant antenna precoding

In this section we will study the linear time-invariant (LTI) prefilter case, where antenna $i$ is preceded by a LTI filter $h_i[n]$ (Figure 3.6). These filters are generally complex-valued.

![Fading Channel](image)

**Figure 3.6: Linear time-invariant antenna precoding.**

**Antenna precoding construction**

Let $x[n]$ be the input symbol sequence and $u_i[n]$ be the output of the prefilter $i$, then

$$u_i[n] = \frac{1}{\sqrt{M}} \sum_{k=-\infty}^{+\infty} h_i[k] x[n - k] = \frac{1}{\sqrt{M}} h_i[n] * x[n]$$

where $1/\sqrt{M}$ is the normalizing factor. The prefilter tap set $\{h_i[n]\}_{i=1}^{M}$ is called the signature of the antenna. The associated Fourier transform of each signature will be

$$H_i^a(\omega) = \sum_{n} h_i[n] e^{-j\omega n} \quad (3.28)$$

These frequency responses $H_i^a(\omega)$ are $2\pi$ periodic. To ensure the total average transmitted power is independent of number of antennas $M$, the signature must satisfy

$$\frac{1}{M} \sum_{i=1}^{M} |H_i^a(\omega)|^2 = 1 \quad \forall \omega \quad (3.29)$$

---

1 The superscript $^a$ denotes frequency response of a sampled (discrete) signal.
since the power of the signal transmitted from the $i$th antenna is

$$ \text{var} \ u_i[n] = \frac{1}{2\pi M} \int_{-\pi}^{\pi} |H_i^s(\omega)|^2 S_{xx}^s(\omega) \, d\omega $$

where $S_{xx}^s(\omega)$ is the power spectrum of the information symbol stream $x[n]$. Thus the total transmit power is

$$ \sum_{i=1}^{M} \text{var} \ u_i[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \frac{1}{M} \sum_{i=1}^{M} |H_i^s(\omega)|^2 \right] S_{xx}^s(\omega) \, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}^s(\omega) \, d\omega = \text{var} \ x[n]. $$

To achieve a certain desirable asymptotic characteristic of the diversity scheme with linear equalization at the receiver, the following condition is imposed on the prefilters,

$$ \rho_M(\omega, n) = \frac{1}{M} \sum_{i=1}^{M} H_i^s(\omega) H_i^s(\nu)^s \to 0 \ \text{as} \ M \to \infty, \ \omega \neq \nu \quad (3.30) $$

which we will call the frequency orthogonality condition. This orthogonality condition will ensure that the variances of the receive signal and noise will asymptotically converge to zero as the number of transmit antennas increases. This in turn has the effect of asymptotically transforming the channel from a Rayleigh fading channel into a non-fading marginal white additive Gaussian noise channel.

Ideal bandpass signatures are infinite length and unrealizable. Thus we consider a family of practical finite length signatures of length $M$. Construct the matrix representation of these signatures as

$$ \mathbf{H} = \begin{bmatrix} h_1[0] & h_1[1] & \ldots & h_1[M - 1] \\ h_2[0] & h_2[1] & \ldots & h_2[M - 1] \\ \vdots & \vdots & \ddots & \vdots \\ h_M[0] & h_M[1] & \ldots & h_M[M - 1] \end{bmatrix}. \quad (3.31) $$

It can be shown that to satisfy both conditions (3.29) and (3.30), it is sufficient that $\mathbf{H}$ is a unitary matrix [18]. From this, we can conveniently choose an arbitrary unitary matrix $\mathbf{H}$ to construct the signature set. For example, if we choose $\mathbf{H} = \mathbf{I}_M$ where $\mathbf{I}_M$ is the $M \times M$ identity matrix so that $h_i[n] = \delta[n - m]$, we obtain the time shifting diversity scheme of Winters [21] (also Seshadri and Winters [20]) where delayed versions of the transmitted signal are sent from different antennas, as described in Section 3.2 of this thesis. Other choices include $\mathbf{H} = \mathbf{F}$, the discrete Fourier transform matrix

$$ [\mathbf{F}]_{k,l} = e^{-j2\pi kl/M} $$
or \( \mathbf{H} = \Xi \), the Hadamard matrix, which is defined recursively when \( M \) is a power of two (i.e., \( M = 2, 4, 8 \ldots \))

\[
\Xi_M = \frac{1}{\sqrt{2}} \begin{bmatrix} \Xi_{M/2} & \Xi_{M/2} \\ \Xi_{M/2} & -\Xi_{M/2} \end{bmatrix}
\]

with \( \Xi_1 = 1 \). The former corresponds to the scheme that divides the available bandwidth into equal fractions for each transmit antenna. When \( M = 2 \), these two schemes reduce to a common scheme which was also studied by Wittneben [22].

In fact, the idea of a unitary matrix here is the same as the equal power and orthogonal impulse response conditions on the FIR prefilters (3.20) in the modulation diversity scheme by Wittneben [23], as described in Section 3.3, and is therefore optimal with respect to minimizing the output error energy (i.e., the variance of the square Euclidean distance) for single symbol inputs. However, in this antenna precoding scheme [18], the unitary condition on \( \mathbf{H} \), or more generally, the orthogonality condition (3.30), is used as a pre-condition without theoretical base, in order to achieve the asymptotic channel transformation. The two conditions imposed on the prefilters, (3.29) and (3.30), are analogous to the equal power (3.18) and orthogonality (3.19) conditions by Wittneben. Therefore it can be said that the linear antenna precoding technique [18] is an extension of modulation diversity [22]. Here, linear equalization is used instead of the Viterbi algorithm based MLSE, thus the scheme has an advantage for practical implementation.

**Equivalent channel**

By introducing the FIR filters in front of each transmit antenna, the receive signal can be expressed in the form

\[
y[n] = a[n] * x[n] + w[n]
\]

where \( w[n] \) is the additive zero-mean special complex Gaussian noise with variance \( \sigma_N^2 \), and \( a[n] \) is the equivalent time-varying channel response created by antenna precoding

\[
a[n] = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i \cdot h_i[n] .
\]

Thus the FIR filters have effectively transformed the channel from a flat fading channel into a frequency selective fading channel (2.26) with frequency response

\[
A^a(\omega) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} a_i \cdot H_i^a(\omega) .
\] (3.32)
Since the fade coefficients \( a_i \) are special complex Gaussian random variables, at each frequency \( \omega \), \( A^s(\omega) \) is a zero-mean, special complex Gaussian random variable with variance \( \sigma_a^2 \) due to (3.29). From this observation, any receiver that exploits spectral diversity such as the traditional approaches to decoding in the presence of intersymbol interference can be applied. Examples include maximum likelihood sequence detection, decision feedback equalization or linear equalization. There are some favorable differences between this synthesized frequency selective channel and a naturally occurring frequency selective channel in that the precoding signatures \( h_i[n] \) chosen according to the two conditions (3.29) and (3.30) ensure that the resulting frequency diversity has strong asymptotic mixing characteristics as will be shown below. Also the channel identification problem in this case is much simpler than in naturally occurring frequency selective channels.

**Asymptotic characteristic and optimum linear equalizer**

Letting \( b[n] \) be the unit-sample response of an LTI equalizer used at the receiver, the signal after equalization can be written as

\[
\hat{x}[n] = \sum_k b[k] \cdot y[n - k] = b[n] \ast a[n] \ast x[n] + b[n] \ast w[n].
\]

Since the receiver is assumed to know the channel coefficients \( a_i \) and the precoding signature \( h_i[n] \), the equalizer \( b[n] \) will depend on the effective channel response \( a[n] \).

Suppose that the equalizer’s frequency response \( B^s(\omega) \) is a function of the channel frequency response \( A^s(\omega) \) in (3.32), which is bounded within any finite region of the complex plane and which grows more slowly than a quadratic exponential \([18]\). Let \( C^s(\omega) \) be the effective frequency response after equalization, i.e. \( C^s(\omega) = A^s(\omega)B^s(\omega) \), then \( C^s(\omega) \) is also a function of \( A^s(\omega) \). Since \( A^s(\omega) \) is a stationary complex Gaussian process, both \( B^s(\omega) \) and \( C^s(\omega) \) will then have finite mean and variance which depends only on \( \sigma_a^2 \) but not \( \omega \). Let

\[
\begin{align*}
\mu_b &= E[B^s(\omega)] \\
\mu_c &= E[C^s(\omega)] \\
\sigma_b^2 &= \text{var} \, B^s(\omega) \\
\sigma_c^2 &= \text{var} \, C^s(\omega).
\end{align*}
\]

Then it can be shown \([18]\) that if the information stream \( x[n] \) is a sequence of zero-mean uncorrelated symbols, each with energy \( \mathcal{E}_a \), then the prefilters satisfying conditions (3.29) and (3.30) together with the linear equalizer at the receiver as defined above,
will exhibit the following asymptotic characteristic

\[ \hat{x}[n] \overset{m.s.}{\to} \mu_v x[n] + v[n] \]  

(3.34)

where m.s. denotes convergence in the mean squared sense, and \( v[n] \) is a complex-valued, marginally Gaussian, zero mean white noise sequence, uncorrelated with the input symbol stream \( x[n] \) and having variance

\[ \text{var} \ v[n] = \mathcal{E}_c \sigma_c^2 + \sigma_N^2 \left( \sigma_b^2 + |\mu_b|^2 \right) . \]  

(3.35)

Relation (3.34) asserts that with this scheme of transmit antenna diversity, the channel seen by the coded symbol stream \( \hat{x}[n] \) is transformed from a fading channel into a non-fading marginally Gaussian white additive noise channel. The noise variance (3.35) shows that there are two components in this equivalent noise: the first one is due to the original receiver noise \( \sigma_N^2 (\sigma_b^2 + |\mu_b|^2) \), whereas the second is due to intersymbol interference (ISI) generated by the introduction of prefilters at the transmit antennas, and hence has a variance that scales with the symbol energy \( \mathcal{E}_c \sigma_c^2 \). By transforming the fading channel into an AWGN channel, this diversity scheme reduces the variance in the system performance. For example, in the case of multiple receivers placed at an approximately equal distance from the transmitter, the variation of the receive signal from receiver to receiver is reduced as the number of transmit antennas increases. It also suggests that some powerful and computationally efficient receiver structure can be used, instead of the complex Viterbi-algorithm-based maximum likelihood sequence detection as used in [21]. A symbol-by-symbol detection is feasible in this case.

From (3.34) and (3.35), the asymptotic SNR after equalization is

\[ \gamma_0 = \frac{\mu_v^2 \mathcal{E}_s}{\mathcal{E}_c \sigma_c^2 + \sigma_N^2 (\sigma_b^2 + |\mu_b|^2)} = \frac{\mathbb{E}[AB]^2}{\mathbb{E}[|AB|^2] \left( \mathbb{E}[|AB|^2] - \mathbb{E}[A]^2 \right) + \xi_0 \mathbb{E}[|B|^2]} \]

where \( \xi_0 = \sigma_N^2 / \mathcal{E}_s \) is the inverse of transmit SNR. In order to maximize this output SNR, applying Schwarz inequality

\[ \mathbb{E}[|AB|^2] \leq \mathbb{E} \left[ \frac{|A|^2}{|A|^2 + \xi_0} \right] \cdot E \left[ |B|^2 (|A|^2 + \xi_0) \right] \]

with equality if and only if

\[ B^s(\omega) \propto \frac{A^s(\omega)^*}{|A^s(\omega)|^2 + \xi_0} \]  

(3.36)

the optimum linear equalizer is therefore given by the above expression. The constant of proportionality is arbitrary. This equalizer not only maximizes the equivalent channel
SNR but also, when suitably normalized, make \( \hat{x}[n] \) a minimum-mean-squared error estimate of \( x[n] \), thus adaptive equalizers based on LMS or recursive least-squares algorithms can be used in practice. The optimum linear equalizer also takes the form of a matched filter with a down factorization, which can be viewed as an additional compensation stage that takes into account the special characteristics of the equivalent noise.

**Performance evaluation**

The SNR after equalization with the optimum equalizer (3.36) is given by

\[
\gamma_0 = \left( \frac{1}{\eta_0} - 1 \right)^{-1}
\]

where \( \eta_0 = E[|A|^2/(|A|^2 + \xi_0)] \).

Performance in terms of the bit error rate for QPSK (quadrature phase shift keying) shows improvement with this linear precoding antenna diversity compared to no diversity system. The BER of a non-fading AWGN QPSK system is given by [1]

\[
P_b = Q(\sqrt{\gamma})
\]

where \( \gamma \) is the receive SNR. After some arithmetic manipulation involving asymptotic expansions of the exponential integral and the \( Q \) function, it can be shown that the BER of antenna precoding diversity with an infinite number of transmit antennas at high SNR \( \rho \) (2.31) is given by [19]

\[
P_{b\infty} \approx \sqrt{\frac{\log_2 \rho}{\rho}} \exp\left( -\frac{\rho/2}{\log_2 \rho} \right)
\]

whereas without transmit diversity, the average BER is

\[
P_{b0} = \frac{1}{2} \left( 1 - \frac{1}{\sqrt{1 + 2/\rho}} \right)
\]

and the BER for infinite receive diversity with optimum combining, based on [1], is

\[
P_{bR} = Q(\sqrt{\rho}) \approx \frac{1}{\sqrt{\rho}} \exp\left( -\frac{\rho}{2} \right).
\]

Thus the average bit error rate with infinite diversity in both transmit and receive diversity cases falls off at a negative exponential rate, faster than BER without diversity. The difference between antenna precoding transmit diversity (3.37) and receive diversity with optimum combining (3.38) amounts to a log-type penalty in the transmit case.
However, the optimum linear receiver in the form (3.36) generally exhibits infinite length and unrealizable unit-sample impulse response. While a finite-length approximation can be used, this still creates excessive delay. A modified scheme using periodically time-varying linear antenna precoding is developed in [18] to overcome this problem.

### 3.5.2 Linear periodically time-varying antenna precoding

**Antenna precoding construction**

The system is depicted in Figure 3.7. In this case, the coded information stream $x[n]$ is first processed by a common linear periodically time-varying (LPTV) prefilters of length $M$ that varies with period $K \geq 2$, and has kernel denoted by $g[n; k]$. The period $K$ in fact plays a relatively minor role in the development, and fixing $K = 2$ suffices in practice [18]. The output $u[n]$ of this prefilter is then processed at each $i$th antenna by a different $M$-periodic sequence, generally complex-valued $\hat{h}_i[n]$ to produce

$$u_i[n] = \hat{h}_i[n] \cdot u[n].$$

Denote a single period of this modulating sequence $\hat{h}_i[n]$ as $h_i[n]$, i.e.

$$h_i[n] = \begin{cases} 
\hat{h}_i[n] & 1 \leq n \leq M \\
0 & otherwise 
\end{cases}.$$

These $h_i[n]$ are again called the signatures of the associated antennas.

For this system, the received signal can be written as

$$y[n] = \bar{a}[n] \cdot u[n] + w[n]$$

(3.39)
where \( \hat{a}[n] \) is the \( M \)-periodic equivalent channel response. Denote a period of this equivalent channel response as \( a[n] \), then
\[
a[n] = \sum_{i=1}^{M} a_i \cdot h_i[n].
\]
(3.40)

From (3.39), the signature modulation has effectively transformed the channel from non-selective into a time-selective one (2.27). The maximum time benefit can be obtained when the fading is independent among time samples within a period. From (3.40) it can be seen that the coefficients \( a[1], a[2], \ldots, a[M] \) are zero-mean and jointly Gaussian. The correlation between an arbitrary pair of these coefficients is proportional to the inner product between the corresponding columns of the matrix \( \mathbf{H} \) in (3.31), i.e.
\[
E[a[n] \cdot a^*[l]] = \sigma_n^2 \sum_{i=1}^{M} h_i[n] \cdot h_i^*[l].
\]

Hence the coefficients \( a[1], a[2], \ldots, a[M] \) will be independent if the columns of \( \mathbf{H} \) are orthogonal. Provided that the prefilter \( g[n; k] \) is an orthonormal transformation as imposed below, the constraint of total transmit power implies that the columns of \( \mathbf{H} \) must have unit norm since
\[
\text{var } u[n] = \sum_{i=1}^{M} \text{var } u_i[n] = \text{var } x[n] \sum_{i=1}^{M} |h_i[n]|^2 = \text{var } x[n].
\]

Thus, it is sufficient and necessary that \( \mathbf{H} \) be a unitary matrix in order to achieve maximum diversity benefit in this case.

**Prefilter design**

We will now consider the prefilter design for \( g[n; k] \). In fact this prefilter can be avoided if coding is used prior to the antenna precoding stage at the transmitter. However, such coding is usually computationally prohibitive in practice, particularly when the number of antennas is large. The best diversity benefit is often achieved by combining, or even replacing coding with a suitably designed prefilter. The role of this time-varying prefilter in combination with the signature set is to create time diversity from spatial diversity.

Let \( g_i[n], \ (i = 0, \ldots, K - 1) \) be the \( M \)-periodic sequences that makes up the \( K \)-period linear periodical time-varying filter (LPTV) \( g[n; k] \). The kernel \( g[n; k] \) represents the response of the filter at time \( n \) to an impulse applied at time \( n - k \), and can be written as
\[
g[n; k] = g[n; lK + i] = g_i[n - lK]
\]

84
where \(0 \leq i \leq K - 1\) satisfies \(k = lK + i\). Then the canonical form of the input-output relationship of a \(K\)-period LPTV filter is given by

\[
u[n] = \sum_{l=-\infty}^{\infty} \sum_{i=0}^{K-1} x[iK + i] \cdot g_l[n - lK] .
\] (3.41)

Denote a vector of \(K\) unit-sample responses of the prefilter as \(g[n] = [g_0[n], g_1[n], \ldots, g_{K-1}[n]]^T\) with its Fourier transform vector \(G(\omega)\) defined as

\[
G(\omega) = \sum_{n=-\infty}^{\infty} g[n] e^{-j\omega n} = \begin{bmatrix}
G_0^s(\omega) & G_1^s(\omega) & \cdots & G_{K-1}^s(\omega)
\end{bmatrix}^T .
\] (3.42)

To ensure that the LPTV prefilter creates an orthonormal transformation such that

\[
\sum_k g[k - nM] g^*[k - mM] = \delta[n - m] I
\]
\[
\sum_k g^*[n - kM] g[m - kM] = \sum_{k,i} G_i^s[m - kM]^* G_i^s[m - kM] = \delta[n - m] \] (3.43)

the Fourier transform vector has to meet certain conditions. The total input power constraint also implies that the power spectral density of the input and output signal of the prefilter must be the same, thus \(|U(\omega)|^2 = |X(\omega)|^2\) where the discrete time Fourier transform is defined as in (3.28). By taking Fourier transforms of both sides of (3.41), the following expression can be derived

\[
\left| \sum_{i=0}^{K-1} G_i^s(\omega) e^{j\omega i} \right|^2 = 1 .
\] (3.44)

Let \(\Delta(\omega)\) be the Fourier transform of the delay chain of order \(K\), i.e.

\[
\Delta(\omega) = \begin{bmatrix} 1 & e^{-j\omega} & \cdots & e^{-j\omega(K-1)} \end{bmatrix}^T .
\]

If we express \(G(\omega)\) (3.42) in the form \(G(\omega) = Q(\omega)\Delta(\omega)\), then from (3.43) and (3.44), it is necessary and sufficient that the \(K \times K\) polyphase matrix \(Q(\omega)\) is paraunitary [17], i.e.

\[
Q^1(\omega) \cdot Q(\omega) = I_K .
\] (3.45)

The condition (3.45) allows us to choose the appropriate polyphase matrix \(Q(\omega)\), and thus construct the prefilter \(g[n; k]\). A convenient way of constructing a maximally spread prefilter, where the unit-sample responses \(g_i[n]\) are binary valued, is to use the recursive Hadamard matrix for \(Q\). An example of \(g_i[n]\) with period \(K = 2\) and length \(M = 8\) is given in [18].

85
Asymptotic characteristic and optimum linear equalizer

Let \( \tilde{b}[n] \) be a suitable linear equalizer used at the receiver to produce

\[
\hat{u}[n] = \tilde{b}[n] \cdot y[n].
\]

This equalizer output is then processed by the inverse of the LPTV prefilter \( g[n; k] \) (see figure 3.7), resulting in symbol estimates

\[
\hat{x}[n] = \sum_k g[-n; -k] \hat{u}[n - k].
\]

Suppose the equalizer \( \tilde{b}[n] \) is a function of the effective channel response \( \hat{a}[n] \) \( (3.40) \) that satisfies the same condition as the equalizer function in the previous time-invariant case. Rewrite its output as

\[
\hat{u}[n] = \hat{c}[n] \cdot u[n] + \tilde{b}[n] \cdot w[n]
\]

where \( \hat{c}[n] = \hat{a}[n] \tilde{b}[n] \). When orthogonal signatures are used, \( \hat{a}[n] \) is a special complex Gaussian sequence with zero mean and variance \( \sigma_a^2 \). It follows that both \( \tilde{b}[n] \) and \( \hat{c}[n] \) have means and variances that are not only finite but are also functions of only \( \sigma_a^2 \) and not of \( n \). Accordingly, denote

\[
\mu_b = E[\tilde{b}[n]] \quad \mu_c = E[\hat{c}[n]]
\]
\[
\sigma_b^2 = \text{var} \tilde{b}[n] \quad \sigma_c^2 = \text{var} \hat{c}[n]
\]

then the same asymptotic results as \( (3.34) \) and \( (3.35) \) are obtained. Also by maximizing the asymptotic SNR after equalization, the optimum filter in this case is found to be

\[
\tilde{b}[n] \propto \frac{\hat{a}^*[n]}{\hat{a}[n]^2 + \xi_0}
\]

where \( \xi_0 = \sigma_v^2 / E_s \). Thus the equalizer in this case has an \( M \)-periodic impulse response sequence. The system delay is finite, since the equalizer \( (3.47) \) introduce no delay while the prefilter \( g[n; k] \) and post-filter \( g[-n; -k] \) each introduces a delay \( M \). The overall delay therefore is proportional to the number of transmit antennas.

Performance evaluation

Some empirical performance evaluation \([18]\), where the BER versus bit SNR was studied for different numbers of transmit antennas, from \( M = 1 \) (no diversity) to the asymptotic case \( M = \infty \), showed that the performance increases with the number of antennas...
as expected, though, there are diminishing returns beyond a moderate value of $M$. Also the hardware cost and the system delay constraints typically limit the practically useful value of $M$. Results also confirm the expectation that given the same number of antennas, receive diversity with optimum combining is always more effective than transmit diversity via linear processing.

In short, the linear antenna precoding scheme as analyzed in this section, although does not achieve the full capacity offered by transmit diversity due to constraints imposed by linear processing [18], does show a performance improvement in term of BER with the number of antennas. It also exhibits the attractive asymptotic characteristic of transforming a flat Rayleigh fading channel into a non-fading, marginally Gaussian, white additive noise channel. This suggests that computationally efficient receivers can be used instead of the Viterbi algorithm based MLSE as in [21].

### 3.5.3 Summary of linear antenna diversity techniques

Up to now, we have presented linear transmit diversity techniques which have in common the idea of imposing some unitary FIR filters before sending signal from each antenna. Various receive techniques have been studied, including linear equalizer (Wornell and Trott [18], Seshadri and Winters [20]), MMSE-DFE (Seshadri and Winters [20]) and MLSE based on Viterbi algorithm (Winters [21], Wittneben [22, 23]). All these techniques improve performance with increasing number of antennas, especially at low BER. The linear antenna precoding technique [18] serves as a generalization of the other linear techniques, although it only provides results on the optimum linear equalizer and BER for the asymptotic case where the number of transmit antennas is infinite. In the next chapter, using the same linear structure, we seek to use the mean square error optimum criterion (which is equivalent to minimizing the variance of squared Euclidean distance of the output sequences corresponding to single symbol inputs) to derive the optimum set of pre.filters $h_i[n]$ and the optimum linear equalizer that can be used for any finite number of transmit antennas. Performance evaluation is also extended to the Rician fading case.

Moving from linear processing to non-linear processing, the combination of transmit diversity, receive diversity and channel coding promises even more performance improvement. Although we do not consider coding in our work, this area deserves to be mentioned with regard to spatial diversity. The new notion of orthogonal coding is
studied in [27, 28] will be described briefly in the next section.

3.6 Space-time orthogonal coding

Recently, space-time coding has been studied [29, 30] which combines signal processing at a multiple antenna receiver with coding applied to a multiple antenna transmitter. Data is encoded using a space-time trellis code, and then is split into $M$ sub-streams which are simultaneously transmitted using $M$ transmit antennas. The received signal at each receive antenna is a linear superposition of the $M$ transmitted signals perturbed by noise. It is claimed that the code design in [29] provides significant gain over the linear transmit diversity techniques in [20, 23], and the time shifting scheme in [20] is a special case of this coding technique. The specific space-time trellis codes designed for 2-4 transmit antennas perform well in slow fading environments and come within 2-3dB of the information theoretic outage capacity computed in [11, 10]. This space-time trellis code also provides the best possible trade-off between constellation size, data rate, diversity advantage and trellis complexity. However, when the number of transmit antennas is fixed, the decoding complexity of the space-time codes, measured by the number of the trellis states in the decoder, increases exponentially with transmission rate.

To reduce the complexity of the decoder, a new paradigm for transmission over Rayleigh fading channels using multiple transmit and receive antennas is introduced by Tarokh, Jafarkhani and Calderbank in [27], and is given the name space-time block coding. The code is designed so that at the receiver, a simple linear detection scheme through de-coupling of the signals transmitted from different antennas can be used to achieve maximum likelihood decoding, instead of complex joint detection. Thus the space-time block code employs an orthogonal structure to obtain maximum diversity of order equal to the product of the number of transmit and receive antennas, and it can be interpreted as providing some less attenuated replica of the transmitted signal to the receiver, with a constraint of having a simple decode algorithm [27].

Diversity criterion

Consider a system with $M$ transmit antennas and $N$ receive antennas. Again slow and flat Rayleigh fading is assumed where the channel coefficients between transmit and receive antennas $\{a_{ij}\}$, $(i = 1 \ldots M; j = 1 \ldots N)$ are modeled as special complex
Gaussian random variables with zero mean and variance $\sigma_n^2$. The channel is assumed to be quasi-static so that the coefficients are constant over a frame of length $l$ symbols and vary from one frame to another. The received signal at antenna $j$ at a sample time $n$ is given by

$$y_n^j = \sum_{i=1}^{M} a_{i,j} x_n^i + \eta_n^j$$

where $x_n^i$ is the $n$th coded symbol sent from the $i$th transmit antenna, and $\eta_n^j$ are independent samples of a zero-mean special complex Gaussian random noise with variance $\sigma_n^2$. All transmit antennas are assumed to transmit with equal power $P/M$.

Assuming perfect channel state information available at the receiver, the receiver would compute the decision metric

$$\sum_{n=1}^{l} \sum_{j=1}^{N} |y_n^j - \sum_{i=1}^{M} a_{i,j} c_n^i|^2$$

over all codewords $c_1^1 c_2^1 \ldots c_M^1 c_1^2 c_2^2 \ldots c_M^2 \ldots c_1^l c_2^l \ldots c_M^l$ and decides in favor of the codeword that minimizes this sum. Denote $e$ as an error signal

$$e = c_1^1 c_2^1 \ldots c_M^1 c_2^2 c_2^2 c_M^2 \ldots c_1^l c_2^l c_M^l$$

given that the transmitted signal was

$$c = c_1^1 c_2^1 \ldots c_M^1 c_2^2 c_2^2 c_M^2 \ldots c_1^l c_2^l c_M^l.$$  

Analysis of the probability that the receiver decides erroneously in favor of signal $e$ instead of the correct one $c$ leads to the diversity criterion that, in order to achieve the maximum diversity of order $M \times N$, the matrix $B(c, e)$, where $B_{i,j} = e_j^i - c_j^i$, has to be full rank for any pair of distinct codewords $c$ and $e$ [29]. If $B(c, e)$ has minimum rank $r$ over the set of pairs of distinct codewords, a diversity of order $r \times N$ is achieved.

**Impact of orthogonal design on diversity and decoding strategy**

A real orthogonal design of size $n$ is an $n \times n$ orthogonal matrix with entries in the indeterminates $\pm s_1, \pm s_2, \ldots, \pm s_n$ such that each row of the matrix is a unique permutation of these indeterminates. Now apply orthogonal design to construct space-time block codes that achieve full spatial diversity $M \times N$. Assume that the transmission at the baseband employs a real signal constellation $A$ with $2^b$ elements, which implies that the maximum transmission rate is $b \text{ bps} / Hz$ [29]. At time slot 1, $Mb$ bits arrive
at the encoder and select constellation signals \( s_1, s_2, \ldots, s_M \), which becomes the first row of the chosen coding matrix

\[
C \triangleq \mathcal{O}(s_1, s_2, \ldots, s_M)
\]

with entries \( \pm s_1, \pm s_2, \ldots, \pm s_M \). At each time slot \( n = 1, 2, \ldots M \), the entries \( C_{ni} \), \( i = 1, 2, \ldots M \) are sent simultaneously from transmit antennas \( 1, 2, \ldots M \) respectively, thus the rate achieved is the maximum \( b \) bps/Hz. The orthogonal codes satisfy

the rank condition required to achieve the maximum spatial diversity \( M \times N \), since the orthogonal matrix \( \mathcal{O}(\bar{s}_1 - s_1, \ldots, \bar{s}_M - s_M) = \mathcal{O}(\bar{s}_1, \ldots, \bar{s}_M) - \mathcal{O}(s_1, \ldots, s_M) \) has determinant equal to \( [\sum_i |\bar{s}_i - s_i|^2]^n / 2 \neq 0 \) for any two distinct code sequences \( (\bar{s}_1 \ldots \bar{s}_M) \neq (s_1 \ldots s_M) \). In other words, the matrix \( B(c, e) \) is non-singular in this case and therefore is full rank. Thus, using a real orthogonal code, the maximum transmission rate is achieved with full spatial diversity.

Next, the decoding algorithm is considered. Since the rows of \( \mathcal{O} \) are all permutations of the first row with possibly different signs, let \( \epsilon_1, \ldots, \epsilon_M \) denote the permutations corresponding to these rows and let \( \xi_n(i) \) denote the sign of \( s_i \) in the \( n \)-th row. Then minimizing the metric (3.48) turns out to be the same as minimizing the following sum [27]

\[
\sum_{i=1}^{M} \left( \left( \sum_{n=1}^{M} \sum_{j=1}^{N} y_j^* \xi_n(i) - s_i \right)^2 + \left( \sum_{k,l} |a_{k,l}|^2 - 1 \right) |s_i|^2 \right)
\]

(3.49)

where \( \epsilon_n(p) = q \) means that \( s_p \) is at the \( (n, q) \) element of \( \mathcal{O} \). The sum

\[
\left( \left( \sum_{n=1}^{M} \sum_{j=1}^{N} y_j^* \xi_n(i) - s_i \right)^2 + \left( \sum_{k,l} |a_{k,l}|^2 - 1 \right) |s_i|^2 \right)
\]

(3.50)

only depends on \( s_i \), the received word, the path coefficients and the structure of the orthogonal design \( \mathcal{O} \). Therefore minimizing (3.49) amounts to minimizing this sum (3.50) for all \( 1 \leq i \leq M \). Thus the maximum likelihood detection rule is applied to form the decision variables

\[
R_i = \sum_{n=1}^{M} \sum_{j=1}^{N} y_j^* a_{e(i),j} \xi_n(i)
\]

for all \( i = 1, 2, \ldots, M \) and then \( s_i \) is chosen from all the constellation symbols \( s \) such that

\[
s_i = \arg \min_{s \in \mathcal{A}} |R_i - s|^2 + \left( \sum_{k,l} |a_{k,l}|^2 - 1 \right) |s|^2
\]

(3.51)

from which the code matrix \( \mathcal{O}(s_1, s_2, \ldots, s_M) \) can be reconstructed. This is indeed a very simple linear decoding strategy.
Generalized orthogonal design and simulation results

The existence problem for orthogonal designs has been studied and completed elsewhere (see reference in [27]), and in fact it has been found that such real orthogonal design exists if and only if $n = 2$, 4 or 8. This real orthogonal design can be expanded to a generalized real orthogonal design, which allows linear processing at the transmitter, giving rise to a non-square coding matrix, and to a generalized complex orthogonal design which contains complex code entries [27]. It is shown in the same paper that a generalized real design can achieve the maximum possible transmission rate for any number of transmit antennas using an arbitrary real constellation such as PAM, while still providing the full spatial diversity. The code construction was described and the specific code designs for cases with up to 8 transmit antennas were given. For complex constellations however, half of maximum rate can always be obtained, and full rate is possible only for the case of two transmit antennas. For cases of three and four transmit antennas, $3/4$ of maximum rate can be obtained using arbitrary complex constellations with specific design codes. Specifically in case of two transmit antennas ($M = 2$), the complex orthogonal code design, first found by Alamouti (see [27]), is given as

$$\mathcal{O}_c = \begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix}.$$  

This code provides full spatial diversity of $2N$ and has the same performance as diversity using an optimum combining with $2N$ receive antennas. The same maximum likelihood detection as (3.49) is used.

Simulation results for transmission rates of 3, 2 and 1 $bps/Hz$, each using 2, 3 and 4 transmit antennas with one receive antenna are provided in [28]. For a 3 $bps/Hz$ transmission rate, two transmit antennas employs 8-PSK constellation with full rate code and the three and four antennas employ 16-QAM constellation with a 3/4 rate code. For a 2 $bps/Hz$ case, two transmit antennas employ 4-PSK with full rate code, whereas three and four antennas employ 16-QAM with a half rate code. For the 1 $bps/Hz$ case, two antennas employ BPSK while three and four antennas employ 4-PSK with a half rate code. All cases showed a performance improvement in the sense that less SNR is required at a given BER as the number of transmit antenna increases. For example, at BER $= 10^{-5}$, the power gains that four transmit antennas have over two antennas are 7 $dB$, 5 $dB$ and 7.5 $dB$ for a 3, 2 and 1 $bps/Hz$ transmission rate respectively. When the number of receive antennas is increased to two, the performance
improves further but the gap between different transmit systems reduces, for example to about 3.5 dB in the last case. This is due to much of the diversity gain being already achieved using two transmit and two receive antennas.

This section has shown that transmit diversity combined with coding can achieve a performance improvement with increasing number of transmit antennas. The improvement here is expected to be more than that achieved by using transmit diversity only. Spatial diversity using multiple antennas can also be combined with other forms of diversity such as spectral diversity (frequency selective channels such as in wideband communication) or temporal diversity (time selective channels such as in fast mobile communication) to increase system capacity and performance. These provide scope for possible further research in these areas.
Chapter 4

Equalization of A Known Wireless Channel using Multiple Transmit Antennas

In the last chapter, we presented an overview of existing transmit diversity schemes. Except where diversity is combined with coding, all of the diversity schemes discussed impose linear processing on the data signal prior to the transmit antennas [20, 21, 23, 22, 25, 17, 18]. The attractiveness of linear processing is not only its simple implementation, but also because it leads to simple linear processing at the receiver while still achieving diversity gain, i.e., improving performance with increasing numbers of transmit antennas. While the asymptotic characteristics and performance of a system having an infinite number of transmit antennas have been established for a Rayleigh fading channel [18] with certain types of transmit antenna (linear) processing, analysis of the optimum linear receiver for any finite number of transmit antennas has not been carried out. In this chapter, we carry out this analysis to obtain the optimum linear diversity technique for multiple transmit antennas, with emphasis on Rician fading channels.

Again we study a single user transmit diversity system with $M$ transmit antennas and a single receive antenna. We also employ the quasi-stationary channel model for multiple antennas as in Chapter 1, where the communication channel is assumed to be frequency non-selective and slow fading such that the channel gain will stay constant during one symbol interval but will change randomly at the next one. Thus the channel is multiplicative with the channel coefficients modeled as special complex Gaussian
random variables in the Rayleigh or Rician fading. Linear processing is imposed on the transmit signal in the form of a linear prefilter in front of each transmit antenna. The information symbols are first linearly modulated using an appropriate modulation scheme, then this modulated signal will be processed by prefilters before simultaneously being passed to the transmit antennas. At the receiver, linear equalization followed by sampling at the symbol rate is used to detect the information symbols. The whole process is depicted in Figure 4.1.

We use a mean square error optimization criterion to derive the optimum linear equalizer and the optimum set of prefilters. This optimum criterion turns out to be similar to minimizing the variance of the squared Euclidean distance of the output sequences corresponding to single symbol inputs, as was also used in Section 3.3 above. However, the modulation diversity technique as described in Section 3.3 minimizes the variance of the squared Euclidean distance of the received signals, whereas we aim to minimize this variance of the equalized output symbols. Also instead of using MLSE with known channel coefficients at the receiver as in Section 3.3, we use a linear equalizer and study both cases where the channel coefficients are known and unknown at the receiver. The transmitter is always assumed not to know the channel coefficients. Results for the case of unknown channel coefficients at the receiver, which have not been derived previously, are presented in the next chapter. The frequency domain analysis presented here differs from current transmit diversity analysis techniques.

We assume ideal equalization with an infinite number of taps and ideal sampling at the receiver to detect the transmitted symbols, hence our results serve as an upper performance bound for any practical linear system of this structure. Results of our analysis for the known channel coefficients case show that the same equalizer as in [18]
(Section 3.5) can be used in both Rayleigh and Rician fading channels for any finite number of antennas, not only in the asymptotic case where the number of antennas goes to infinity, and that choice is MSE optimal. For a Rician fading channel, we derive the optimum set of prefilters in which all filters have the same frequency response which is the square root of a Nyquist pulse frequency response. This is different from the optimum set of prefilters for a Rayleigh fading channel established previously in the literature [22] (Section 3.3), where the tap coefficients of different (FIR) filters make up rows of a unitary matrix. It will also be shown that with the optimum set of prefilters for a Rician fading channel, the MSE will converge in a mean square sense to a minimum lower bound on the average MSE as the number of transmit antennas grows. This further implies that the symbol estimate for any input symbol will converge to its correct value in the mean square sense. A bound on the probability of symbol error for a rectangular QAM (quadrature amplitude modulation) signal is developed. For a Rician fading channel, it is shown that this error bound approaches zero exponentially as the number of transmit antennas increases, whereas for a Rayleigh fading channel, the bit error rate performance reduces to the results in [18].

The chapter is organized as follows: Section 4.1 develops the signal model, Section 4.2 explains the MSE optimization criterion and how it applies to our problem. The optimum linear equalizer is derived in Section 4.3, and the optimum set of prefilters for a Rician fading channel is derived in Section 4.4 through minimizing the MSE obtained with the optimum equalizer. In Section 4.5, we develop a bound on the probability of symbol error and provide an analysis for a Rician fading channel. Section 4.6 states our conclusions for the known channel case.

### 4.1 Signal model

Since the channel bandwidth is usually much smaller than the carrier frequency, the channel can be effectively modeled by the equivalent lowpass signal through an equivalent lowpass channel. Using a Rician fading model with Rayleigh fading as a special case, the multiplicative channel coefficients $a_i$ between the $i$th transmit antenna and the receiver ($i = 1 \ldots M$) are special complex Gaussian random variables with the same mean $\mu_a$ and variance $\sigma_a^2$. The transmit antennas are placed at least half a wavelength apart so that the fading between different transmit antennas and the receive antenna are uncorrelated, thus the coefficients $a_i$ are independent due to the Gaussian
distribution.

In the following sections we will develop signal models for our linear structure. Starting from a continuous model, we will build the discrete model which is used in the subsequent analysis. These signal models are similar to the general model developed in Chapter 1 but are re-established here specifically for the linear structure in Figure 4.1.

### 4.1.1 Continuous signal model

Let $c_n$ be the coded complex information data symbols which are assumed to be wide-sense stationary with zero mean and variance $\sigma^2$. Assume that the communication bandwidth is $W$ Hz and these information symbols are sent at the rate $T = 1/W$. We are interested in linear spectral efficient modulation methods such as quadrature amplitude modulation (QAM). For these linear modulation schemes, the complex baseband signal is generally represented as

$$s(t) = \sum_k c_k g(t - kT) \tag{4.1}$$

where $g(t)$ is the pulse shaping signal, which is usually chosen to have a spectrum such as the square root of the spectrum of a Nyquist pulse, i.e.

$$\sum_k \left| G(\omega + \frac{2\pi k}{T}) \right|^2 = T \tag{4.2}$$

where $G(\omega)$ is the Fourier transform of $g(t)$.

The modulated signal $s(t)$ is processed by a linear prefilter $h_i(t)$ before transmission from antenna $i$. Since the channel is multiplicative with coefficients $\{a_i\}$, the receive signal is $^1$

$$r(t) = \sum_{i=1}^{M} s(t) * h_i(t) \cdot a_i + n(t) = \sum_{i=1}^{M} a_i \int_{-\infty}^{\infty} s(\tau) h_i(t - \tau) d\tau + n(t) \tag{4.3}$$

where $n(t)$ is the white Gaussian noise added at the receiver with power spectral density $N_0$. Let $h(t)$ denote the equivalent channel which includes all the prefilters then

$$h(t) = \sum_{i=1}^{M} a_i \cdot h_i(t) \tag{4.4}$$

This equivalent channel model is depicted in Figure 4.2.

---

$^1$We use $\ast$ to denote convolution.
Figure 4.2: Equivalent signal model in continuous time

At the receiver, the signal is processed by a linear equalizer and then sampled at the symbol rate to obtain estimates of the information symbols. The output of the equalizer

\[ \hat{c}(t) = r(t) \ast b(t) \]  

(4.5)

passes through an ideal low-pass filter with cut-off frequency \( \pm W/2 \) before being sampled at rate \( T = 1/W \). The output of the sampler is an estimated sequence \( \{\hat{c}_n\} \) of the information symbols.

The overall input-output equation can thus be written as

\[ \hat{c}(t) = \left( \sum_k c_k g(t - kT) \right) \ast h(t) + n(t) \ast b(t) \]  

(4.6)

4.1.2 Discrete signal model

Since we wish to reconstruct the symbols by sampling the output of the equalizer, it is convenient to use a discrete time signal model. Let \( s[n] \) be the representation of \( s(t) \) in terms of its Nyquist T-spaced samples, i.e. \( s[k] = s(kT) \), and let \( r[k] \) denote the sequence arising from filtering the received signal \( r(t) \) by an ideal low-pass filter with cutoff frequency \( \pm W/2 \) and then sampling at symbol rate \( T \). The received sequence can be written as

\[ r[k] = s[k] \ast h[k] + n[k] = \sum_l s[l] \cdot h[k - l] + n[k] \]  

(4.7)

where \( n[k] = n(kT) \) are samples of the additive white Gaussian noise, which is a special complex Gaussian random variable with zero mean and variance \( \sigma_N^2 = N_0 \cdot W = N_0/T \), and \( h[k] \) is the discrete model of the equivalent channel

\[ h[k] = T \cdot h(kT) = \sum_{i=1}^{M} a_i \cdot h_i[k] \]  

(4.8)
which has frequency spectrum \(^2\) \(H^*(\omega)\) related to spectrum \(H(\omega)\) of the continuous channel by

\[
H^*(\omega) = \sum_k h[k] e^{-j \omega T k} = \sum_k H(\omega - \frac{2 \pi k}{T}). \tag{4.9}
\]

The estimate of the information symbol \(c_k\) is then the output of the discrete equalizer \(b[k] = b(kT)\), with \(r[n]\) as the input, i.e.

\[
\hat{c}_k = r[k] * b[k] = \sum_l r[l] b[k - l]. \tag{4.10}
\]

The representation of the modulated signal \(s(t)\) by its Nyquist samples also suggest the discrete representation for the linear modulation scheme

\[
s[k] = \sum_l c_l g[k - l] = c_k * g[k] \tag{4.11}
\]

where \(g[k] = g(kT)\). Together with (4.7) and (4.10), the estimated symbols can be written as

\[
\hat{c}_k = (c_k * g[k] * h[k] + n[k]) * b[k]. \tag{4.12}
\]

The discrete equivalent model is depicted in Figure 4.3.

Figure 4.3: Equivalent signal model in discrete time

Let \(a[k] = g[k] * h[k]\), then the processing of the information symbols by first linearly modulating the symbols by pulse shape \(g[k]\) and then passing them through the channel \(h[k]\) is equivalent to passing the unmodulated symbols through channel \(a[k]\) (Figure 4.4). The estimated symbols at the output become

\[
\hat{c}_k = (c_k * a[k] + n[k]) * b[k]. \tag{4.13}
\]

\(^2\)We use the superscript * to denote the frequency spectrum of a sampled (discrete) sequence
Frequency response of the discrete channel model

Note that all signals are now in discrete time, and we will use the discrete Fourier transform to represent the spectrum of these signals. For any discrete sequence \( \{ \xi[k] = \xi(kT) \} \), the frequency response is defined by

\[
\Xi^\ast(\omega) = \sum_k \xi[k] e^{-j\omega k T} . \tag{4.14}
\]

This frequency response is periodic with period \( 2\pi/T \), and it is related to the frequency response \( \Xi(\omega) \) of the continuous signal by

\[
\Xi^\ast(\omega) = \frac{1}{T} \sum_k \Xi \left( \omega + \frac{2\pi k}{T} \right) . \tag{4.15}
\]

We observe that within a period \([-2\pi/T, 2\pi/T]\), the frequency response of the discrete signal is equal to the frequency response of the continuous signal over the whole spectrum, except for a scale factor \( 1/T \), i.e.

\[
\Xi^\ast(\omega) = \frac{1}{T} \Xi(\omega_c) , \quad -\frac{2\pi}{T} \leq \omega \leq \frac{2\pi}{T} , \quad -\infty < \omega_c < \infty . \tag{4.16}
\]

To ensure that the total transmit power is independent of the number of transmit antennas, the frequency spectrum of the discrete prefilters \( H_i^\ast(\omega) \) must satisfy

\[
\sum_{i=1}^M |H_i^\ast(\omega)|^2 = 1 \quad \forall \omega \in [-\pi W, \pi W] . \tag{4.17}
\]

This is derived from a straightforward comparison between the input energy and the total energy transmitted from the \( M \) antennas

\[
\frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |S^\ast(\omega)|^2 d\omega = \sum_{i=1}^M \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |S^\ast(\omega)H_i^\ast(\omega)|^2 d\omega = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |S^\ast(\omega)|^2 \sum_{i=1}^M |H_i^\ast(\omega)|^2 d\omega
\]

where \( S^\ast(\omega) \) is the spectrum of the discrete modulated signal \( s[k] \).

Turning attention to the frequency response of the equivalent channel, we have

\[
A^\ast(\omega) = G^\ast(\omega) \cdot H^\ast(\omega) = G^\ast(\omega) \cdot \sum_{i=1}^M a_i H_i^\ast(\omega) . \tag{4.18}
\]
Since the $a_i$ are special complex Gaussian random variables, the channel response $A^s(\omega)$ is also a special complex Gaussian random variable at each frequency point $\omega$ with mean and variance given by

$$
\mu_A^s(\omega) = G^s(\omega) \cdot \mu_a \sum_{i=1}^{M} H_i^s(\omega) \quad (4.19)
$$

$$
\sigma_A^2(\omega) = |G^s(\omega)|^2 \cdot \sigma_a^2 \sum_{i=1}^{M} |H_i^s(\omega)|^2. \quad (4.20)
$$

Since the modulation pulse $g(t)$ is assumed to have a frequency spectrum equal to the square root spectrum of a perfect Nyquist pulse, from (4.2) and (4.16), we have

$$
|G^s(\omega)|^2 = \frac{1}{T^2} |G(\omega_c)|^2 = \frac{1}{T^2} \sum_{k} |G(\omega - \frac{2\pi}{T} k)|^2 = \frac{1}{T} \quad (4.21)
$$

with $-\frac{\pi}{T} \leq \omega \leq \frac{\pi}{T}$ and $-\infty < \omega_c < \infty$, where $\omega_c$ denote the continuous spectrum. Combined with the constraint on the prefilters $H_i^s(\omega)$ in (4.17), the variance of $A^s(\omega)$ becomes

$$
\sigma_A^2(\omega) = \frac{1}{T} \sigma_a^2 \quad (4.22)
$$

for all frequencies within a period, whereas the square spectrum of the equivalent channel is

$$
|A^s(\omega)|^2 = \frac{1}{T} \sum_{i=1}^{M} a_i H_i^s(\omega)|^2. \quad (4.23)
$$

Let $\gamma$ denote the receive signal-to-noise (SNR) ratio in the general discrete channel (Figure 4.4), then

$$
\gamma = \frac{E[c_k * a[k]]^2}{E[n[k]]^2} = \frac{\sigma_c^2}{\sigma_N^2} \cdot \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |A^s(\omega)|^2 d\omega. \quad (4.24)
$$

Again this receive SNR is a function of channel coefficients and therefore is a random variable.

### 4.2 Optimization criterion

The mean square error (MSE) criterion has been widely used in optimizing equalizer coefficients for communication systems. Using the MSE criterion, we aim to minimize the square of the error between a symbol estimate and the correct input symbol, which is equivalent to minimizing the symbol error energy. Assuming the data sequence and
the additive noise samples are wide-sense stationary, the error sequence \( \xi_k = \hat{c}_k - c_k \) is also stationary. The MSE value therefore does not depend on the symbol index \( k \) and can be written as

\[
\epsilon = E_{c_k,N}[\xi_k^2] = E_{c_k,N}[\hat{c}_k - c_k]^2
\]

where the expectation is taken over the transmit symbol \( c_k \) and the additive noise \( N \). Thus the optimization problem becomes one of finding a linear equalizer \( b[k] \) and a set of linear prefilters \( h_i[k] \) to minimize the MSE expression, i.e.

\[
\epsilon_{\text{min}} = \min_{b,h_i} E_{c_k,N} \left[ |\hat{c}_k - c_k|^2 \right].
\] (4.26)

We will now construct the expression for minimizing this performance index. Using the following definition and theorem, the MSE is transformed to the frequency domain.

**Definition 4.1.** For a symbol sequence \( \{\xi_k\} \) at rate \( T \), define the \( T \)-interval frequency response as

\[
\Xi_T(\omega) = \int_{-T/2}^{T/2} \xi(t)e^{-j\omega t} dt = \int_{-T/2}^{T/2} \sum_k \xi_k \delta(t - kT)e^{-j\omega t} dt,
\]

where \( \xi(t) = \sum_k \xi_k \delta(t - kT) \) is the continuous representation of the symbol sequence \( \xi_k \).

**Theorem 4.1.** Given a wide-sense stationary symbol sequence \( \xi_k \) at rate \( T \), its average symbol power can be written in the frequency domain as

\[
E[|\xi_k|^2] = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} E[|\Xi_T(\omega)|^2] d\omega
\] (4.27)

where \( \Xi_T(\omega) \) is the \( T \)-interval frequency response symbol sequence \( \{\xi_k\} \).

**Proof.** Since the symbol sequence \( \xi_k \) is stationary, let \( I_m = E[\xi_k \xi_{k+m}^*] \) denote the
autocorrelation of the sequence. Then

$$E[\Xi_T(\omega)]^2 = E\left[\left(\int_{-T/2}^{T/2} \sum_k \xi_k \delta(t - kT)e^{-j\omega t} dt\right)\left(\int_{-T/2}^{T/2} \sum_k \xi_k^* \delta(t - kT)e^{j\omega t} dt\right)\right]$$

$$= \sum_k \sum_m E[\xi_k \xi_m^*] \left(\int_{-T/2}^{T/2} \delta(t - kT)e^{-j\omega t} dt\right)\left(\int_{-T/2}^{T/2} \delta(t - mT)e^{j\omega t} dt\right)$$

$$= \sum_m I_m \sum_k \left(\int_{-T/2}^{T/2} \delta(t - kT)e^{-j\omega t} dt\right)\left(\int_{-T/2}^{T/2} \delta(t - mT)e^{j\omega t} dt\right)$$

$$= \sum_m I_m \left(\int_{-T/2}^{T/2} \delta(t)e^{-j\omega t} dt\right)\left(\int_{-T/2}^{T/2} \delta(t)e^{j\omega t} dt\right)$$

$$= \sum_m I_m e^{-j\omega mT} \quad (4.28)$$

On the other hand, $E[\Xi_T(\omega)]^2$ is periodic with period $\frac{2\pi}{T}$ since

$$E[\Xi_T(\omega + \frac{2\pi}{T})]^2 = \sum_m I_m e^{-j(w + \frac{2\pi}{T})mT} = \sum_m I_m e^{-j\omega mT}$$

which is the same expression as (4.28). Thus its Fourier series exists. The expression (4.28) is in the same form as a Fourier series expansion, and therefore $I_m$ are the Fourier coefficients which, by definition, are given by

$$I_m = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} E[\Xi_T(\omega)]^2 e^{-j\frac{2\pi}{T}\omega} d\omega.$$ 

With $m = 0$, this establishes the theorem. \hfill \Box

Applying Theorem 4.1 to our problem, the optimization criterion (4.26) can be rewritten in the frequency domain as

$$\epsilon_{\text{min}} = \min_{B, H_T} \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} E_{c_k, N} \left[|\hat{C}_T(\omega) - C_T(\omega)|^2\right] d\omega \quad (4.29)$$

where $\hat{C}_T(\omega)$ and $C_T(\omega)$ are the $T$-interval frequency responses of $\{\hat{c}_k\}$ and $\{c_k\}$ as defined in definition 4.1, and $B$ and $H_T$ are normal frequency responses of the discrete
linear equalizer \( b[k] \) and discrete prefilters \( h_i[k] \) as in (4.14) respectively. Since the function under the integral in (4.29) is non-negative, the value of the integral can be minimized by minimizing this function. Thus, the problem becomes one of finding the equalizer \( B \) and prefilters \( H_i \) to minimize the following expression in the frequency domain

\[
J(\omega) = E_{c_k,N} \left[ |\hat{C}_T(\omega) - C_T(\omega)|^2 \right] 
\]  \hspace{1cm} (4.30)

and the MSE (4.25) can be written as

\[
\epsilon = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} J(\omega) \, d\omega . 
\]  \hspace{1cm} (4.31)

Noticing that the \( T \)-interval frequency response of a convolution between two signals \( f(t) = u(t) * v(t) \) can be written as \( F_T(\omega) = V_T(\omega) \cdot U(\omega) \) as a straightforward expansion of the standard convolution result, from (4.13), the \( T \)-interval frequency response of the symbol estimate sequence \( \{\hat{c}_k\} \) becomes

\[
\hat{C}_T(\omega) = \left[ C_T(\omega)A^*(\omega) + N_T^*(\omega) \right] B^*(\omega) 
\]  \hspace{1cm} (4.32)

where \( A^*(\omega) \) and \( B^*(\omega) \) are the Fourier transform of the discrete channel \( a[k] \) and equalizer \( b[k] \). To simplify notation, we use the shorthand \( C_T(\omega) = C_T, \hat{C}_T(\omega) = \hat{C}_T, \)

\( A^*(\omega) = A, \) \( B^*(\omega) = B \) and \( N_T^*(\omega) = N_T \) to denote the frequency responses. From (4.28), when the information symbols \( \{c_k\} \) are mutually uncorrelated with zero mean and variance \( \sigma_c^2 \), the symbol covariance will be \( I_m = \sigma_c^2 \delta_m \). Therefore

\[
E|C_T(\omega)|^2 = \sigma_c^2 \quad \forall \omega .
\]

Similarly, \( E|N_T|^2 = \sigma_N^2 = N_0/T \). Substituting (4.32) into (4.30) and taking expectations with respect to the information symbols and noise, noting that \( E|N_T| = 0 \), we obtain

\[
J(\omega) = \sigma_c^2 |A|^2 |B|^2 + \sigma_N^2 |B|^2 + \sigma_c^2 - \sigma_c^2 AB - \sigma_c^2 A^* B^* .
\]  \hspace{1cm} (4.33)

Also recall that \( A = A^*(\omega) \), the frequency response of the general equivalent channel \( a[k] = g[k] * h[k] \), is a random special complex Gaussian variable in frequency \( \omega \) with mean and variance given by (4.20) and (4.22). Thus at each frequency point \( \omega \), \( J(\omega) \) given by (4.33) is a random variable.

We seek to minimize this expression for \( J(\omega) \) in (4.33) by first finding the optimum choice of equalizer \( B \), then finding the optimum set of the prefilters \( H_i \) given the optimal
equalizer, subject to the constraint (4.17), i.e.

$$\min_{B,H_i} J(\omega), \quad \omega \in [-\pi W, \pi W]$$

$$\text{s.t.} \quad \sum_{i=1}^{M} |H_i^*(\omega)|^2 = 1 . \quad (4.34)$$

The value of the equalizer \( B \) in turn depends on the value of the channel coefficients \( a_i \), or their distributions if the values are unknown at the receiver, and the choice of the prefilters \( H_i \). When the values of \( a_i \) are not known at the receiver, we will assume that the statistical mean and variance of these coefficients are still available, and this case will be treated in the next chapter. The transmitter is always assumed to have no knowledge of the values of \( a_i \) but only the Gaussian distribution characteristic of these channel coefficients. In the following sections, we will analyze the case of known channel coefficients at the receiver.

### 4.3 Optimum equalizer

With known channel coefficients, the equivalent channel response \( A^*(\omega) \) is known at the receiver. Thus the value of the equalizer response \( B^*(\omega) \) will depend on the value of \( A^*(\omega) \), in other words, \( B^*(\omega) \) is a function of the equivalent channel \( A^*(\omega) \). We have the following theorem.

**Theorem 4.2.** The optimum linear equalizer that minimizes the mean square error (4.31) for the system described in (4.13), has the frequency spectrum of its sampled taps given by

$$B^*_0(\omega) = \frac{(A^*(\omega))^*}{|A^*(\omega)|^2 + \sigma_c^2 + \sigma_N^2} \quad (4.35)$$

where \( A^*(\omega) \) is the frequency spectrum of the discrete channel and \( \sigma_c^2 \) and \( \sigma_N^2 \) are the variances of the input symbols and the sampled noise respectively.

**Proof.** Here we aim to find \( B \) that minimizes the expression for \( J(\omega) \) given by (4.33). Employing the inequality

$$AB + A^*B^* \leq 2|A||B| \quad (4.36)$$

where equality happens only when (see appendix for proof)

$$A_1B_R = -A_RB_1 \quad , \quad (4.37)$$
it can be seen that $J(\omega)$ satisfies

$$J(\omega) \geq \sigma_c^2 |A|^2 |B|^2 + \sigma_N^2 |B|^2 - 2\sigma_c^2 |A||B| + \sigma_c^2 = f(|B|).$$ \hspace{1cm} (4.38)

We will minimize the right hand side expression first, which is a function of $|B|$. The value of $|B|$ that minimizes this function, is one that makes the function’s first derivative equal to zero while the second derivative is positive. Equating the first derivative to zero, we have

$$\frac{df(|B|)}{d|B|} = 2\sigma_c^2 |A|^2 |B| + 2\sigma_N^2 |B| - 2\sigma_c^2 |A| = 0$$

which gives

$$|B| = \frac{\sigma_c^2 |A|}{\sigma_c^2 |A|^2 + \sigma_N^2}.$$ \hspace{1cm} (4.39)

A simple check shows that the second derivative at this point is positive. Thus (4.39) is the minimum point of $f(|B|)$ in (4.38). Combined with the equality condition (4.37), we obtain the optimum linear equalizer (4.35).

The resulting equalizer is in the same form as the MSE optimum equalizer in [1], and is also the optimum equalizer for the limiting case where number of transmit antennas goes to infinity as derived in [18].

Notice that the denominator expression of the optimum equalizer (4.35) contains two terms, one is the equivalent channel power gain $|A^*(\omega)|^2$ which depends on the channel coefficients and the prefilters, and the other is the inverse of the input signal to noise ratio ($\rho_0 = \sigma_c^2 / \sigma_N^2$). Thus, it is interesting to look at these results in the two asymptotic cases.

First, when the SNR is relatively large compared to equivalent channel gain, i.e., $\rho_0$ is small compared to $|A^*(\omega)|^2$, the dominating term in the denominator is $|A^*(\omega)|^2$. The optimum equalizer therefore comes close to $A^*(\omega)^{-1}$, which is the inverse filter.

The second case is when the SNR is relatively small compared to the equivalent channel gain. In this case, the denominator is dominated by term $\rho_0$. The optimum equalizer in this case comes close to $\rho_0 \cdot (A^*(\omega))^*$, which is a matched filter.

These two limiting cases can be summarized as below.

$$\rho_0 = \frac{\sigma_c^2}{\sigma_N^2} \text{ large } \implies B^*_0(\omega) \approx A^*(\omega)^{-1} \text{ (inverse filter)}$$

$$\rho_0 = \frac{\sigma_c^2}{\sigma_N^2} \text{ small } \implies B^*_0(\omega) \sim (A^*(\omega))^* \text{ (matched filter)}.$$ 

105
From the expression for the optimum equalizer (4.35), if we let \( u[k] \) be the sample sequence that has a frequency response given by

\[
U^*(\omega) = \sum_k u[k] e^{-j\omega kT} = \frac{1}{|A^*(\omega)|^2 + \frac{\sigma^2_q}{\sigma^2_e}}
\]

then the sequence \( u[k] \) can be uniquely determined at the receiver. The optimum equalizer then has the frequency spectrum of the sample sequence in the form \( B^*(\omega) = U^*(\omega) \cdot (A^*(\omega))^* \), which suggests that the sampled taps of this equalizer are

\[
b[k] = \sum_l u[l] a^*[l - k] .
\]

In other words, the optimum equalizer structure can be interpreted as consisting of a matched filter matched to the channel, which is first sampled with period \( T \), then these samples are passed through an infinite tapped-delay-line, T-spaced, with tap coefficients given by \( u[k] \).

### 4.4 Minimum mean square error and optimum prefilters

Having found the optimum linear equalizer, we can now establish the value of the MSE with this equalizer. Subsequently, we will find the set of optimum prefilters \( H^*_t(\omega) \) by minimizing this MSE.

Since the optimum equalizer (4.35) is a function of channel response \( A^*(\omega) \), the MSE will also be a function of the channel response, and therefore will be a random variable. Replacing the optimum equalizer (4.35) in the expression for \( J_o(\omega) \) in (4.33) we obtain

\[
J_o(\omega) = \frac{\sigma^2_q}{|A^*(\omega)|^2 + \frac{\sigma^2_q}{\sigma^2_e}} = \frac{N_0}{\sum_{i=1}^M a_i H^*_i(\omega) + \frac{N_0}{\sigma^2_e}}
\]

where the second equality results from (4.23). Denote \( \zeta_0 = N_0/\sigma^2_e \), then the mean square error (4.31) obtained as a result of the optimum equalizer (4.35) is

\[
\epsilon_0 = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{N_0}{\sum_{i=1}^M a_i H^*_i(\omega) + \zeta_0} d\omega .
\]

We will now analyze this expression to find the optimum set of prefilters \( H^*_t(\omega) \) which minimizes the optimum MSE. Since the channel coefficients \( a_i \) are random and unknown at the transmitter, so that the prefilters \( H^*_t(\omega) \) cannot be functions of these channel coefficients, the optimum MSE is also a random variable. It then makes sense to find
\( H_\alpha^*(\omega) \) that minimize the expected value of this random variable, which corresponds to minimizing the average MSE. We have the following theorem.

**Theorem 4.3.** For a Rician fading channel \((\mu_\alpha \neq 0)\), the set of linear prefilters which have the same frequency responses satisfying

\[
\left| H_\alpha^*(\omega) \right| = \frac{1}{\sqrt{M}} \quad \forall \quad i = 1 \ldots M, \quad -\pi W \leq \omega \leq \pi W \quad (4.42)
\]

will make the optimum mean square error (4.41) (which is obtained as a result of the optimum linear equalizer given in Theorem 4.2), converge in mean-square to its average minimum lower bound

\[
\varepsilon_0 \cdot \frac{N_0}{M |\mu_\alpha|^2 + \sigma^2_\alpha + \frac{\sigma^2_\alpha}{M}} \quad (4.43)
\]

where \( \mu_\alpha \) and \( \sigma^2_\alpha \) respectively are the mean and variance of the complex fading channel coefficients.

**Proof.** Applying Jensen’s inequality for the convex function \((x + \alpha)^{-1}\) with \( x \geq 0 \) and constant \( \alpha > 0 \), a lower bound for the average \( J_\alpha(\omega) \) is found to be

\[
E_{a_i} \left[ \frac{N_0}{\left| \sum_i a_i H_i \right|^2 + \zeta_0} \right] \geq \frac{N_0}{E_{a_i} \left[ \left| \sum_i a_i H_i \right|^2 \right] + \zeta_0} = \frac{N_0}{|\mu_\alpha|^2 |\sum_i H_i|^2 + \sigma^2_\alpha + \zeta_0}. \quad (4.44)
\]

Thus the average optimum MSE satisfies

\[
\varepsilon_0 \geq \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{N_0}{\sum_{i=1}^{M} H_\alpha^*(\omega) |\mu_\alpha|^2 + \sigma^2_\alpha + \zeta_0} \, d\omega \quad (4.45)
\]

where the equality happens if and only if

\[
\left| \sum_i a_i H_\alpha^*(\omega) \right|^2 = E_{a_i} \left[ \left| \sum_i a_i H_\alpha^*(\omega) \right|^2 \right] \quad \forall \omega \in \left[ -\frac{\pi}{T}, \frac{\pi}{T} \right]
\]

i.e., \( \left| \sum_i a_i H_\alpha^*(\omega) \right|^2 \) is a constant for all \( \omega \) within one period \([-\pi/T, \pi/T]\). However, unless the transmitter knows the channel coefficients \( a_i \), it is not possible to choose the set of prefilter \( \{ H_\alpha^*(\omega) \} \) that makes \( \left| \sum_i a_i H_\alpha^*(\omega) \right|^2 \) constant at every frequency point \( \omega \). Thus, when the channel coefficients are not known at the transmitter, the inequality (4.45) will serve as a lower bound for the average optimum MSE. It will be of interest to find out how close we can approach this error bound with a certain design of the prefilters \( H_\alpha^*(\omega) \).

In order to find the set of prefilters \( \{ H_\alpha^*(\omega) \} \) that minimizes the lower bound on the average MSE in (4.45), we need to maximize the square of the sum \( H_\alpha^*(\omega) \) term in
the denominator, subject to the constraint (4.17). Employing the inequality

\[ \left| \sum_{i=1}^{M} H_i^*(\omega) \right|^2 \leq M \sum_{i=1}^{M} |H_i^*(\omega)|^2 \]  

(4.46)

where equality happens if and only if all \( H_i^*(\omega) \) are the same, provided that \( \mu_a \neq 0 \), we obtain the minimum of the lower bound on the optimum MSE as

\[ \tilde{\varepsilon}_o \geq \frac{N_0}{M} |\mu_a|^2 + \sigma_a^2 + \zeta_0 = \tilde{\varepsilon}_{min} . \]  

(4.47)

The set of prefil ters \( \{ H_i(\omega) \} \) that minimizes this lower bound is given by (4.42). With this set of \( \{ H_i^*(\omega) \} \), the optimum mean square error (4.41) becomes

\[ \varepsilon_o = \frac{N_0}{\sum_{i=1}^{M} a_i^2 + \sum_{i=1}^{M} a_i^2 + \zeta_0} = \frac{\sigma_e^2}{\gamma + 1} \]  

(4.48)

where \( \gamma \) is the receive SNR (4.24), which, with the set of prefil ters in (4.42), is given by

\[ \gamma = \frac{\sigma_e^2}{N_0 \cdot M} \sum_{i=1}^{M} a_i^2 = \frac{\sum_{i=1}^{M} a_i^2}{M \zeta_0} . \]  

(4.49)

Let \( a_M = (\sum_{i=1}^{M} a_i)/M \). Since the \( a_i \) are independent special complex Gaussian random variables with mean \( \mu_a \) and variance \( \sigma_a^2 \), \( a_M \) is also a special complex Gaussian random variable with mean \( \bar{a}_M = \mu_a \) and variance \( var[a_M] = \sigma_a^2/M \). Let \( x_M = |a_M|^2 \) then \( x_M \) is a chi-square random variable with two degree with freedom, which has mean \( \mu_M = E[|a|^2] = |\mu_a|^2 + \sigma_a^2/M \) and variance \( var[x_M] = \text{var}[|a|^2] = \sigma_a^2/M^2 + 2|\mu_a|^2 \sigma_a^2/M \) (see Appendix). Since \( var[x] \to 0 \ as \ M \to \infty \), \( x_M \) converges in mean-square to its mean value \( \mu_M \) as the number of antennas goes to infinity. We can write

\[ \lim_{M \to \infty} E[|x_M - \mu_M|^2] = 0 . \]  

(4.50)

The optimum MSE can be written as

\[ \varepsilon_o = \frac{N_0}{M x_M + \zeta_0} . \]

It then can be shown that as the number of transmit antennas \( M \) grows large, the optimum mean square error \( \varepsilon_o \) will converge in mean-square to the lower bound in (4.47). Consider

\[ \left| \frac{N_0}{M x_M + \zeta_0} - \frac{N_0}{M \mu_M + \zeta_0} \right|^2 = \left| \frac{N_0^2}{M x_M + \zeta_0} - \frac{N_0^2}{M \mu_M + \zeta_0} \right|^2 \leq \left| \frac{N_0^2}{\zeta_0 \cdot \mu_M} \right| \left[ x_M - \mu_M \right]^2 \]  

108
Therefore, provided that $\mu_a \neq 0$,

$$0 \leq \lim_{M \to \infty} E[\epsilon_o - \bar{\epsilon}_{\text{min}}]^2 \leq \lim_{M \to \infty} E \left[ \frac{|N_o|^2}{\epsilon_o \cdot |\mu_a|^2} |x_M - \mu_M|^2 \right] = 0$$

(4.51)

where the last inequality follows from (4.50). In other word, $\epsilon_o$ converges to $\bar{\epsilon}_{\text{min}}$ in the mean-square sense.

Thus for Rician fading, the set of prefilters (4.42) is the optimum set that makes the optimum MSE (4.41) converges to the minimum lower bound of the average MSE (4.47) in a mean-square sense as the number of transmit antennas grows large.

Note that the convergence in mean-square of the MSE in (4.51) also implies convergence in mean-square sense of a symbol estimate to its correct value, since $\bar{\epsilon}_{\text{min}} \to 0$ as $M \to \infty$. Indeed

$$0 = \lim_{M \to \infty} E[\epsilon_o - \bar{\epsilon}_{\text{min}}]^2 = \lim_{M \to \infty} E[\epsilon_o^2 - 2\bar{\epsilon}_{\text{min}} \epsilon_o + \bar{\epsilon}_{\text{min}}^2] = \lim_{M \to \infty} E[\epsilon_o^2] = \lim_{M \to \infty} E[\hat{\epsilon}_k - \epsilon_k]^2$$

(4.52)

Thus the symbol estimate converges to the symbol value for any input symbol in 4th order, i.e.

$$\hat{\epsilon}_k \xrightarrow{k} \epsilon_k \quad \forall \quad k$$

which implies mean-square convergence (i.e., converges in 2nd order), where the expectation is taken over the input symbol, noise and the fading channel coefficients.

Since $H^*_i(\omega)$ is the frequency response of the samples $h_i[k] = T \cdot h_i(kT)$ of the prefilter $h_i(t)$, let $X_i(\omega) = |H_i(\omega)|^2$, then the optimum condition of $H^*_i(\omega)$ in (4.42) is equivalent to

$$\sum_k X_i(\omega + k \frac{2\pi}{T}) = \sum_k |H_i(\omega + k \frac{2\pi}{T})|^2 = \frac{1}{M} \quad \forall \quad -\pi W \leq \omega \leq \pi W .$$

Thus $X_i(\omega)$ has a frequency response the satisfies the Nyquist condition, which leads to the prefilters $\{h_i(t)\}$ having the same frequency spectrum as the square root of a Nyquist pulse spectrum. Since the sample of a Nyquist pulse contains only one non-zero pulse $x_i[0] = 1$ (normalized to signal power $2\pi/(MT)$), this implies

$$x_i[n] = \sum_k h_i[k] h^*_i[n - k] = \frac{2\pi}{MT} \delta_n \quad \forall \quad i = 1 \ldots M .$$

(4.53)

Moreover, since all the optimum prefilters have the same discrete time frequency response, the discrete time taps of these prefilters must be of the same sequence. Thus
we can replace the prefilter set by a single filter that satisfies (4.42) or equivalently (4.53).

Contrast this result for a Rician flat fading channel with the result for a Rayleigh flat fading channel previously found in the literature [22, 18]. Using the equivalent criterion of minimizing the variance of the square Euclidean distance between the output sequences corresponding to single symbol inputs, Wittneben [22] showed that the optimum set of prefilters for a Rayleigh flat fading channel is obtained when the tap sequences of these prefilters make up rows of a unitary matrix (3.31), assuming they are FIR filters (Section 3.3). Wornell and Trott [18] then generalized this result and proved that with this optimum prefilter set for a Rayleigh flat fading channel, when the number of antennas goes to infinity, the same optimal linear equalizer as (4.35) is obtained, and the system converges to a non-fading marginally white Gaussian noise channel in a mean-square sense (Section 3.5). In fact, our result shows that the optimum linear equalizer is the same for both Rayleigh and Rician fading channels, and it is MSE optimum for any number of antennas, not only in the asymptotic case with an infinite number of transmit antennas.

### 4.5 Evaluating performance by bound on probability of error

Here we adopt a result by Balaban and Salz [34], which is an application of a more general result by Saltzberg [36], on an upper bound on the probability of symbol error for rectangular QAM. When rectangular QAM is used, i.e. the complex points \( c_k \) lie on a lattice and decisions are made independently on the real and imaginary axes, then an upper bound on the probability of symbol error can be derived from the optimum MSE. Below we apply these results to our problem and obtain Theorem 4.4.

**Definition 4.2.** A rectangular complex symbol is a random multiple level symbol \( c_k = a_k + jb_k \) that has zero-mean and variance \( \sigma_c^2 \). The real and imaginary parts \( a_k \) and \( b_k \) are identically distributed random variables with zero mean and variance \( \sigma_a^2 = \sigma_b^2 = \frac{1}{2}\sigma_c^2 \), and each may independently take any value of \( N \) discrete values in the set \( \{ \pm1, \pm3, \ldots, \pm(N - 1) \} \) with equal probability.

**Theorem 4.4.** Let \( c_k \) be a sequence of wide-sense stationary, mutually uncorrelated rectangular complex information symbols that is the input to a channel with kernel \( a[k] \),
which is linearly equalized at the receiver (4.13) by the optimum equalizer \( b[k] \) given by

\[
P_e \leq 4 \exp \left( -\frac{1}{\epsilon_o} + \frac{1}{\sigma_c^2} \right)
\]

where \( \epsilon_o \) is the optimum mean square error obtained with this optimum equalizer given in (4.41), and \( \sigma_c^2 = E[|c_k|^2] \).

**Proof.** The equation for the symbol estimate, i.e., the output of the equalizer, in (4.13) is repeated here for convenience

\[
\hat{c}_k = c_k \ast a[k] \ast b[n] + n[k] \ast b[k]
\]

where \( n[k] \) is sample of additive white Gaussian noise. Let \( f[k] = a[k] \ast b[k] \) and \( w[k] = n[k] \ast b[k] \), then \( w[k] \) is also a zero-mean complex Gaussian noise sequence, with variance \( \sigma_w^2 \). The symbol estimate of the \( k^{th} \) input symbol can then be written as

\[
\hat{c}_k = c_k \ast f[k] + n[k] = c_k f[0] + \sum_{l \neq 0} c_{k-l} f[l] + w[k].
\]

Since the real and imaginary parts of \( c_k \) are detected independently, we express the output of the detector in terms of real and imaginary parts. Without loss of generality, assume \( f[0] \) is real, and let \( f[k] = f_r[k] + j f_i[k] \), we obtain

\[
\hat{a}_k = a_k f[0] + \sum_{l \neq 0} \left( a_{k-l} f_r[l] - b_{k-l} f_i[l] \right) + w_r[k] = a_k f[0] + \sum_{l \neq 0} z_l + w_r[k]
\]

\[
\hat{b}_k = b_k f[0] + \sum_{l \neq 0} \left( a_{k-l} f_i[l] + b_{k-l} f_r[l] \right) + w_i[k] = b_k f[0] + \sum_{l \neq 0} y_l + w_i[k]
\]

where \( z_l \) and \( y_l \) are random variables which represent the corresponding expressions for \( a_{k-l} \) and \( b_{k-l} \), \( w_r[k] \) and \( w_i[k] \) are the real and imaginary parts of the noise, which have zero mean and variance \( \sigma_w^2/2 \).

Since \( a_k \) and \( b_k \) are independent, identically distributed random variables, and each can take any value

\[
\alpha_j = 2j - N - 1, \quad j = 1, 2, \ldots, N \tag{4.54}
\]

with equal probability, \( z_l \) and \( y_l \) are independent and have the same probability density

\[
p(z_l) = p(y_l) = \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \delta(z_l - \alpha_j f_r[l] - \alpha_k f_i[l]) . \tag{4.55}
\]
Therefore the probability of error in detecting \( a_k \) will be the same as that for \( b_k \). Let \( P_a \) denote the probability of error in detecting \( a_k \) then the probability of symbol error in detecting \( c_k \) will be \( P_e = 2P_a \).

Let

\[
z = \sum_{l \neq 0} z_l + w_r[k]
\]

then an estimate of the real part can be written as

\[
\hat{a}_k = a_k \cdot f[0] + z.
\]

Since \( a_k \) can take one of the multiple levels \( \alpha_j \), detecting \( a_k \) amounts to placing thresholds at 0, \( \pm 2 \), \( \ldots, \pm (N - 2) \) and comparing these with the value of \( \hat{a}[k] \). When \( a_k = -N + 1 \), an error results if \( z > f[0] \), and when \( a_k = N - 1 \), an error results if \( z < -f[0] \). With all other symbol values, either condition results in an error.

It can be seen that the probability density of \( z \) is symmetric about zero, therefore

\[
P[z > f[0]] = P[z < -f[0]].
\]

Thus the error probability for detecting \( a_k \) can be approximated as

\[
P_a = \frac{2(N - 1)}{N} Pr[z > f[0]] \approx 2 Pr[z > f[0]].
\]

Applying the Chernoff bound [36] yields the upper bound (the derivation is in the Appendix)

\[
Pr[z > f[0]] < \exp\left(\frac{-f[0]^2}{2(\sigma_w^2 + \sigma_a^2 \sum_{k \neq 0} |f[k]|^2)}\right)
\]

where \( \sigma_w^2 = E[w_r[k]^2] = \sigma_w^2/2 \) and \( \sigma_a^2 = E[a_k^2] = \sigma_a^2/2 \). Therefore the probability of symbol error in detecting \( c_k \) can be upper bounded by

\[
P_e < 4 \exp\left(\frac{-f[0]^2}{(\sigma_w^2 + \sigma_a^2 \sum_{k \neq 0} |f[k]|^2)}\right).
\]

Now notice that by definition, the mean square error of the system can be written as

\[
\epsilon = E[\hat{c}_k - c_k]^2 = E[c_k f[0] + \sum_{l \neq 0} c_{k-l} f[l] + n[k] - c_k]^2 = \sigma_c^2 (1 - f[0])^2 + \sigma_a^2 \sum_{l \neq 0} |f[l]|^2 + \sigma_w^2
\]

due to the wide-sense stationary and mutually uncorrelated symbol sequence \( \{c_k\} \).

Thus the probability of error satisfies

\[
P_e < 4 \exp\left(\frac{-f[0]^2}{\epsilon - \sigma_c^2 (1 - f[0])^2}\right).
\]

112
With the optimum equalizer given in Theorem 4.2, the optimum MSE is given by (4.41). Thus we can express the optimum \( J_o(\omega) \) in (4.40) as

\[
J_o(\omega) = \sigma_c^2 \left[ 1 - A^*(\omega)B_o^*(\omega) \right] = \sigma_c^2 \left[ 1 - F^*(\omega) \right]
\]

where \( F^*(\omega) \) is the frequency response of discrete signal \( f[k] \). The optimum MSE can therefore be rewritten as

\[
\varepsilon_o = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \sigma_c^2 \left[ 1 - F^*(\omega) \right] d\omega = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \sigma_c^2 \left[ 1 - F^*(\omega) \right] d\omega = \sigma_c^2 \left[ 1 - f[0] \right]
\]

where the last equality results from taking the inverse discrete-time Fourier transform of \( F^*(\omega) \). Substituting this into (4.59), we get the result of the theorem. \( \square \)

Apply this theorem to our problem, using the optimum set of prefilters for Rician fading channel given by Theorem 4.3, the optimum mean square error is then given by (4.48). Thus we have

\[
P_e \leq 4 \exp \left( -\frac{\left| \sum_{k=1}^{M} a_k \right|^2}{MN_0} \right) = 4 \exp \left( -\frac{\gamma}{\sigma_c^2} \right)
\]

with \( \gamma \) is the receive SNR in (4.49). Note that \( \gamma \) is then a chi-squared random variable with two degrees of freedom. Let

\[
\mu = \frac{M\sigma_c^2 |\mu_o|^2}{N_0}, \quad \sigma^2 = \frac{\sigma_c^2 \cdot \sigma_a^2}{N_0}
\]

then \( \mu \) is the noncentrality parameter of the chi-squared random variable \( \gamma \). The average receive SNR is

\[
\gamma_0 = \mu + \sigma^2 = \frac{\sigma_c^2}{N_0} \left( M |\mu_o|^2 + \sigma_a^2 \right)
\]

where \( \mu \) represents average SNR of the direct path and \( \sigma^2 \) represents average SNR of the scattering components.

Averaging the upper error bound with respect to \( \gamma \) (the derivation is in the Appendix), we get an upper bound on the average probability of symbol error as

\[
\bar{P}_e \leq 4 \left( 1 + \frac{\sigma^2}{\sigma_c^2} \right)^{-1} \exp \left( -\frac{\mu}{\sigma_c^2 + \sigma^2} \right).
\]

The non-exponential factor in the bound above is an upper bound on the average symbol error probability for a single transmit and receive antenna system in a Rayleigh
fading channel. Replacing $\mu$ and $\sigma^2$ by their expressions above (4.61), we obtain the upper bound
\[ P_e \leq 4 \left( \frac{N_0}{\sigma_a^2 + N_0} \right) \exp \left( -\frac{M|\mu_a|^2}{\sigma_a^2 + N_0} \right). \] (4.63)

This upper bound for the average probability of symbol error shows that in a Rician fading channel ($\mu_a \neq 0$), the average probability of error decreases exponentially and approaches zero as the number of transmit antennas increases. For the Rayleigh fading case ($\mu_a = 0$), the above upper bound does not change with the number of antennas. However, as shown in [18], and also in [20, 22], the use of linear multiple transmit antennas in Rayleigh fading channels, with linear prefilters which form a unitary tap matrix, does result in a decreased bit error rate, although this error rate does not approach zero (3.37) as the number of transmit antennas goes to infinity, and there are diminishing returns beyond a certain number of antennas. Our analysis shows that multiple transmit antennas are much more effective in Rician fading channels, as can be expected due to the power gain of the direct or strong paths between the transmit and receive antennas.

The reason for this distinction between Rician and Rayleigh flat fading channels can be made more obvious by examining the receive power variance normalized to the average receive power as in Section 3.1. In a Rician flat fading channel, since $\mu_a \neq 0$, the equivalent fade coefficient for transmit diversity (3.1) is a special complex Gaussian random variable with the variance $\sigma_a^2$ and non-zero mean $\mu = \sqrt{M} \mu_a$. Therefore the normalized receive power variance $\sigma_P^2$ for transmit diversity is
\[ \sigma_P^2 = \frac{\text{var} \left( |a_i|^2 \right)}{(E|a_i|^2)} = \frac{2|\mu|^2\sigma_a^2 + \sigma_a^4}{(\mu^2 + \sigma_a^2)^2} = 1 - \left( 1 + \frac{\sigma_a^2}{|\mu|^2} \right)^{-1}. \] (4.64)

Since $|\mu|^2 \to 0$ as $M \to \infty$, the normalized receive power variance in a Rician fading channel decreases to zero as the number of transmit antennas increases, i.e., $\sigma_P^2 \to 0$ as $M \to \infty$, unlike in a Rayleigh fading channel (3.3) where this variance does not depend on the number of transmit antennas. Thus multiple transmit antennas in a Rician fading channel reduce the variance of the instantaneous receive signal power as the number of antennas increases, hence they drive the error rate to zero.

### 4.6 Summary of known channel coefficients case

In this chapter we have developed the minimum mean square error optimization criterion in the frequency domain and derived the optimum linear equalizer when channel
coefficients are known at the receiver for both Rayleigh and Rician fading channels. We then established the optimum set of linear prefilters for multiple transmit antennas in a Rician fading channel, which minimizes a lower bound on the average MSE and makes the MSE converge to this minimum lower bound in a mean-square-sense as number of transmit antenna increases. For rectangular QAM signals, an upper bound on the probability of symbol error was developed which shows that for a Rician flat fading channel, with the optimum linear equalizer and optimum set of linear prefilters, the error probability will decrease to zero exponentially. Thus multiple transmit antennas in a Rician flat fading channel can significantly improve system performance even with simple linear processing at both the transmitter and receiver.

This result concerning the effect of multiple transmit antennas on Rician fading channels is quite different from that for Rayleigh fading channels, which, as shown in the previous two chapters, tends to mitigate the fading and drive the channel to the equivalent non-fading AWGN channel in both capacity and performance. But the bit error rate for a Rayleigh fading channel cannot be driven to zero by only increasing the number of transmit antennas. In a Rician fading channel, multiple transmit antennas also have the same mitigating effect on fading, but in addition, the non-zero power gain of the strong paths (often the direct paths if they exist) drive the error probability towards zero as the number of antennas increases, even when this power gain is not large.

Nevertheless, results in both fading cases support the conjecture that multiple transmit antennas reduce the variation in the received signal. The intuitive reason is that since the channel is random, with multiple antennas using appropriate signal processing, the probability that all paths are in a deep fade is less than the probability of a single path being in a deep fade, therefore the probability of the receive signal falling below a certain threshold when using multiple antennas is less than when using a single transmit antenna.
Chapter 5

Equalization of An Unknown Wireless Channel using Multiple Transmit Antennas

In this chapter we consider the case where the channel coefficients are not known at the receiver. The motivation for studying this case is that, in practice, it may not always be possible to measure the channel, especially if it is changing rapidly. Therefore we are interested in determining if multiple transmit antennas can also improve system performance when the channel coefficients are not available at the receiver. To simplify the problem, we assume that the channel is wide sense stationary, and the mean and variance of the channel coefficients are known. These parameters can be obtained, for example, by measuring the channel extensively at the start to derive the mean and variance, then these parameters are used throughout the communication duration without re-measuring the channel, which can be more easily accomplished in practice than measuring the channel during every frame. Again we will assume an ideal case where the mean and variance of the channel coefficients have been accurately measured and are available at the receiver, while the transmitter does not have this information.

We study the same linear structure as in previous chapters where the transmit signal is processed by separate linear prefilters before being sent from each transmit antenna. At the receiver, linear equalization is used, followed by sampling at the symbol rate to detect the transmit symbols. We focus on frequency non-selective Rician fading channels, where the channel is multiplicative with random complex Gaussian coefficients. Since only the mean and the variance of channel coefficients are acquired
at the receiver, our analysis will apply to both slow and fast fading, unlike in the
previous chapter where the channel was assumed to be reasonably slow fading (quasi-
stationary) in order to enable coefficient measurements. However, the signal model
developed in the last chapter is fully applicable in this case, thus we will not repeat it
here.

The same optimization criterion of minimizing the mean square error is used. Re-
results of our analysis show that the optimum linear equalizer is a function of the mean
and variance of the channel coefficients, and is therefore deterministic. Moreover, these
results are only valid when the channel mean is non-zero, which corresponds to the
Rician fading case. The same set of optimum linear prefilters for a Rician fading chan-
nel as in the known channel coefficients case is obtained, which minimizes the average
mean square error obtained with the optimum equalizer, where all the prefilters have
the same frequency response equal to the square root of a Nyquist pulse spectrum. It
is found that an error bound on the average probability of symbol error for rectangu-
lar QAM signals also decreases exponentially towards zero as the number of antennas
increases. This is a powerful result, since it shows that in a Rician fading channel,
even when the channel coefficients are not fully known at the receiver, using multiple
transmit antennas can still drive the error rate arbitrarily small.

The chapter is organized as follows: Section 5.1 analyzes the optimum linear equal-
ze for a mean square error criterion. In Section 5.2, we derive the optimum set of
linear prefilters that minimize the average mean square error. An upper bound on the
probability of symbol error for rectangular QAM signals is developed in Section 5.3.
Finally, we conclude this chapter in Section 5.4.

5.1 Optimum equalizer

We use the discrete signal model and optimization criterion developed in the previous
chapter. The aim is to find a linear equalizer that minimizes the mean square error
expression (4.31), which is equivalent to minimizing $J(\omega)$ in (4.33). In this case, the
equivalent channel $A^e(\omega)$ is unknown and random, therefore we cannot find the equal-
ze to minimize $J(\omega)$ instantaneously. Instead we can minimize the average value of
$J(\omega)$, and by doing so, effectively seeking to minimize the average mean square error
with respect to the channel coefficients.

Since the channel coefficients’ mean $\mu_n$ and variance $\sigma_n^2$ are known at the receiver,
the mean $\mu^*_A(\omega)$ (4.20) and variance $\sigma^2_A(\omega)$ (4.22) of the equivalent channel response $A^*(\omega)$ are known. We have the following theorem.

**Theorem 5.1.** The optimum linear equalizer that minimizes the average of the mean square error (4.31) with respect to fading channel coefficients for the system described in (4.13), has a frequency spectrum of its sampled taps given by

$$B^*_0(\omega) = \frac{\mu^*_A(\omega)}{|\mu^*_A(\omega)|^2 + \sigma^2_A(\omega) + \frac{\sigma^2_N}{\sigma^2_c}}$$

(5.1)

where $\mu^*_A(\omega)$ and $\sigma^2_A(\omega)$ are the mean and variance of the discrete channel frequency spectrum $A^*(\omega)$ and $\sigma^2_c$ and $\sigma^2_N$ are variances of the input symbols and the sampled noise respectively.

**Proof.** Here we use the shorthand notation $\mu_A = \mu^*_A(\omega)$ and $\sigma^2_A = \sigma^2_A(\omega)$. Taking the expected value of $J(\omega)$ (4.33) with respect to channel coefficients, we obtain

$$E_A[J(\omega)] = \left(\sigma^2_c \left(|\mu_A|^2 + \sigma^2_A\right) \sigma^2_N\right) |B|^2 - \sigma^2_c \left(\mu_A B + \mu_A^* B^*\right) + \sigma^2_c.$$  (5.2)

Again, applying the inequality of the form (4.36), we have

$$E_A[J(\omega)] \geq \left(\sigma^2_c \left(|\mu_A|^2 + \sigma^2_A\right) \sigma^2_N\right) |B|^2 - 2 \sigma^2_c |\mu_A| |B| + \sigma^2_c = f(|B|)$$

with equality only when

$$\mu_A B = -\mu_A B_R.$$  (5.3)

To find $|B|$ that minimize the right hand side expression $f(|B|)$, set the first derivative of this function with respect to $|B|$ to zero, i.e.

$$\frac{df(|B|)}{d|B|} = 2|B| \left(\sigma^2_c \left(|\mu_A|^2 + \sigma^2_A\right) \sigma^2_N\right) - 2 \sigma^2_c |\mu_A| = 0$$

which gives

$$|B| = \frac{\sigma^2_c |\mu_A|}{\sigma^2_c \left(|\mu_A|^2 + \sigma^2_A\right) \sigma^2_N}.$$  

It can be seen that the second derivative of $f(|B|)$ is positive, thus this value of $|B|$ minimizes $f(|B|)$. Coupled with the equality condition (5.3), we obtain the optimum equalizer in (5.1).

The equalizer in this case is deterministic. Result (5.1) shows that the equalizer is valid only when the equivalent channel mean $\mu_A$ is non-zero, consequently the channel
coefficients’ mean $\mu_\alpha$ must be non-zero. Thus the result is only applicable for the Rician fading channel but not the Rayleigh fading case. With our linear structure, where the equivalent channel mean and variance are given in (4.20) and (4.22), the optimum equalizer spectrum becomes

$$B^*_\alpha(\omega) = \frac{T \left( G^*(\omega) \cdot \mu_\alpha \sum_{i=1}^{M} H^*_i(\omega) \right)}{|\mu_\alpha|^2 + \sigma^2_c(\omega) + \zeta_0}$$  \hspace{1cm} (5.4)

where $\zeta_0 = N_0/\sigma^2_c$. Thus the equalizer is in the form of a matched filter which is matched to the frequency response of the prefilters and the modulation pulse.

### 5.2 Minimum mean square error and optimum prefilters

Having established the optimum equalizer, again we set out to find the optimum set of prefilters which minimizes the MSE obtained with this optimum equalizer. With the equalizer (5.1), the average of $J(\omega)$ in (5.2) becomes

$$E[J(\omega)] = \frac{\sigma^2_c \sigma^2_\alpha(\omega)^2 + \sigma^2_N}{|\mu^*_\alpha(\omega)|^2 + \sigma^2_A(\omega)^2 + \frac{\sigma^2_c}{\sigma^2_c}} = \frac{\sigma^2_c \sigma^2_\alpha + N_0}{|\mu_\alpha|^2 \left| \sum_{i=1}^{M} H^*_i(\omega) \right|^2 + \sigma^2_\alpha + \zeta_0}$$  \hspace{1cm} (5.5)

where the second equality results from replacing $\mu^*_\alpha(\omega)$ and $\sigma^*_A(\omega)$ by their expressions (4.20) and (4.22), taking note of square spectrum of the modulation pulse in (4.21). Thus the average mean square error obtained with the optimum equalizer (5.1) is

$$\bar{\varepsilon}_o = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{\sigma^2_c \sigma^2_\alpha + N_0}{|\mu_\alpha|^2 \left| \sum_{i=1}^{M} H^*_i(\omega) \right|^2 + \sigma^2_\alpha + \zeta_0} \, d\omega.$$  \hspace{1cm} (5.6)

Comparing this average MSE expression with the lower bound on the average optimum MSE (4.45) in the case of known channel coefficients at the receiver, we see that there is an extra component $\sigma^2_c \sigma^2_\alpha$ in the numerator, caused by fluctuation in the channel and which is therefore proportional to the input power and the channel's variance. This is the penalty when we do not know the channel coefficients at every point in time and thus cannot adjust the equalizer accordingly to mitigate the effect of channel fluctuation.

We now focus on finding the set of $H^*_i(\omega)$ that minimizes this average MSE (5.6). Obviously this is equivalent to maximizing the square term of the sum of $H^*_i(\omega)$ at the denominator. Again applying inequality (4.46), we obtain the minimum average MSE as

$$\bar{\varepsilon}_{min} = \frac{\sigma^2_c \sigma^2_\alpha + N_0}{M |\mu_\alpha|^2 + \sigma^2_\alpha + \zeta_0}$$  \hspace{1cm} (5.7)

119
where this minimum value is obtained if and only if the $H_i^+(\omega)$ are all equal. Thus the same set of $H_i^+(\omega)$ as in (4.42) is optimal.

As a result, the set of optimum prefilters is the same for both cases of known and unknown channel coefficients at the receiver for a Rician fading channel. The average mean square error in case of unknown coefficients (5.7) is larger than the lower bound on the average MSE in the known channel case (4.47) by a term in the numerator that equals the fluctuation in the received power due to the unknown fluctuation in the channel. However, since in the known channel coefficients case, the MSE converges to its lower bound in a mean square sense when the number of transmit antennas goes to infinity, it is conjectured that the difference between the average values of the MSE in both cases given a finite number of transmit antennas will be less than this power fluctuation.

From (5.7), it is also clear that more transmit antennas helps to reduce the average mean square error. As the number of transmit antennas $M$ increases to infinity, the average MSE approaches zero, which indicates that the symbol estimate converges to the symbol value in the mean square sense, i.e.

$$
\hat{c}_k \xrightarrow{m,n} c_k \quad \forall k.
$$

Note that in the case of known channel coefficients, convergence of the instantaneous MSE to its lower average bound in mean square sense (4.51) also implies symbol estimate convergence in 4th order (4.52), therefore convergence in the known channel coefficients case is stronger. Here we do not attempt to analyze the convergence characteristic of the instantaneous MSE. In the next section, we are interested in how the number of transmit antennas affects the system performance in terms of probability of symbol error.

### 5.3 Evaluating performance by bound on probability of error

Again we study an upper bound on probability of symbol error for a rectangular QAM signal. The general error bound (4.58) still applies, though the optimum mean square error in this case is of different form and theorem 4.4 therefore can not be applied. However, noticing special the form of square spectrum of equivalent channel in (4.23) that, with the optimum prefilters (4.42), $|A^e(\omega)|^2$ is independent of frequency $\omega$. From
this, we have the following theorem.

**Theorem 5.2.** Let $c_k$ be a sequence of wide-sense stationary, mutually uncorrelated rectangular complex information symbols that is the input to a channel $a[k]$ with frequency response $A^s(\omega)$ given in (4.18). The receive signal is linearly equalized (4.13) by the optimum equalizer $b[k]$ given in Theorem 5.1. Then, with the optimum set of prefilter given in (4.42), the probability of symbol error is upper bounded by

\[ P_e \leq 4 \exp\left(-\frac{|A|^2}{\sigma_N^2}\right) \]

where $|A|^2$ is the square spectrum of the channel and $\sigma_N^2$ is the noise variance.

**Proof.** The arguments are similar to the proof of Theorem 4.4 up to the general bound in (4.58). Now note that with the optimum prefilter (4.42), the optimum equalizer (5.4) becomes

\[ B_0^s(\omega) = \frac{TM \mu_a^* \cdot G^s(\omega)^* H^s(\omega)^*}{M|\mu_a|^2 + \sigma_a^2 + \zeta_0} \]  

(5.8)

where $H^s(\omega)$ is the frequency response of the optimum prefilter in (4.42). From the expression for $A^s(\omega)$ in (4.18), together with the Nyquist condition of $G^s(\omega)$ (4.21) and the optimum prefilter (4.42), we have

\[ F^s(\omega) = A^s(\omega) \cdot B^s(\omega) = \frac{\mu_a^* \sum_{i=1}^M a_i}{M|\mu_a|^2 + \sigma_a^2 + \zeta_0}. \]  

(5.9)

Thus $F^s(\omega)$ is independent of frequency $\omega$, which leads to the samples

\[ f[k] = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} F^s(\omega)e^{-j\omega kT} d\omega = \begin{cases} F^s & k = 0 \\ 0 & \text{otherwise} \end{cases} \]  

(5.10)

where $F^s$ denotes the value of $F^s(\omega)$ given in (5.9).

Thus the error bound in (4.58) becomes

\[ P_e < 4 \exp\left(-\frac{|F^s|^2}{\sigma_w^2}\right) \]  

(5.11)

where $F^s$ in (5.9) is a function of the random variables $a_i$ and therefore is also random. But $f[k] = a[k] \ast b[k]$ and $w[k] = n[k] \ast b[k]$, thus $F^s(\omega) = A^s(\omega) \cdot B^s(\omega)$ and $\sigma_w^2 = \sigma_N^2 \cdot |B^s(\omega)|^2$. Substituting these into (5.11) we obtain the theorem.

The key argument leading to the error bound in Theorem 5.2 is that the frequency response $F^s(\omega) = A^s(\omega) \cdot B^s(\omega)$ is independent of the frequency $\omega$. Since with the
optimum set of prefilters (4.42), \(|A^*(\omega)|^2\) is independent of frequency \(\omega\), this means that any linear filter that is matched to the equivalent channel \(A^*(\omega)\) will satisfy this bound. The linear equalizer in the case of known channel coefficients (4.35) therefore also satisfies the error bound in the above theorem. However, the error bound in Theorem (4.4) is more general in that it is applicable for any channel \(A^*(\omega)\) and any set of prefilters \(H_i^*(\omega)\), given the optimum equalizer in Theorem 4.2, whereas the error bound in Theorem 5.2 is specific to the optimum set of prefilters in (4.42), and only for case of flat fading channel, i.e. the channel coefficients \(a_i\) are frequency independent.

Applying this theorem, noticing the square channel spectrum in (4.23) and the optimum prefilters in (4.42), the error bound becomes

\[
P_e < \exp 4 \left( -\frac{1}{MN_0} \sum_{i=1}^{M} |a_i|^2 \right). \tag{5.12} \]

Since in both cases of known and unknown channel coefficients, the set of optimum prefilters are the same and the optimum equalizers are matched to the channel, the symbol error probability bounds are also the same (4.60). Thus the same bound on average error probability (4.62) is obtained, and it decreases exponentially to zero as the number of transmit antennas increases. Although the actual error probabilities were not compared, this shows that in a Rician flat fading channel, system performance is not greatly affected when the channel coefficients are not available at the receiver.

### 5.4 Summary of the unknown channel coefficients case

In this chapter we have developed the optimum equalizer and optimum set of prefilters for a Rician fading channel, when the channel coefficients are not known at the receiver but only the mean and variance are available. The equalizer is in the form of a deterministic matched filter which is matched to the mean of the equivalent channel, whereas the optimum prefilters are the same as in the case of known channel coefficients. This leads to the same bound on the probability of symbol error holding in both cases, which shows that the average error probability will approach zero exponentially as number of transmit antennas increases. These results imply that in a Rician flat fading channel, compatible performance can be obtained without the overhead of measuring the channel coefficients, thus the transmission rate can be higher than when some fraction of the transmission time is needed for channel measurement, as may be required in the known channel coefficient case.
Our results here are only for flat fading channels where the channel coefficients are frequency independent. In a frequency-selective channel, a different error bound in the case of unknown channel coefficients is expected. However, with a frequency-selective channel, additional gain arises due to the diversity in the spectrum, and it can be reasonably expected that better performance will be obtained in such a way that the error probability will decrease at a faster rate with either more frequency taps or more antennas. Analysis of this case is left to further research.
Chapter 6

Conclusion

6.1 Thesis summary

In this thesis we have analyzed transmit antenna diversity in a flat Rayleigh or Rician fading channel. Results for the Rayleigh fading case include the information theoretic capacity presented in Section 2.3, and a comprehensive review of various diversity techniques in Chapter 3. Using a linear antenna processing structure where the data is processed by linear filters prior to the transmit antennas, together with a linear equalizer at the receiver, results show that the optimum prefilters have tap weights that form the rows of a unitary matrix, and the flat and slow Rayleigh fading channel converges to a non-fading AWGN channel in a mean-square sense. The asymptotic bit error rate for QPSK signals when the number of transmit antennas goes to infinity is a negative exponential function of only the average receive SNR. With the optimum equalizer derived in Chapter 4, which agrees with the optimum equalizer in the asymptotic transmit diversity Rayleigh fading case, the optimum linear antenna processing structure for a Rayleigh flat and slow fading channel is completed.

For a Rician flat fading channel, optimum results are derived in Chapters 4 and 5 for both known and unknown channel cases. For the known channel case, the optimum linear equalizer is the same for both Rayleigh and Rician fadings. However, the set of optimum linear prefilters in Rician fading channels is different from that in Rayleigh fading channels, and consists of linear filters that have a spectrum equal to the square root of a Nyquist pulse spectrum. With these optimum prefilters and equalizer, the MSE converges in mean-square to a minimum average lower bound, implying convergence of symbol estimates in the 4th moment. An upper bound on the
average probability of symbol error for rectangular QAM signals decreases to zero ex-
ponentially as the number of transmit antennas increases. This shows that transmit

diversity has a much more significant effect for Rician fading channels than for Rayleigh
fading channels.

For the unknown channel case, similar results for Rician flat fading channels are ob-
tained, assuming the channel coefficient mean and variance are available at the receiver.
With the same set of the optimum prefilters as in the known channel case, the same
upper bound on the average probability of symbol error for rectangular QAM signals is
obtained. This shows that transmit diversity can still achieve significant diversity gain
in Rician flat fading channels even when the channel is not fully known at the receiver.

6.2 Further research directions

The work here gives rise to some possible further research directions

- The same linear structure can be extended to frequency selective channels. Since
  in frequency selective channels, additional diversity arises in the spectral domain,
  it can reasonably be expected that more diversity gain will be obtained. The
  optimum linear equalizer in both cases of known and unknown channel will be in
  the same forms as derived in this thesis (Chapters 4 and 5). Assuming that the
delays between resolvable paths between different pairs of transmit and receive
antennas are the same and fixed, the set of optimum linear prefilters will be
the same as for flat fading channels. The error bound evaluation will be more
complicated since the square spectrum of the equivalent channel is now dependent
on frequency.

- Another possibility is to investigate other types of equalization such as the deci-
sion feedback equalizer (DFE). This also promises better performance than linear
equalizers but in exchange for more complicated processing. Decision feedback
equalizers have been applied to receive diversity [34, 35] and achieved a smaller
MSE compared to linear equalizers. The application of DFE in transmit diversity
therefore is conjectured to reduce the MSE or increase its convergence rate.

- Combination of transmit and receive diversity holds the potential for substantial
capacity and performance improvement as shown in Chapter 2. There has been
some work done in this area for Rayleigh slow and flat fading channels [26, 30,
but there is still scope for further research into combining spatial diversity. A transmit and receive diversity system using linear antenna prefilter at the transmitter and a linear equalizer for the receive signal at each receive antenna can be established. Since transmit diversity stabilizes the fading channel and also reduces the error probability in case of Rician fading, and receive diversity using linear equalizers has been shown to improve the system performance in term of the error probability and outage probability [34, 35], the combination of both transmit and receive diversity using linear structure can substantially improve the performance and transmission rate. And again, non-linear processing or additional spectral diversity can be applied to further enhance the performance at the trade off of simplicity.
Appendix

Lemma 1. A special complex Gaussian random variable as defined in Definition 2.1 has the following properties

a. A linear combination of independent special complex Gaussian random variables is also a special complex Gaussian random variable.

b. Let $X$ be a special complex Gaussian random variable with mean $\mu$ and variance $\sigma^2$. Then the random variable $|X|^2$ will have mean and variance given by

$$E(|X|^2) = |\mu|^2 + \sigma^2$$

$$\text{var}(|X|^2) = 2|\mu|^2\sigma^2 + \sigma^4$$

Proof. Let $\mathcal{N}(\mu, \sigma^2)$ denote a real Gaussian random variable, and $\mathcal{N}^c(\mu, \sigma^2)$ denote a special complex Gaussian random variable, with mean $\mu$ and variance $\sigma^2$.

a. We will prove the result for the sum of two independent special complex Gaussian random variables, which can be generalized for linear combination of an arbitrary number of such variables.

Suppose $X \in \mathcal{N}^c(\mu_x, \sigma^2_x)$ and $Y \in \mathcal{N}^c(\mu_y, \sigma^2_y)$ are independent. Express $X$ and $Y$ in terms of their real and imaginary parts as

$$X = X_1 + jX_2, \quad Y = Y_1 + jY_2$$

then $X_1 \in \mathcal{N}(\mu_{x1}, \sigma^2_{x1}/2)$ and $X_2 \in \mathcal{N}(\mu_{x2}, \sigma^2_{x2}/2)$ are independent, and $Y_1 \in \mathcal{N}(\mu_{y1}, \sigma^2_{y1}/2)$ and $Y_2 \in \mathcal{N}(\mu_{y2}, \sigma^2_{y2}/2)$ are also independent. The sum of the two random variables $X$ and $Y$ is

$$Z = X + Y = (X_1 + Y_1) + j(X_2 + Y_2).$$

Since $X$ and $Y$ are independent, the real and imaginary parts of $Z$ which are $Z_1 = X_1 + Y_1$ and $Z_2 = X_2 + Y_2$ are independent Gaussian random variables with variances
equal to sum of individual component variances respectively, and therefore are both
equal to \((\sigma_1^2 + \sigma_2^2)/2\). Thus \(Z\) is also a special complex Gaussian random variable
according to Definition 2.1.

\(b.\) Express \(X\) in term of its real and imaginary parts \(X = X_1 + jX_2\), then \(X_1 \in \mathcal{N}(\mu_1, \sigma_1^2/2)\) and \(X_2 \in \mathcal{N}(\mu_2, \sigma_2^2/2)\) are independent and

\[ \mu_1^2 + \mu_2^2 = |\mu|^2. \]

The \(k\)th moments of a Gaussian random variable \(X_1\) is given by [1]

\[ E[X^k] = \sum_{i=0}^{k} \binom{k}{i} \mu_1^i m_{k-i} \]

where \(m_k\) is the \(k\)th central moment of \(X_1\) given by

\[ m_k = E[(X - \mu_k)^k] = \begin{cases} 1 \cdot 3 \cdots (k-1) \left(\frac{\sigma^2}{2}\right)^k & k \text{ even} \\ 0 & k \text{ odd} \end{cases} \]

Since the fourth moment of \(X\) is \(E(|X|^4) = E[(X_1^2 + X_2^2)^2]\), substituting the appropriate expressions of second and fourth moments of \(X_1\) and \(X_2\), we get

\[ E(|X|^4) = |\mu|^4 + 4\sigma^2|\mu|^2 + 2\sigma^4. \]

Together with \(E(|X|^2) = E[X_1^2 + X_2^2] = |\mu|^2 + \sigma^2\), we arrive at the results of (b). \(\square\)

**Lemma 2.** Given \(f(t)\) as a convolution between two other signals \(f(t) = u(t) * v(t)\),
the \(T\)-interval frequency response of \(f(t)\) can be written as \(F_T(\omega) = V_T(\omega) \cdot U(\omega)\).

**Proof.** By Definition 4.1, the frequency response of a signaling interval of \(f(t)\) is

\[ F_T(\omega) = \int_{-T/2}^{T/2} f(t)e^{-j\omega t}dt = \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} u(\tau)v(t-\tau)d\tau \cdot e^{-j\omega t}dt \]

\[ = \int_{-\infty}^{\infty} \int_{-T/2}^{T/2} v(t-\tau)e^{-j\omega(t-\tau)}d\tau \cdot u(\tau)e^{-j\omega t}dt = V_T(\omega) \cdot U(\omega) \]

The order of \(U\) and \(V\) can be interchanged. This completes the proof. \(\square\)

**Lemma 3.** Let \(A\) and \(B\) be two complex numbers, then the following inequality holds

\[ AB + A^*B^* \leq 2|A||B| \]

with equality if and only if \(A_1B_R = -A_RB_1\).
Proof. Express $A$ and $B$ in the form $A = A_R + jA_I$ and $B = B_R + jB_I$ then


Consider the difference between the square of the two sides of the inequality

$$4|A|^2|B|^2 - (AB + A^*B^*)^2 = 4(A_R^2 + A_I^2)(B_R^2 + B_I^2) - 4(A_RB_R - A_IB_I)^2$$

$$= 4(A_I^2B_R^2 + A_R^2B_I^2 + 2A_RB_R A_IB_I) = 4(A_IB_R + A_RB_I)^2 \geq 0$$

This leads to

$$|AB + A^*B^*| \leq 2|A||B|$$

with equality only when $A_IB_R = -A_RB_I$. \hfill \Box

**Lemma 4.** Proof of inequality (4.57)

$$Pr[z > f(0)] < \exp\left(\frac{-f[0]^2}{2(\sigma_{w_r}^2 + \sigma_a^2 \sum_{k \neq 0} |f[k]|^2)}\right).$$

Proof. This proof is derived from similar proof in [36] for case when the $c_k$ are real. Since $z$ is the sum of independent, zero-mean random variables (4.56), the Chernoff bound applies

$$Pr[z > x] \leq e^{-\lambda x} E[e^{\lambda w_r[k]} \prod_{l \neq 0} E[e^{\lambda f[l]}] \forall \lambda > 0. \tag{6.1}$$

Since $w_r[k]$ is a zero-mean Gaussian noise random variable with variance $\sigma_{w_r}^2$, we have

$$E[e^{\lambda w_r[k]}] = \exp\left(\frac{1}{2} \lambda^2 \sigma_{w_r}^2\right),$$

and using the probability density function of $z[l]$ in (4.55)

$$E[e^{\lambda z}] = \frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \exp\left(\alpha_j f_r[l] + \alpha_k f_i[l]\right) \lambda = \frac{2}{N} \sum_{j=\frac{N}{2}+1}^{N} \cosh(\lambda f_r[l] \alpha_j) + \frac{2}{N} \sum_{k=\frac{N}{2}+1}^{N} \cosh(\lambda f_i[l] \alpha_k)$$

with $\alpha_j$ as defined in (4.54). Employ the following inequality [36]

$$\frac{2}{N} \sum_{j=\frac{N}{2}+1}^{N} \cosh(\lambda f_r[l] \alpha_j) < \exp\left(\frac{1}{2} \sigma_a^2 \lambda^2 f_r[l]^2\right)$$

where $\sigma_a^2$ is the variance of the set $\{\alpha_j\}$, i.e.

$$\sigma_a^2 = \frac{1}{N} \sum_{j=1}^{N} \alpha_j^2 .$$
Then (6.1) becomes
\[
Pr[z > x] < \exp \left( -\lambda x + \frac{1}{2} \lambda^2 \sigma^2_{w_i} + \frac{1}{2} \lambda^2 \sigma^2_a \sum_{l \neq 0} (f_l[l]^2 + f_i[l]^2) \right) \quad \forall \lambda > 0.
\] (6.2)

Finding the value of \( \lambda \) which maximizes the right hand side of above expression, by taking first derivative of the expression with respect to \( \lambda \), setting it to zero, and verifying that the second derivative is positive, we obtain

\[
\lambda = \frac{x}{\sigma^2_{w_i} + \sigma^2_a \sum_{l \neq 0} |f[l]|^2}.
\]

Substituting this value of \( \lambda \) into (6.2) yields the inequality (4.57).

**Lemma 5.** Derive the upper bound on average probability of error in (4.62)
\[
P_e \leq 4 \left( 1 + \frac{\sigma^2}{\sigma^2_c} \right)^{-1} \exp \left( -\frac{\mu}{\sigma^2_c + \sigma^2} \right).
\]

**Proof.** Since the \( a_i \) are special complex Gaussian random variables with the same mean \( \mu_a \) and variance \( \sigma^2_a \), the receive SNR

\[
\gamma = \frac{\sigma^2_c}{N_0 \cdot M} \sum_{i=1}^M a_i^2
\]
is a chi-square random variable with two degree with freedom. The probability density function of \( \gamma \) is given by

\[
p(\gamma) = \frac{1}{\sigma^2} \exp \left( -\frac{\gamma + \mu}{\sigma^2} \right) I_0 \left( \frac{2\sqrt{\mu\gamma}}{\sigma^2} \right)
\]
where \( \mu \) and \( \sigma^2 \) are given in (4.61), and \( I_0(\cdot) \) is the zero-th order modified Bessel function of first kind, which has the series expansion

\[
I_0(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{4^k k! \Gamma(k+1)}, \quad x \geq 0
\]
with \( \Gamma(\cdot) \) is the Gamma function, \( \Gamma(k+1) = k! \) when \( k \) is an integer.

Therefore the upper bound on the average error probability is

\[
U = E \left[ 4 \exp \left( -\frac{\gamma}{\sigma^2_c} \right) \right] = \int_{0}^{\infty} \frac{4 \exp \left( -\frac{\gamma + \mu}{\sigma^2} \right) \exp \left( -\frac{\gamma}{\sigma^2_c} \right)}{\sigma^2} \sum_{k=0}^{\infty} \frac{\gamma^k \mu^k}{4^k k! \Gamma(k+1)} d\gamma.
\]

Using integration by parts and noting the recursion of the following integral, we have

\[
\int_{0}^{\infty} e^{-ax} x^b \frac{dx}{k!} = \frac{e^{-b}}{a^{k+1}} \quad \forall a, b.
\]

Thus, using the series expansion of \( e^x \), we obtain

\[
U = \frac{4}{\sigma^2} e^{-\mu/\sigma^2} \sum_{k=0}^{\infty} \frac{\mu^k}{\sigma^2 k! \Gamma(k+1)} \left( \frac{\sigma^2_c \cdot \sigma^2}{\sigma^2_c + \sigma^2} \right)^{k+1} = \frac{4}{\sigma^2} e^{-\mu/\sigma^2} \left( \frac{\sigma^2_c \cdot \sigma^2}{\sigma^2_c + \sigma^2} \right) \exp \left( \frac{\mu \sigma^2_c}{\sigma^2 (\sigma^2_c + \sigma^2)} \right).
\]

Rearranging the exponential components we obtain the upper bound of (4.62). \( \square \)
Bibliography


